# How to determine structures when single crystals cannot be grown: opportunities for structure determination of molecular materials using powder diffraction data

Kenneth D. M. Harris* and Eugene Y. Cheung

*School of Chemistry, Cardiff University, P.O. Box 912, Cardiff, Wales, UK CF10 3TB.*
*E-mail: HarrisKDM@cardiff.ac.uk; Fax: +44–2920–874–030; Tel: +44–2920–870–133*

Many crystalline solids cannot be prepared as single crystals of sufficient size and/or quality for structure determination to be carried out using single crystal X-ray diffraction techniques. In such cases, when only polycrystalline powders of a material are available, it is necessary instead to tackle structure determination using powder X-ray diffraction. This article highlights recent developments in the opportunities for determining crystal structures directly from powder diffraction data, focusing on the case of molecular solids and giving particular attention to the most challenging stage of the structure determination process, namely the structure solution stage. In particular, the direct-space strategy for structure solution is highlighted, as this approach has opened up new opportunities for the structure determination of molecular solids. The article gives an overview of the current state-of-the-art in structure determination of molecular solids from powder diffraction data. Relevant fundamental aspects of the techniques in this field are described, and examples are given to highlight the application of these techniques to determine crystal structures of molecular materials.

## 1 Introduction

Few chemists would argue against the assertion that single crystal X-ray diffraction is the most powerful approach for the determination of structural information at the atomic level, and indeed, many important advances in chemical, biological and physical sciences have arisen through the use of this technique. Nevertheless, it is important to recall that the requirement for a single crystal sample of appropriate size and quality imposes a natural limitation on the scope of this technique. Unfortunately, many crystalline solids can be prepared only as microcrystalline powders, and therefore cannot be studied using single crystal X-ray diffraction techniques (including the rapidly developing synchrotron-based microcrystal X-ray diffraction techniques). How then

do we progress towards understanding the structural properties of these materials? The most direct approach is to use powder X-ray diffraction data, although it is important to recognize that the process of carrying out structure determination from powder diffraction data is substantially more challenging than structure determination from single crystal diffraction data.

Although single crystal and powder diffraction patterns contain essentially the same information, the diffraction data are distributed in three-dimensional space in the single crystal diffraction pattern, whereas the diffraction data are "compressed" into one dimension in the powder diffraction pattern. As a consequence, there is usually considerable overlap of peaks in the powder diffraction pattern. Such peak overlap obscures information on the positions and intensities of the diffraction maxima, and the intrinsic difficulty of obtaining

*Kenneth D. M. Harris is a graduate of the University of St. Andrews (BSc, 1985) and the University of Cambridge (PhD, 1988, supervisor Professor Sir John Meurig Thomas, FRS). He held academic positions at the Universities of St. Andrews, London (UCL) and Birmingham before moving in 2003 to his present appointment at Cardiff University. His research covers several areas relating to the physical chemistry of solids, with some focus at present on the development and application of new techniques for structure determination from powder diffraction data. He has been awarded the Meldola (1991), Marlow (1996), Corday-Morgan (1999) and Structural Chemistry (2001) Medals of the RSC, and the Physical Crystallography Prize (1997) of the BCA. He has held positions as Visiting Professor in Spain, Taiwan, Japan and the USA.*

**Kenneth Harris**

*Eugene Y. Cheung studied for a combined B.Sc. (Hons) in Chemistry and Biochemistry at the University of British Columbia in Canada, and continued to complete his PhD in solid state reactions and X-ray crystallography under the supervision of Professors John R. Scheffer and James Trotter. In 2001 he moved to the United Kingdom to join Professor Kenneth D. M. Harris in the study of powder X-ray diffraction.*

*His current research is focused on developing new techniques to determine crystal structure directly from powder diffraction data, and to rationalize the structural behaviour of materials based upon information from X-ray diffraction, thermal analysis, and NMR techniques.*

**Eugene Cheung**

reliable information of this type from the powder diffraction pattern can impede (or in some cases prohibit) the process of carrying out crystal structure determination. As molecular solids typically have large unit cells and low symmetry, the problem of peak overlap is often particularly severe for such materials, and the process of carrying out structure determination from powder diffraction data is correspondingly more challenging. Indeed, as discussed below, it is only within the last 15 years or so that it has been possible to determine organic molecular structures directly from powder X-ray diffraction data.[1] Clearly, progress in overcoming the peak overlap problem relies either upon finding improved techniques for extracting individual peak intensities from the overlapped data (an area in which significant progress has been made), or the formulation of new structure solution strategies (see Section 4) that allow the experimental powder diffraction data to be used directly "as measured", without the requirement to extract the intensities of individual diffraction maxima.

As many important materials can be prepared only as microcrystalline powders, the availability of reliable procedures for determining crystal structures directly from powder X-ray diffraction data has the potential to make considerable impact in structural sciences. For this reason, much research activity in recent years has been devoted to the development and application of new techniques for carrying out structure determination directly from powder diffraction data, and has led to significant advances in the scope and power of techniques in this field.[1–5]

This article highlights recent developments in the opportunities for determining crystal structures directly from powder diffraction data, focusing on the case of molecular solids and giving particular attention to the most challenging stage of the structure determination process – the structure solution stage. In particular, the direct-space strategy for structure solution[6] is highlighted, as this approach has proven to be particularly applicable in the case of molecular solids. The article gives an overview of the current state-of-the-art in structure determination of molecular solids from powder diffraction data. Relevant fundamental aspects of the techniques used to carry out structure determination from powder diffraction data are described, and examples highlighting the application of these techniques to determine crystal structures of molecular materials are given. While the discussion of fundamentals and the highlighted examples focus on the case of powder X-ray diffraction, it is relevant to note that the techniques discussed are also generally applicable in the case of structure determination from powder neutron diffraction data. Other aspects of the application of powder diffraction techniques are covered in another article within this Special Issue.[7]

## 2  Background to structure determination from powder diffraction data

### 2.1  The relationship between a crystal structure and its diffraction pattern

In the diffraction pattern from a crystalline solid, the *positions* of the diffraction maxima depend on the periodicity of the structure (*i.e.* the dimensions of the unit cell), whereas the *relative intensities* of the diffraction maxima depend on the distribution of scattering matter (*i.e.* the atoms, ions or molecules) within the repeating unit. Each diffraction maximum is characterized by a unique set of integers $h$, $k$ and $l$ (the Miller indices) and is defined by a scattering vector $\boldsymbol{H}$ in 3-dimensional space, given by $\boldsymbol{H} = h\boldsymbol{a}^* + k\boldsymbol{b}^* + l\boldsymbol{c}^*$. The 3-dimensional space in which the diffraction pattern is measured is called "reciprocal space", whereas the 3-dimensional space defining the crystal structure is called "direct space". The basis vectors $\boldsymbol{a}^*$, $\boldsymbol{b}^*$ and $\boldsymbol{c}^*$ are called the reciprocal lattice

vectors, and depend on the crystal structure. A given diffraction maximum $\boldsymbol{H}$ is completely defined by the structure factor $F(\boldsymbol{H})$, which has amplitude $|F(\boldsymbol{H})|$ and phase $\alpha(\boldsymbol{H})$. In the case of X-ray diffraction, the structure factor $F(\boldsymbol{H})$ is related to the electron density $\rho(\boldsymbol{r})$ within the unit cell by the equation

$$F(\boldsymbol{H}) = |F(\boldsymbol{H})| \exp(i\alpha(\boldsymbol{H})) = \int \rho(\boldsymbol{r}) \exp[2\pi i \boldsymbol{H} \cdot \boldsymbol{r}] \, \mathrm{d}\boldsymbol{r} \quad (1)$$

where $\boldsymbol{r}$ is the vector $\boldsymbol{r} = x\boldsymbol{a} + y\boldsymbol{b} + z\boldsymbol{c}$ in direct space ($\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are the lattice vectors that define the periodicity of the crystal structure). The integration is over all vectors $\boldsymbol{r}$ in the unit cell. From equation (1), it follows that

$$\rho(\boldsymbol{r}) = (1/V) \sum_{\boldsymbol{H}} |F(\boldsymbol{H})| \exp[i\alpha(\boldsymbol{H}) - 2\pi i \boldsymbol{H} \cdot \boldsymbol{r}] \quad (2)$$

where $V$ denotes the volume of the unit cell, and the summation is over all vectors $\boldsymbol{H}$ with integer coefficients $h$, $k$ and $l$. If the values of both $|F(\boldsymbol{H})|$ and $\alpha(\boldsymbol{H})$ could be measured directly from the experimental diffraction pattern, then $\rho(\boldsymbol{r})$ (*i.e.* the "crystal structure") could be determined directly from equation (2) by summing over the measured diffraction maxima $\boldsymbol{H}$. However, while the values of $|F(\boldsymbol{H})|$ can be obtained experimentally [they are obtained from the measured diffraction intensities $I(\boldsymbol{H})$], the values of $\alpha(\boldsymbol{H})$ cannot be determined directly from the experimental diffraction pattern. The lack of experimental information on the phases $\alpha(\boldsymbol{H})$ constitutes the "phase problem in crystallography". In order to determine a crystal structure from experimental diffraction data (*i.e.* $|F(\boldsymbol{H})|$ data) by making use of equation (2), it is necessary to make use of techniques (*e.g.* direct methods or Patterson methods) that provide ways of estimating the phases $\alpha(\boldsymbol{H})$. The use of such estimated phases $\alpha(\boldsymbol{H})$ together with the experimentally determined values of $|F(\boldsymbol{H})|$ in equation (2) provides a means of determining (at least approximately) the electron density distribution $\rho(\boldsymbol{r})$ and hence the crystal structure.

It is important to emphasize that the reverse procedure – *i.e.* calculating the diffraction pattern for any given structure – is an "automatic" calculation. Thus, the diffraction pattern (*i.e.* the $|F(\boldsymbol{H})|$ data) can always be calculated automatically for *any* crystal structure simply by using the positions of the atoms in the crystal structure in equation (1) [in essence, a form of equation (1) is used in which the electron density $\rho(\boldsymbol{r})$ is approximated by a function that depends on the positions of the atoms in the unit cell]. This approach forms the basis of the *direct-space strategy* for structure solution discussed in Section 3.4. In this strategy, a large number of trial structures are generated by computational procedures, the diffraction pattern for each trial structure is then calculated using equation (1), and these calculated diffraction patterns are then compared with the experimental diffraction pattern in order to assess the degree of "correctness" of each trial structure.

### 2.2  Comparison of experimental and calculated powder diffraction patterns

First, we consider how we define the different features of a powder diffraction pattern. The complete powder diffraction profile (either for an experimental or calculated powder diffraction pattern) may be described in terms of the following components: (i) the peak positions, (ii) the background intensity distribution, (iii) the peak widths, (iv) the peak shapes, and (v) the peak intensities. The peak shape depends on characteristics of both the instrument and the sample, and different peak shape functions are appropriate under different circumstances. The most common peak shape for powder X-ray diffraction is the pseudo-Voigt function, which represents a hybrid of Gaussian and Lorentzian character, although several other types of peak shape function may be applicable under different circumstances. These peak shapes and the types of

analytical functions that are commonly used to describe the $2\theta$-dependence of the peak width are described in detail elsewhere.[8]

There are two basic strategies for comparing the experimental powder diffraction data with the powder diffraction data calculated for structures during the process of structure determination: (1) comparison of the complete powder diffraction profile, and (2) comparison of integrated peak intensities. We now consider each of these approaches in turn.

Comparison of the complete powder diffraction profile uses the whole digitized *experimental* powder diffraction pattern "as measured", and requires a digitized *calculated* powder diffraction pattern for comparison to it. Construction of this calculated powder diffraction pattern for a trial structure requires not only the intensities of all peaks in the powder diffraction pattern (which are readily calculated for any structural model using equation (1)), but also requires information on the peak positions, peak widths, peak shapes and the background intensity distribution in the powder diffraction pattern. In order to carry out a reliable comparison between calculated and experimental powder diffraction patterns, it is essential that the variables that describe these aspects of the calculated powder diffraction pattern faithfully reproduce those in the experimental powder diffraction pattern. Methods for determining the values of the variables that describe these features of the experimental powder diffraction pattern are discussed in Section 3.3. Once the complete digitized powder diffraction pattern for the structural model has been calculated, it can be compared directly with the experimental powder diffraction using an appropriate whole-profile figure-of-merit. The most commonly used figure-of-merit is the weighted profile R-factor $R_{wp}$, which is defined as

$$R_{wp} = 100 \times \sqrt{\frac{\sum_i w_i (y_i - y_{ci})^2}{\sum_i w_i y_i^2}} \qquad (3)$$

where $y_i$ is the intensity of the $i$th point in the digitized experimental powder diffraction pattern, $y_{ci}$ is the intensity of the $i$th point in the calculated powder diffraction pattern, and $w_i$ is a weighting factor for the $i$th point. One notable advantage of using a figure-of-merit of this type during the structure determination process is that it uses the experimental powder diffraction data (*i.e.* the digitized data points $\{y_i\}$) directly "as measured" without further manipulation. Such figures-of-merit are used in several implementations of direct-space techniques for structure solution and in Rietveld refinement.

Comparison of integrated peak intensities, on the other hand, involves analysis of the experimental powder diffraction pattern to extract a set of integrated peak intensities $I(H)$ [and hence $|F(H)|$ values] analogous to those obtained (directly) from a single crystal X-ray diffraction experiment. In view of the problem of peak overlap in the powder diffraction pattern, the extraction of a reliable set of integrated peak intensities from heavily overlapped data is a non-trivial task. The basic techniques for extracting integrated peak intensities are discussed in Section 3.3, and we note that methods have been developed for enhancing the reliability of the intensity extraction process, and for using the extracted intensities in a manner that takes the reliability of the extraction process into consideration (for example, by making use of the covariance matrix[9]). Once a set of integrated peak intensities has been obtained, comparison between experimental and calculated $I(H)$ data can be carried out using figures-of-merit analogous to those employed in the analysis of single crystal diffraction data. However, an intrinsic disadvantage of this approach is that any errors or uncertainties that arise during the process of extracting the integrated peak intensities from the experimental powder diffraction pattern (for example, originating from

ambiguities in handling the peak overlap problem) are inevitably propagated into the structure determination process, and may ultimately limit the reliability of the derived structural information, or may even prohibit successful structure determination. As discussed below, comparison of integrated peak intensities is a central feature of the traditional approach for structure solution. In addition, some implementations of the direct-space approach for structure solution have been based on comparison of integrated peak intensities, presumably with the aim of maximizing speed (such figures-of-merit are faster to calculate than those based on comparison of the complete powder diffraction profile) at the possible expense of reliability.

## 3 An overview of structure determination from powder diffraction data

### 3.1 Stages of the structure determination process

The three stages involved in crystal structure determination from diffraction data are: (i) unit cell determination and space group assignment, (ii) structure solution, and (iii) structure refinement. The aim of *structure solution* is to obtain an initial approximation to the structure, using the unit cell and space group determined in the first stage, but starting with 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 can then be obtained by *structure refinement*. For powder diffraction data, structure refinement can be carried out fairly routinely using the Rietveld profile refinement technique,[8,10] and unit cell determination is carried out using standard indexing procedures.[11–13]

### 3.2 Unit cell determination (indexing)

The first stage of crystal structure determination from powder diffraction data involves determination of the unit cell dimensions $\{a, b, c, \alpha, \beta, \gamma\}$ by analysis of the peak positions in the powder diffraction pattern. This process is usually referred to as "indexing" the powder diffraction pattern. Clearly it is possible to proceed to structure solution and refinement only if the correct unit cell is found at the indexing stage, and difficulties encountered in the reliable indexing of powder diffraction patterns can often be the limiting step in the structure determination process.

The most widely used programs for indexing powder diffraction data (ITO,[11] TREOR,[12] DICVOL[13] and CRYS-FIRE[14]) typically consider the measured positions of the peak maxima for about 20 selected peaks at low diffraction angles. However, the occurrence of peak overlap can lead to problems in indexing, and certain peaks that may be important for correct indexing may be obscured or completely unresolved due to peak overlap. Indeed, the use of low-angle peaks in the indexing process is dictated by the fact that the peak overlap at high diffraction angles is usually so extensive that the data in the high-angle region cannot be used reliably in the indexing process.

### 3.3 Preparing the intensity data for structure solution – pattern decomposition/profile fitting

After the unit cell has been determined from the powder diffraction pattern, the next stage is to prepare the intensity data for space group determination and structure solution using an appropriate "pattern decomposition" or profile fitting technique. The two most commonly applied techniques for this purpose are those developed by Pawley[15] and Le Bail.[16] The aim of these techniques is to fit the complete experimental powder diffraction profile by refinement of variables that describe: **(a)** the peak positions (the variables that determine

the peak positions include the unit cell parameters and the zero-point shift parameter), **(b)** the background intensity distribution, **(c)** the peak widths, **(d)** the peak shapes, and **(e)** the peak intensities. With regard to **(a)**, the unit cell parameters obtained in the indexing procedure are used as the initial values, although the refined values from the pattern decomposition/profile fitting procedure represent a more accurate set of unit cell parameters. It is important to emphasize that no structural model is used in the pattern decomposition/profile fitting procedure (except in so far as the unit cell parameters determined in the indexing stage are used as input variables), and the intensities **(e)** represent a set of intensity variables that are refined to give optimal fit to the experimental powder diffraction pattern without reference to any structural model. Thus, the aim of this process is not to determine the structure but rather to obtain reliable values of the variables that describe different features of the powder diffraction profile (*i.e.* **(a)**–**(e)** above) in preparation for subsequent stages of the structure determination process. It is also important to emphasize that different approaches for structure solution (*i.e.* the traditional and direct-space approaches) make use of different combinations of the variables in **(a)**–**(e)** as input information.

The traditional approach for structure solution requires, as input data, the integrated peak intensities extracted from the experimental powder diffraction pattern – *i.e.* the intensity values **(e)**. In addition, some implementations of the direct-space approach for structure solution are based on comparison of integrated peak intensities, and thus also make use of the intensity values **(e)** as the input intensity data. It is noteworthy that, after extraction of the integrated peak intensities, these approaches do not make any further use of the experimental powder diffraction profile during the structure solution process.

Alternatively, many implementations of the direct-space approach for structure solution are based on comparison of the complete powder diffraction profile, with the experimental and calculated data compared using a whole-profile figure-of-merit such as $R_{wp}$. In this case, the intensity data **(e)** extracted from the powder diffraction pattern in the pattern decomposition/profile fitting procedure are not used in the structure solution process. Instead, the variables **(a)**–**(d)** determined in the pattern decomposition/profile fitting procedure are required (together with intensities calculated for the trial structure) in order to construct the calculated powder diffraction pattern for the trial structure.

Following pattern decomposition/profile fitting, the space group can be assigned by identifying the conditions for systematic absences in the intensity data **(e)**. If the space group cannot be assigned uniquely, structure solution calculations should be carried out separately for each of the plausible space groups. Knowledge of the unit cell volume and space group, together with density considerations, should allow the contents of the asymmetric unit to be established. Information obtained from other experimental techniques (particularly high-resolution solid state NMR) may be particularly helpful in confirming the number of molecules in the asymmetric unit and elucidating other structural aspects that may be used to assist the structure solution process.

## 3.4 Structure solution

The techniques that are currently used for structure solution from powder diffraction data can be subdivided into two categories – the traditional and direct-space approaches.

The *traditional* approach follows a close analogy to the analysis of single crystal diffraction data, in that the intensities $I(H)$ of individual reflections are extracted directly from the powder diffraction pattern and are then used in the types of structure solution calculation (*e.g.* direct methods or Patterson methods) that are used for single crystal diffraction data. As

discussed above, however, peak overlap in the powder diffraction pattern can limit the reliability of the extracted intensities and can therefore lead to difficulties in subsequent attempts to solve the structure using these intensity data. As noted above, such problems may be particularly severe in cases of large unit cells and low symmetry, as encountered for most molecular solids. In spite of these intrinsic difficulties, however, there have been several reported successes in the application of traditional techniques for structure solution of molecular solids from powder diffraction data.

The *direct-space* approach,[6] on the other hand, follows a close analogy to global optimization procedures, which find applications in many areas of science. In the direct-space approach, trial structures are generated in direct space, independently of the experimental powder diffraction data, and the suitability of each trial structure is assessed by direct comparison between the powder diffraction pattern calculated for the trial structure and the experimental powder diffraction pattern (see Section 2.2). This comparison is quantified using an appropriate figure-of-merit. Our implementations of the direct-space strategy have used the weighted powder profile R-factor $R_{wp}$ (the R-factor normally employed in Rietveld refinement), which considers the entire digitized intensity profile point-by-point, rather than the integrated intensities of individual diffraction maxima. Thus, $R_{wp}$ takes peak overlap implicitly into consideration. Furthermore, $R_{wp}$ uses the digitized powder diffraction data directly as measured, without further manipulation of the type required when individual peak intensities $I(H)$ are extracted from the experimental powder diffraction pattern.

The basis of the direct-space strategy for structure solution is to find the trial crystal structure that corresponds to lowest R-factor, and is equivalent to exploring a hypersurface $R(\Gamma)$ to find the global minimum, where $\Gamma$ 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/Simulated Annealing[6,17–32] and Genetic Algorithm[33–45] methods in this field. In addition, grid search[46–50] and differential evolution[51] methods have also been employed. As discussed in Section 2.2, some of these implementations of the direct-space approach (for example[22,26,32]) have used figures-of-merit that are based on the use of extracted peak intensities, rather than a comparison of the complete profile by means of $R_{wp}$. Specific details are discussed in the papers cited.

To date, most reported crystal structure determination of organic molecular solids from powder diffraction data has used the direct-space strategy, although there have also been several reports of successful structure determination of such materials using the traditional approach (see Section 6.6 and examples discussed in refs. 1,5). One feature of the direct-space approach (discussed in more detail in Section 4.1) is that it makes maximal use of information on molecular geometry (such as bond lengths, bond angles and the geometries of well-defined structural units such as phenyl rings) that is already reliably known, independently of the powder diffraction data, prior to commencing the structure solution calculation. The traditional approach for structure solution, on the other hand, does not rely on prior knowledge of the geometry of a well-defined structural fragment. However, for "equal-atom" structures (*e.g.* organic compounds containing no atom heavier than oxygen) in which there are no dominant scatterers, difficulties can be encountered in the application of traditional structure solution techniques.

By way of historical context, we note that the first demonstration[52] of structure solution of a molecular crystal from powder diffraction data was for the *previously known* structure of cimetidine using the traditional approach for structure solution (direct methods) and using data recorded at a

synchrotron X-ray radiation source (see Section 5.1). The first *previously unknown* equal-atom molecular crystal structure to be solved from powder diffraction data was formylurea, again using the traditional approach (direct methods) but from laboratory powder X-ray diffraction data.[53] The first material of unknown crystal structure to be solved using a direct-space strategy was $p\text{-BrC}_6\text{H}_4\text{CH}_2\text{CO}_2\text{H}$[6] using the Monte Carlo method, followed by other examples (including 3-chloro-*trans*-cinnamic acid[18] and 1-methylfluorene[54]) using the same technique.

## 3.5 Structure refinement

In Rietveld refinement of a crystal structure from powder diffraction data,[8,10] the variables that define the structural model and the variables that define the powder diffraction profile (*i.e.* the variables discussed under **(a)**–**(d)** in Section 3.3) are adjusted by least squares methods in order to obtain an optimal fit between the experimental and calculated powder diffraction patterns. In general, the weighted powder profile R-factor $R_{wp}$ (defined in Section 2.2) is used to assess the fit between experimental and calculated powder diffraction patterns (an example of the fit obtained in a typical Rietveld refinement calculation is shown in Fig. 3(b) in Section 6.1). The structural variables that are involved in the Rietveld refinement are analogous to those (*e.g.* atomic coordinates, atomic displacement parameters, site occupancies, *etc.*) that are used in refinement from single crystal diffraction data, and are thus different from the variables that are involved in direct-space structure solution (see Section 4.1). Thus, while bond lengths and bond angles are generally fixed during direct-space structure solution calculations, these constraints are relaxed during Rietveld refinement.

For successful Rietveld refinement, the initial structural model (taken from the structure solution stage) must be a sufficiently good representation of the correct structure. As Rietveld refinement can often suffer from problems of instability, it is generally necessary to use geometric restraints (soft constraints) based on standard molecular geometries to ensure stable convergence of the refinement calculation. As in structure refinement from single crystal X-ray diffraction data, the structural model obtained in the structure solution stage can sometimes be an incomplete representation of the true structure (particularly when structure solution is carried out using the traditional approach). In such cases, difference Fourier techniques can be used in conjunction with Rietveld refinement in order to complete the structural model. Finally, it is relevant to compare the quality of structural information that can be obtained by structure determination from powder *versus* single crystal X-ray diffraction data. In general, the final structural parameters obtained from powder diffraction data are not as accurate or precise as those that could be determined for the same material from single crystal diffraction data (assuming that single crystals were available). Nevertheless, a properly refined crystal structure from powder diffraction data (for example, giving the quality of fit shown in Fig. 3(b)) provides reliable information on the arrangement of atoms and molecules in the crystal structure, and allows an understanding of most aspects of the crystal structure that are of interest to chemists (such as details of the molecular packing arrangement and identification of the intermolecular interactions).

# 4 Methodology for direct-space structure solution from powder diffraction data

## 4.1 Introduction

Within the last 10 years or so, there has been a significant upsurge in research activity relating to the application of techniques to determine the structures of organic molecular solids from powder X-ray diffraction data. This activity has been catalyzed, to a large extent, by the availability of direct-space techniques for structure solution[6], both because the structure determination of molecular materials is particularly well suited to these techniques and because of continual improvements in computer power. In this section, we present a more detailed overview of the essential features of the direct-space strategy for structure solution.

In the direct-space strategy, trial structures are generated in direct space, independently of the experimental powder diffraction data, and the suitability of each trial structure is 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 figure-of-merit. For reasons elaborated in Section 2.2, our implementations of the direct-space strategy have used a figure-of-merit ($R_{wp}$) that is based on consideration of the complete powder diffraction profile, rather than integrated peak intensities.

We now consider the way in which trial structures are defined within the context of direct-space structure solution calculations. In principle, the set of structural variables could be taken to comprise the coordinates of each individual atom within the asymmetric unit, but this approach discards any prior knowledge of molecular geometry and corresponds to the maximal number of structural variables ($3N_a$ variables for $N_a$ atoms in the asymmetric unit). Instead, it is advantageous to make direct use of all information on molecular geometry that is already known reliably beforehand [in studying molecular materials, the identity of the molecule is generally known before starting the structure solution calculation, and if ambiguities remain on issues concerning the atomic connectivity (*e.g.* tautomeric form), other techniques such as solid state NMR spectroscopy may be useful to resolve these ambiguities before starting the structure solution calculation (for an example, see ref. 19)]. Thus, it is common practice to fix bond lengths and bond angles at standard values during the direct-space structure solution calculation, and to fix the geometries of well defined structural units (such as phenyl rings). In general, the only aspects of intramolecular geometry that are not known reliably beforehand are the values of some (or all) of the torsion angles that define the molecular conformation. Under these circumstances, each trial structure in a direct-space structure solution calculation is defined by a set ($\Gamma$) of structural variables that represent, for each molecule in the asymmetric unit, the position of the molecule in the unit cell (defined by the coordinates $\{x, y, z\}$ of the centre of mass or a selected atom), the orientation of the molecule in the unit cell (defined by rotation angles $\{\theta, \phi, \psi\}$), and the unknown torsion angles $\{\tau_1, \tau_2, ..., \tau_n\}$. Thus, in general, there are $6 + n$ variables, $\Gamma = \{x, y, z, \theta, \phi, \psi, \tau_1, \tau_2, ..., \tau_n\}$, for each molecule in the asymmetric unit.

For structure solution by the direct-space approach, the complexity of the structure solution problem is dictated to a large extent by the total number of structural variables to be determined, and thus the greatest challenge in the application of direct-space techniques arises when the number of structural variables is large. This situation occurs when there is considerable molecular flexibility (*i.e.* when the molecule has a large number of variable torsion angles) and/or when there are several independent molecules in the asymmetric unit.

As discussed above, the basis of the direct-space strategy for structure solution is to find the trial crystal structure that corresponds to optimal agreement (*i.e.* lowest R-factor) between calculated and experimental powder diffraction patterns, and is equivalent to exploring a hypersurface $R(\Gamma)$ to find the global minimum. Of the different global optimization techniques that have been used in direct-space structure solution from powder diffraction data, the two most widely

applied have been the Monte Carlo/Simulated Annealing and Genetic Algorithm techniques. We now discuss the essential basis of each of these search algorithms.

## 4.2 The Monte Carlo/Simulated Annealing technique

In the Monte Carlo/Simulated Annealing technique, a sequence of structures (denoted $\Gamma_i$ for $i = 1, 2, ..., N$) is generated for consideration as potential structure solutions. Each structure is derived from the previous structure by a small random displacement of the molecule(s) within the unit cell. The procedure for generating structure $\Gamma_{j+1}$ from structure $\Gamma_j$ is summarized as follows.

Starting from structure $\Gamma_j$, a trial structure $\Gamma_{j,\text{trial}}$ is generated by making small random displacements to each of the structural variables in $\Gamma_j$. The agreement between the powder diffraction pattern calculated for the trial structure and the experimental powder diffraction pattern is then assessed by calculating an appropriate figure-of-merit, such as $R_{\text{wp}}$. The trial structure is then accepted or rejected by considering the difference $[Z = R(\Gamma_{j,\text{trial}}) - R(\Gamma_j)]$ between the values of R-factor for structures $\Gamma_{j,\text{trial}}$ and $\Gamma_j$ and invoking the Metropolis importance sampling algorithm. Thus, if $Z \leqslant 0$, 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 $\Gamma_{j+1}$ is taken to be the same as $\Gamma_{j,\text{trial}}$. If the trial structure is rejected, structure $\Gamma_{j+1}$ is taken to be the same as $\Gamma_j$. The parameter $S$ may either be fixed or varied in a controlled manner during the calculation. The higher the value of $S$, the higher the probability of accepting trial structures for which $Z > 0$.

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

The essential distinction between Monte Carlo and Simulated Annealing techniques is 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 a well-defined annealing schedule or temperature reduction procedure. Different implementations of Simulated Annealing methods in this field employ a range of different ways of handling the annealing procedure. We note that a variant of the Simulated Annealing technique – parallel tempering – allows the simulated annealing search to benefit from the implementation of parallel computing concepts.

## 4.3 The Genetic Algorithm technique

The Genetic Algorithm (GA) technique is based on the principles of evolution and involves familiar evolutionary operations such as mating, mutation and natural selection. An important feature of the GA technique is that it operates in a parallel manner, with many different regions of the $R(\Gamma)$ hypersurface investigated simultaneously. Furthermore, information concerning these different regions of the $R(\Gamma)$ hypersurface is passed actively between different members of the population by the mating procedure. Clearly, the intrinsic parallel nature of the GA technique confers additional efficiency in searching the $R(\Gamma)$ hypersurface.

Our GA technique[33–44] for structure solution from powder diffraction data is implemented in the program EAGER.[55] The GA structure solution strategy investigates the evolution of a population of trial structures, with each member of the population defined by a set of variables $\Gamma$, as defined in Section 4.1. As each member of the population is uniquely characterized by the values of these variables, the set $\Gamma$ can be regarded to define its "genetic code". The initial population $P_o$ 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 (population $P_{j+1}$). The number $N_p$ of structures in the population is constant for all generations, and $N_m$ mating 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-factor (lower R-factor represents higher fitness), and it is advantageous to define fitness as an appropriate decreasing function of R-factor. A schematic flow chart describing the procedure for combining the operations of mating, mutation and natural selection in the evolution of the population within our GA for structure solution is shown in Fig. 1.
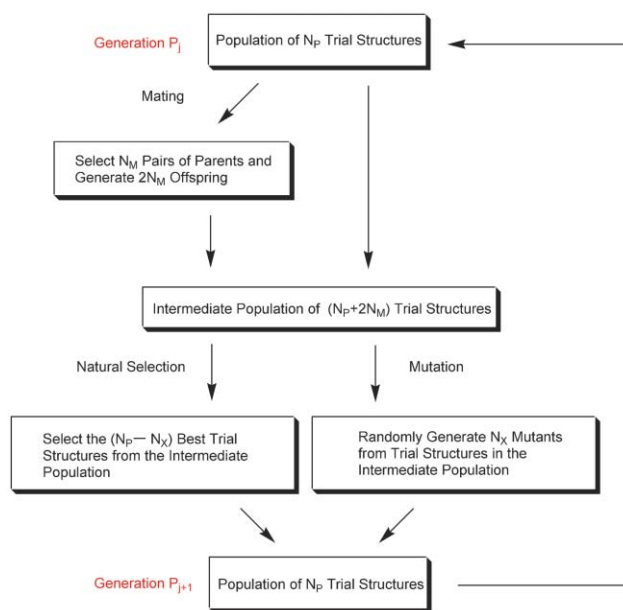


**Fig. 1** Flow chart representing the procedure for evolution of the population from one generation (population $P_j$) to the next generation (population $P_{j+1}$) in the GA technique for powder structure solution.

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 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 a

larger number of variables, more complex rules may be adopted for the mating procedure.

It is important to recognise that the mating operation generates new structures by redistributing the existing genetic information in different ways, but does not actually create any new values of the individual genetic variables. New values of the genetic variables are instead introduced into the population by 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 codes to create mutant structures. The changes that are made to selected variables in generating the mutants may either be new random values (static mutation) or small random displacements from the existing values (dynamic mutation). The original structures from which the mutants are derived are still retained within the population.

In the *natural selection* procedure, only the structures of highest fitness (lowest R-factor) are allowed to pass from one generation to the next generation in the GA calculation. After the population has evolved for a sufficiently large number of generations, the structure with lowest R-factor should be close to the correct structure.

Other features of our GA technique[55] include an implementation of Lamarckian evolution[37] (in which each new structure generated in the GA calculation is subjected to local minimization of $R_{wp}$ with respect to the variables in $\Gamma$), a parallel GA[41] (involving the separate evolution of different sub-populations, with migration of structures between sub-populations allowed to occur in a controlled manner), and an algorithm for rapid evaluation of $R_{wp}$ within the context of direct-space structure solution.[56]

# 5 Experimental considerations

## 5.1 Synchrotron *versus* laboratory powder X-ray diffraction data

We now consider the relative merits of using synchrotron X-ray powder diffraction data *versus* conventional laboratory powder X-ray diffraction data in the field of structure determination from powder diffraction data, recognizing that the use of synchrotron radiation generally gives rise to powder diffraction data of higher resolution and improved signal/noise ratio. With high resolution, problems due to peak overlap can be alleviated, at least to some extent, allowing more reliability in determining accurate peak positions (which is advantageous in unit cell determination) and more reliability in extracting the intensities of individual diffraction maxima from the powder diffraction pattern. In this regard, synchrotron radiation can be advantageous when traditional techniques (or those direct-space techniques that use figures-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 structure solution techniques that employ a figure-of-merit based on a profile R-factor (such as $R_{wp}$), the important requirement is not high resolution itself, 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, and many examples demonstrate that the use of a good quality, well optimized laboratory powder X-ray diffractometer is usually perfectly adequate for research in this field. We note that, within the context of Rietveld refinement, the use of synchrotron data generally leads to structural results of greater accuracy, as a consequence of the fact that the data in the high

$2\theta$ region of the powder diffraction pattern are usually of higher quality for synchrotron data than laboratory data.

## 5.2 Preferred orientation

In general, structure solution from powder diffraction data has a good chance of success only if the experimental powder diffraction pattern contains reliable information on the intrinsic relative intensities of the diffraction maxima, which requires that there is no "preferred orientation" in the powder sample. Preferred orientation arises when the crystallites in the powder are oriented preferentially in certain directions, and can be particularly severe when the crystal morphology is strongly anisotropic (*e.g.* long needles or flat plates). When there is a non-random distribution of crystallite orientations in the sample, the measured relative peak intensities differ from the intrinsic relative diffraction intensities, limiting the prospects for determining reliable structural information from the powder diffraction pattern. In order to circumvent this difficulty, it is recommended that appropriate procedures[57] are carried out to screen powder samples for preferred orientation, and to take steps to ensure that the sample is free of preferred orientation before recording high quality powder diffraction data for use in structure determination calculations. If preferred orientation is detected, a variety of experimental approaches may be used to alleviate the effects of preferred orientation, such as using a capillary or end-loading sample holder, mixing the sample with an amorphous material, preparing the sample by spray-drying, or using an appropriate grinding procedure to induce a crystal morphology that is as isotropic as possible.

## 5.3 Phase purity

Another issue that may potentially limit the successful application of techniques for structure determination from powder diffraction data concerns the phase purity of the powder sample. Thus, the discussion in this article so far has been under the implicit assumption that the powder sample comprises only one crystalline phase. If the powder sample contains a second crystalline phase (*e.g.* an impurity phase or a second polymorph of the material of interest) and is not known to contain this second phase, then the structure determination process will almost certainly fail at the indexing stage (it will be impossible to find a single unit cell that predicts all the peak positions in the powder diffraction pattern). However, if the existence and identity of an impurity or second phase are known beforehand, the peaks due to this phase may be recognized and handled in an appropriate manner that allows structure determination of the main phase of interest to proceed successfully. Clearly it is advantageous to use other experimental techniques (such as solid state NMR spectroscopy) to provide an independent assessment of the phase purity of a powder sample *before* embarking upon the process of structure determination from powder diffraction data, and to optimize the sample preparation procedures such that the powder diffraction data are recorded for a phase-pure sample. Nevertheless, in favourable cases, careful inspection of the powder diffraction data alone may enable progress to be made, as illustrated by the structure determination of cyclopentadienyl rubidium[58] from the powder diffraction pattern of a sample comprising a mixture of two polymorphs. In this case, initial attempts to index the powder diffraction pattern failed due to the presence of the two phases, but closer inspection revealed that the data could be sub-divided into two sets of peaks with appreciably different linewidths, which were attributed to the two polymorphs. By subdividing the experimental data in this way, indexing and structure determination were carried out successfully for both polymorphs.

# 6 Examples of structure determination from powder diffraction data

In this section we present a selection of examples that highlight the application of the techniques discussed above for structure determination from powder X-ray diffraction data. Although we focus on examples of structure determination of molecular materials using direct-space techniques for structure solution, the examples also include structure determination of a framework material (Section 6.7) and illustrations of structure determination using the traditional approach (Sections 6.5 and 6.6).

## 6.1 Structural rationalization of oligopeptides

Knowledge of the conformational properties and interactions in oligopeptides can yield important insights concerning the structural properties of polypeptide sequences in proteins. In many cases, however, the target materials cannot be prepared in the form of single crystals appropriate for single crystal X-ray diffraction studies, and in such cases structure determination from powder diffraction data represents the only viable route towards structural understanding and rationalization[59]. We now describe three examples of oligopeptide structures determined from powder X-ray diffraction data using our GA technique for structure solution.

The first example concerns Phe-Gly-Gly-Phe,[38] for which the GA structure solution calculation involved 11 variable torsion angles, with the peptide groups constrained to be planar units and the O–C–N–H torsion angle fixed at 180°. The structure was solved within 50 generations for a population size of 50 trial structures. The structure (space group $P4_1$) comprises ribbons that run along the $c$-axis, with adjacent molecules in these ribbons interacting through three N–H···O hydrogen bonds in a manner directly analogous to an anti-parallel β-sheet. Intermolecular N–H···O hydrogen bonds involving the end-groups of the oligopeptide chains give rise to two intertwined helical chains running along the $4_1$ screw axis.

Next we consider structure determination of the peptides Piv-$^L$Pro-Gly-NHMe and Piv-$^L$Pro-γ-Abu-NHMe, with particular interest in the potential for these molecules to form β-turn
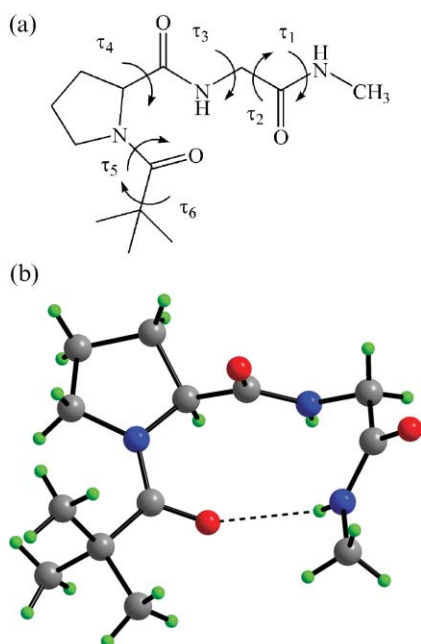
conformations (these structural features allow polypeptide chain reversals in proteins).

In the GA structure solution calculation[39] for Piv-$^L$Pro-Gly-NHMe (Fig. 2(a)), the genetic code comprised 9 variables $\{\theta, \phi, \psi, \tau_1, \tau_2, ..., \tau_6\}$ (in space group $P1$, the position $\{x, y, z\}$ of the molecule is fixed arbitrarily). The variable torsion angles were allowed to take any value, except $\tau_5$ which was allowed to take only the values 0° or 180°; the O–C–N–H torsion angle between $\tau_3$ and $\tau_4$ was fixed at 180°. The structure was solved within 50 generations for a population size of 100 trial structures. Fig. 2(b) shows the final refined structure of Piv-$^L$Pro-Gly-NHMe, in which it is clear that the molecule adopts a Type II β-turn conformation stabilized by an intramolecular $4 \rightarrow 1$ hydrogen bond between the C=O group of the Piv residue and the methylamide N–H group (N···O, 2.99 Å; N···O–C, 140.6°).

In the GA structure solution calculation[40] for Piv-$^L$Pro-γ-Abu-NHMe (Fig. 3(a)), which differs from Piv-$^L$Pro-Gly-NHMe
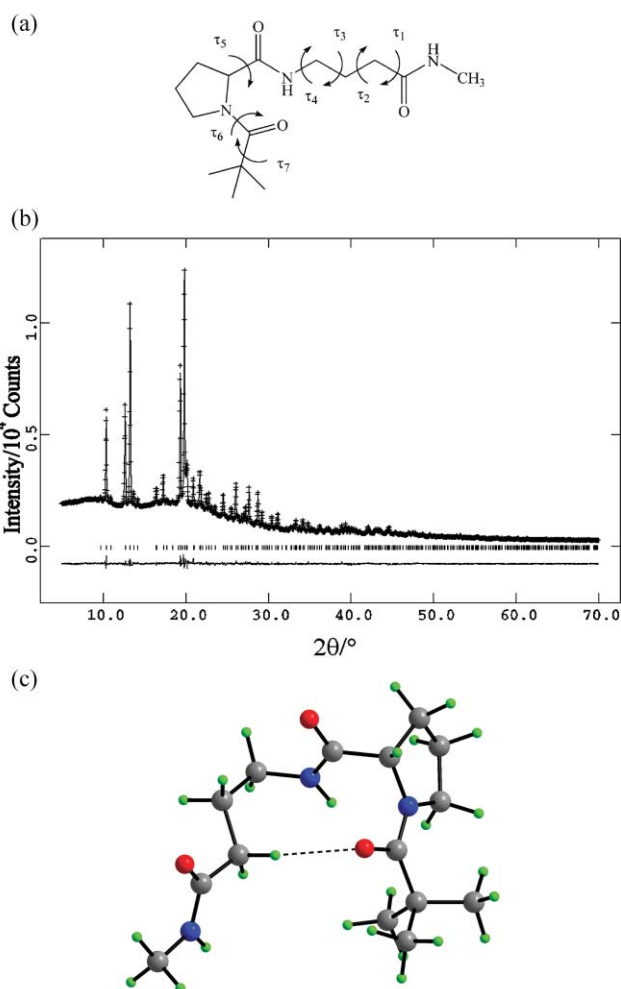


Fig. 3 (a) Molecular structure of Piv-$^L$Pro-γ-Abu-NHMe indicating the torsion angles considered as variables in the GA structure solution calculation. (b) Final Rietveld refinement for Piv-$^L$Pro-γ-Abu-NHMe, showing the experimental powder X-ray diffraction pattern (+marks), calculated powder X-ray diffraction pattern (solid line) and difference between experimental and calculated powder X-ray diffraction patterns (lower line). The tick marks indicate reflection positions. (c) Conformation of Piv-$^L$Pro-γ-Abu-NHMe in the final refined crystal structure, showing the formation of an intramolecular C–H···O=C hydrogen bond.



Fig. 2 (a) Molecular structure of Piv-$^L$Pro-Gly-NHMe indicating the torsion angles considered as variables in the GA structure solution calculation. (b) Conformation of Piv-$^L$Pro-Gly-NHMe in the final refined crystal structure, showing the formation of a type II β-turn.

by the introduction of two additional CH$_2$ units within the peptide chain, the genetic code comprised 13 variables (7 variable torsion angles). The torsion angle of the peptide bond of the $^L$Pro residue was restricted to be 0° or 180°, and the

other two amide linkages CO–NH were maintained as planar units with O–C–N–H torsion angle fixed at 180°. All other torsion angles were considered as variables. The structure was solved within 20 generations for a population size of 100 trial structures. The final fit to the experimental powder X-ray diffraction pattern following Rietveld refinement is shown in Fig. 3(b). In the crystal structure (Fig. 3(c)), Piv-$^L$Pro-$\gamma$-Abu-NHMe adopts a folded conformation, with a short C–H···O interaction [H···O, 2.51 Å; C···O, 3.59 Å; C–H···O, 172°; hydrogen atom position normalized according to standard geometries from neutron diffraction] observed between one of the methylene hydrogen atoms of $\gamma$-Abu and the C=O group of the Piv residue. This C–H···O interaction defines an intramolecular cyclic 10-atom motif, similar to that observed in the classical $\beta$-turn (which involves an intramolecular N–H···O hydrogen bond), as discussed above for Piv-$^L$Pro-Gly-NHMe.

## 6.2 Structure determination of a multi-component co-crystal

In recent years, there has been much interest in the structural rationalization of materials containing aryl and perfluoroaryl groups, initiated by the discovery that the 1 : 1 co-crystal formed between benzene and hexafluorobenzene contains stacks of alternating benzene and hexafluorobenzene molecules. Subsequently, the alternating stack motif of aryl and perfluoroaryl rings has been recognized and exploited as a structural element that may be used as the basis for crystal design strategies. Recently, we have been interested in the formation and structural properties of co-crystals containing molecules of the types $C_6H_5X$ and $C_6F_5Y$, in which X and Y are hydrogen bond donor and/or acceptor groups. As part of our work in this area, we recently[43] studied the 1 : 1 co-crystal containing benzoic acid ($C_6H_5CO_2H$; **BA**) and pentafluorobenzoic acid ($C_6F_5CO_2H$; **PFBA**), with structure determination carried out from powder X-ray diffraction data using the parallel GA technique for structure solution. Interestingly, the high-resolution solid state $^{13}C$ NMR spectrum of the co-crystal indicates that there are two crystallographically inequivalent molecules of **BA** and two crystallographically inequivalent molecules of **PFBA** in the asymmetric unit. In space group $Cc$, the values of $x$ and $z$ for one molecule can be fixed arbitrarily, and thus a total of 26 structural variables are required in the GA calculation (with one variable torsion angle for each of the four independent molecules in the asymmetric unit). The structure was solved within 170 generations for four sub-populations each containing 50 trial structures. The structure of the co-crystal (Fig. 4) is
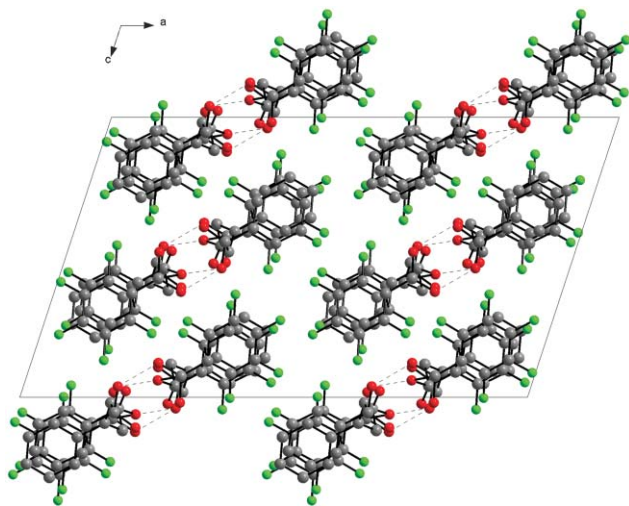
found to comprise stacks of alternating **BA** and **PFBA** molecules, with two crystallographically independent types of stack. Molecules in the two crystallographically independent types of stack interact with each other by hydrogen bonding of their carboxylic acid groups. All inter-stack hydrogen bonding interactions of this type are heteromolecular in nature, involving a **BA** molecule in one stack and a **PFBA** molecule in the other stack. The two independent **BA** molecules and the two independent **PFBA** molecules differ appreciably in molecular conformation, concerning the torsion angle between the carboxylic acid and aryl units. The fact that one molecule of each type has a comparatively large value for this torsion angle can be attributed to the avoidance of repulsive F···O interactions in the structure.

## 6.3 Structure determination of a new co-crystal phase produced by a solid state grinding procedure

In addition to the preparation of molecular co-crystals by conventional solution phase crystallization, as done for the **BA/PFBA** co-crystal described above, many molecular co-crystals can be prepared only by grinding together the "pure" solid phases of the constituent molecules. In many cases, single phase co-crystals are obtained following sufficient grinding. Materials prepared by the solid state grinding procedure are virtually always microcrystalline powders, and are therefore not amenable to structural characterization by single crystal X-ray diffraction. We have demonstrated[42] the use of powder X-ray diffraction to determine the structure of a co-crystal material prepared by the solid state grinding route. The material contains three molecular components – racemic bis-$\beta$-naphthol (**BN**), benzoquinone (**BQ**) and anthracene (**AN**). Grinding a physical mixture of the pure crystalline phases of **BN**, **BQ** and **AN** produces a polycrystalline material with reddish purple colour (crystallization from solution, on the other hand, gives a different co-crystal with bluish black colour). Structure solution was carried out using our parallel GA technique in space group $C2/c$. The contents of the asymmetric unit (confirmed on the basis of high-resolution solid state $^{13}C$ NMR data) comprise one **BN** molecule, one **BQ** molecule and half of an **AN** molecule, with the **AN** molecule residing on a two-fold rotation axis. Thus, the structure solution calculation involved a total of 17 structural variables. The structure was solved within 50 generations for two sub-populations each containing 100 trial structures. The structure (Fig. 5) is rationalized in terms of three different interaction motifs: edge-to-face interactions between **BQ** (edge) and **AN** (face) molecules,
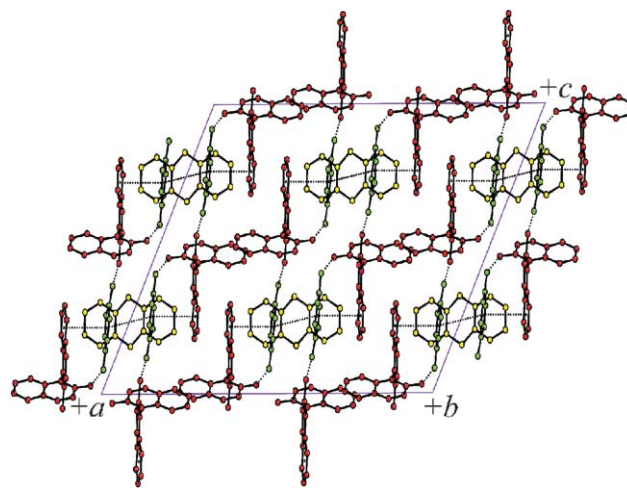
**Fig. 4** Structure of the **BA/PFBA** co-crystal viewed along the stacking axis. Inter-stack hydrogen bonding involving the carboxylic acid groups is clearly evident.

**Fig. 5** Crystal structure of the **BN/BQ/AN** co-crystal material (**BN** – red; **BQ** – green; **AN** – yellow). Dotted lines indicate $\pi$-stacking interactions and hydrogen bonded chains.

face-to-face interactions between **BQ** and **BN** molecules, and chains of O–H···O hydrogen bonds involving **BN** and **BQ** molecules. Knowledge of the structural properties of this material creates the possibility to understand its interesting colouristic properties. Hitherto, structural characterization of co-crystal materials prepared by grinding procedures has been limited by the fact that the preparation procedure intrinsically leads to polycrystalline powders. Clearly, as demonstrated in this case, structure determination from powder diffraction data has an important role to play in the structural characterization of new co-crystal phases prepared by such procedures.

## 6.4 Structure determination of intermediates in a solid state reaction

An interesting example of the application of structure solution from powder diffraction data concerns structural characterization of reactants and products of chemical reactions, including structure determination of intermediate phases that may be difficult to obtain as pure phases. The decomposition reaction of ammonia trimethylalane ($Me_3AlNH_2$) to give aluminium nitride (AlN) *via* the intermediates $(Me_2AlNH_2)_x$ and $(MeAlNH)_y$ was first discovered over fifty years ago.[60] The structure of the intermediate $(Me_2AlNH_2)_x$ produced in this reaction has been determined directly from powder diffraction data[61] using a simulated annealing technique. Interestingly, this structure is different from that of the trimer $(Me_2AlNH_2)_3$ prepared by a different route, and solved from single crystal X-ray diffraction data.[62] The unit cell of the $(Me_2AlNH_2)_x$ material obtained from indexing the powder diffraction pattern was consistent with the asymmetric unit comprising a trimeric unit, which complicated the structure solution process, as this trimer may exist in various boat and twist-boat conformations. Three independent molecules were joined together to form trimers of various conformations, with the trimers treated as rigid bodies in the structure solution calculations. The best structure solution was obtained for the boat conformation (Fig. 6), in contrast to the structure determined from single crystal diffraction data, in which the trimer adopts a twist-boat conformation.
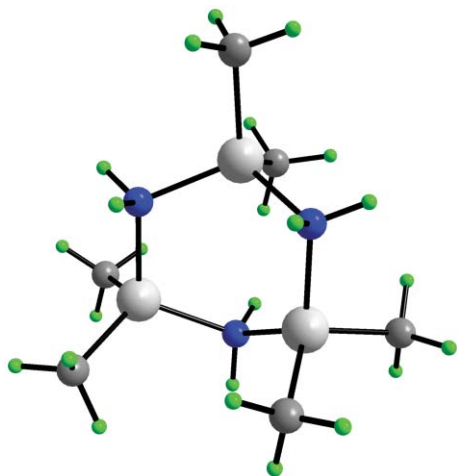


**Fig. 6** Trimeric association of $Me_2AlNH_2$ units in the polymorph of $(Me_2AlNH_2)_3$ prepared by the decomposition reaction of $Me_3AlNH_2$.

## 6.5 Structure determination of $VOCl_2(H_2O)(C_6H_8O_2)_2$

The crystal structures of other metal-organic compounds have also been solved using powder diffraction data. Vanadium complexes are a popular area of research because of their use in medicinal chemistry, and many vanadyl compounds have not

been able to be prepared as single crystals appropriate for structure determination using single crystal X-ray diffraction techniques. One such material is $VOCl_2(H_2O)(C_6H_8O_2)_2$, for which structure determination has been carried out from powder X-ray diffraction data.[63] Structure solution was carried out by the traditional approach. Following extraction of individual peak intensities from the powder diffraction data, the structure was solved by direct methods and expanded using Fourier techniques. The crystal structure reveals valuable information about the geometry of the complex. Instead of the $C_6H_8O_2$ ligand acting in a bidentate manner to form a chelate with the vanadium centre, only one oxygen atom of each ligand is bonded to the vanadium (Fig. 7). An interesting feature of
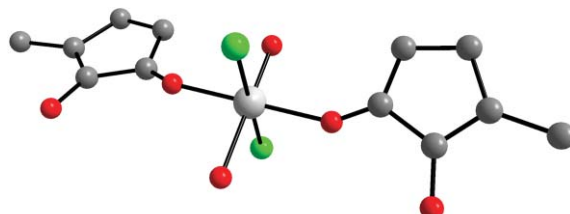


**Fig. 7** The conformation of $VOCl_2(H_2O)(C_6H_8O_2)_2$ in the crystal structure determined from powder diffraction data.

the structure concerns the shape of the cyclopentene ring, which is interpreted in terms of the existence of both enol and keto forms, and represents a disordered crystal structure.

## 6.6 Exploiting anisotropic thermal expansion in structure solution

As elaborated above, structure solution from powder diffraction data by the traditional approach relies upon the availability of accurate values of the extracted intensities of individual reflections in the powder diffraction pattern. The difficulty in obtaining reliable intensities in this way was one of the primary factors that motivated the development of direct space strategies, which, as discussed above, do not require the use of extracted peak intensities. However, it is important to emphasize that structure solution by the traditional approach can still be successful when the powder diffraction pattern suffers from substantial peak overlap, as illustrated by the case of 9-ethylbicyclo[3.3.1]nonan-9-ol.[64] Powder X-ray diffraction patterns were recorded for this material at several different temperatures. At each temperature, the peaks in the powder diffraction pattern are substantially overlapped, and the peak intensities extracted from the data recorded at any one of these temperatures did not allow the structure to be solved. To overcome this problem, the authors took advantage of anisotropic thermal expansion[65,66] in order to improve the intensity extraction process. Due to anisotropic thermal expansion, different peaks in the powder diffraction pattern shift to a different extent as temperature is varied, and hence the nature of the peak overlap changes as a function of temperature. Thus, by carrying out a combined analysis of the data recorded at all the temperatures studied, a more reliable extraction of the integrated peak intensities can be obtained. Using the accurate set of integrated peak intensities obtained from this multi-pattern peak extraction process, successful structure solution of 9-ethylbicyclo[3.3.1]nonan-9-ol was achieved using direct methods (together with analysis of difference Fourier maps). In this crystal structure, there are four independent molecules in the asymmetric unit, which assemble into a tetrameric unit that is held together by O–H···O hydrogen bonds (Fig. 8). The observation of strong hydrogen bonds in the crystal structure may provide an explanation for the ordered nature of this material, in contrast to the disordered structures that are found
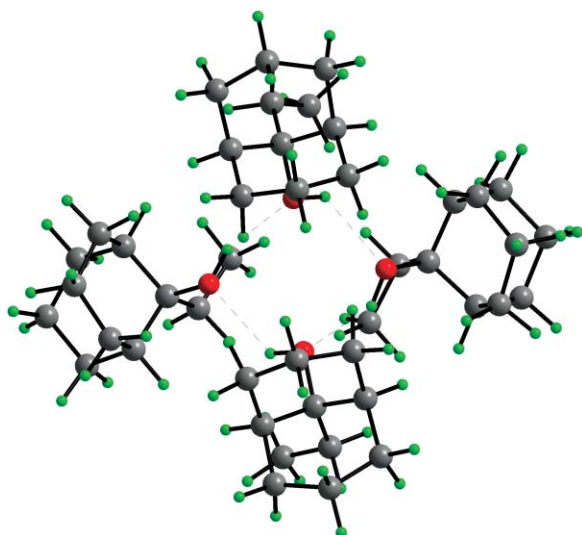
**Fig. 8** The tetrameric association of 9-ethylbicyclo[3.3.1]nonan-9-ol in its crystal structure.

for several analogues that do not interact through strong hydrogen bonds.

### 6.7 Structure determination of a novel aluminium methylphosphonate

In general, traditional techniques for structure solution from powder diffraction data have been applied widely for structure determination of inorganic materials, but recent reports demonstrate that direct-space strategies can provide a viable alternative for such materials. Thus, structure solution of the γ-phase of $Al_2(CH_3PO_3)_3$ proved to be difficult by direct methods, and a direct-space approach (employing a simulated annealing algorithm) was used instead.[67] For this material, thermal and spectroscopic data indicated an anhydrous asymmetric unit containing two inequivalent aluminium atoms and three inequivalent phosphonate groups, representing a total of five structural fragments in the simulated annealing structure solution calculation. The packing was found to be lamellar, with the methyl groups protruding outwards from each sheet (Fig. 9). The atomic positions obtained from lattice energy
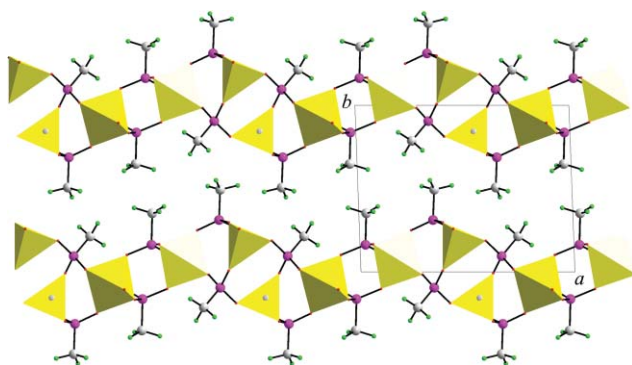


**Fig. 9** The crystal structure of the γ-phase of $Al_2(CH_3PO_3)_3$. Note that the methyl groups protrude outwards from each sheet.

minimization using semi-empirical methods were found to be in good agreement with those in the structure determined from the powder diffraction data.

### 6.8 Rationalization of a solid state reaction after 140 years

It is well known[68,69] that to understand the chemical reactivity of solids relies upon knowledge of the structural properties of

the solid. Determination of the crystal structures of reactive materials is therefore a pre-requisite for understanding the chemical transformations that occur within them. It has been known since the 1850s[70,71] that solid sodium chloroacetate undergoes a polymerization reaction at high temperature to produce polyglycolide and sodium chloride.

$$n\ ClCH_2COONa \rightarrow n\ NaCl + [CH_2COO]_n$$
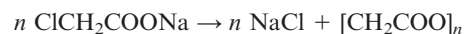
However, an understanding of this reaction and its mechanism could not be established, as sodium chloroacetate is microcrystalline and the structure of this material could not be determined using single crystal X-ray diffraction. However, with the advent of the direct-space approach for structure determination from powder diffraction data, the structure of sodium chloroacetate has now been determined,[72] using the Monte Carlo technique for structure solution. The crystal structure contains rows of chloroacetate anions (Fig. 10) within
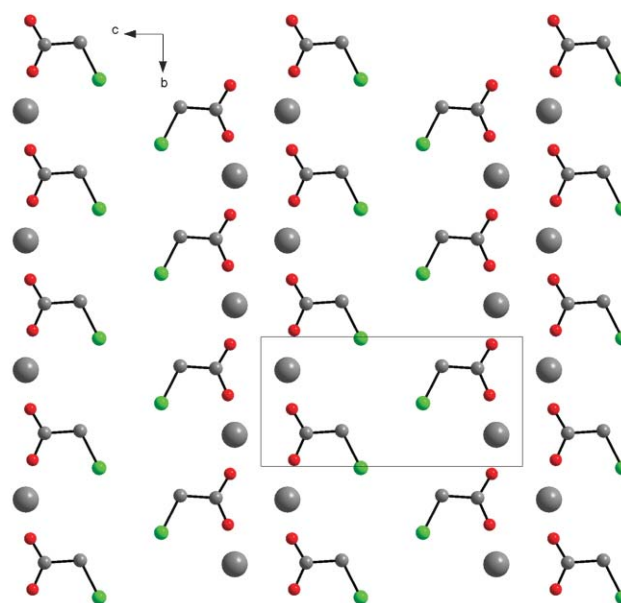


**Fig. 10** A section of the crystal structure of sodium chloroacetate (hydrogen atoms not shown). The polymerization reaction is propagated within rows of chloroacetate anions along the *b*-axis.

which one of the two oxygen atoms of the carboxylate group is ideally positioned to attack the α-carbon atom of a neighbouring chloroacetate anion to expel a $Cl^-$ anion; propagation of this attack along the row of chloroacetate anions results in polymerization to produce polyglycolide. Thus, from knowledge of the crystal structure, the production of polyglycolide may be rationalized directly on the basis of a topochemical reaction pathway. The crystal structures of various lithium halogenoacetates have also been determined from powder diffraction data,[73] and provide a basis for understanding the chemical reactions in these materials.

### 6.9 Rationalizing the structural properties of pharmaceutical materials

For pharmaceutical materials, which are often administered in the form of polycrystalline powders, knowledge of the crystal structure is crucial for fully understanding and optimizing a range of properties, including solubility, bioavailability, and the conditions for handling and administration. In addition, the quest to produce and structurally characterize all accessible polymorphs of a given drug substance has become an area of intense activity within the pharmaceuticals industry (motivated

in part by patenting and registration issues). In many cases, powder diffraction provides the only possible route for determining the crystal structures of pharmaceutical materials, and there has been a significant upsurge in recent years in the use of powder diffraction within pharmaceutical sciences (see a recent review for more details[74]). Nevertheless, the structures of pharmaceutical materials can sometimes be difficult to solve from powder diffraction data because the processes to which they are subjected during manufacture can lead to significant line-broadening in the powder diffraction pattern, or give rise to crystal morphologies that enhance the effects of preferred orientation.

The crystal structures of a number of substances of pharmaceutical interest have been determined from powder diffraction data, of which we highlight the case of bupivacaine,[75] a local anaesthetic that acts upon sodium ion channel receptors in the cell membrane. This material suffers severely from the effects of preferred orientation, and reliable powder X-ray diffraction data were recorded only after screening samples for preferred orientation using the method discussed in ref. 57. The crystal structure was solved using the GA technique, with the bupivacaine molecule in the 1,4-di-equatorial conformation. In the crystal structure, bupivacaine forms intermolecular N–H⋯O hydrogen bonds between the carbonyl oxygen atom of one molecule and the amine N–H group of a neighbouring molecule, resulting in a one-dimensional hydrogen bonded network involving columns of molecules along the c-axis (Fig. 11). This structure of the free base form of bupivacaine provides interesting contrasts to those of its two hydrochloride salts.[76,77]
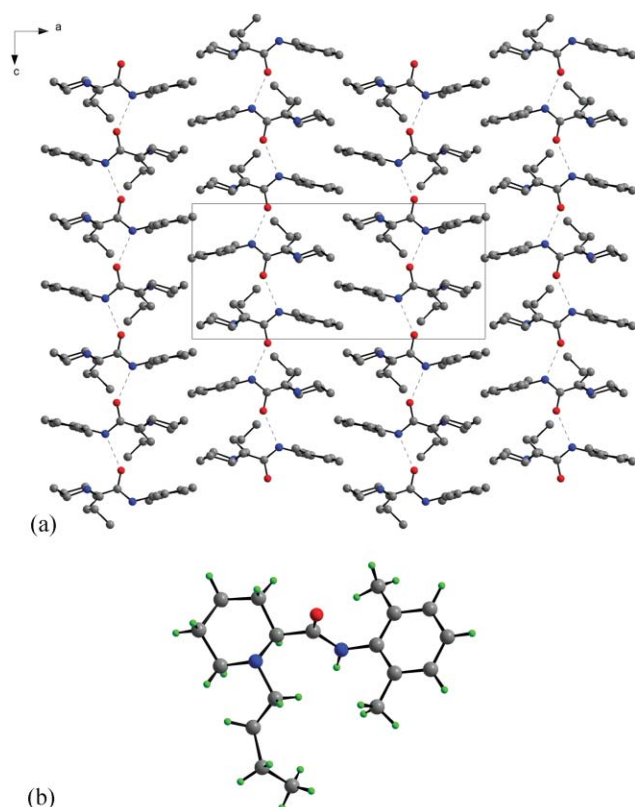


**Fig. 11** (a) The crystal structure of bupivacaine, which contains columns of molecules linked by N–H⋯O hydrogen bonds parallel to the c-axis. (b) The conformation of bupivacaine in the crystal structure.

Several other examples of structure determination of materials of pharmaceutical relevance from powder X-ray diffraction data have been reported in recent years, mainly employing direct-space techniques for structure solution. Examples include the structure determination of a new polymorph of fluticasone propionate,[36] 2-{[4-(4-fluorophenoxy)phenyl]-methylene}-hydrazinecarboxamide,[78] promazine hydrochloride and thio-thixene,[22] four polymorphic forms of fananserine,[79] tetracaine hydrochloride,[80] polymorphs of telmisartan[81] and form II of enalapril maleate.[82] In addition, the traditional approach for structure solution has also been used in a few cases, including structure determination of chlorothiazide[66] and polymorph V of sulfathiazole.[83]

## 7 Concluding remarks

As recently as 15 years ago, no previously unknown molecular crystal structure had been determined directly from powder X-ray diffraction data. In the intervening period of time, advances in methodology (particularly the development of the direct-space strategy for structure solution) have allowed the field to progress to the current situation in which molecular crystal structures of moderate complexity can now be determined directly from powder X-ray diffraction data, as illustrated by the examples presented in this article. Nevertheless, while some structures can be determined straightforwardly using these techniques, other examples can still present challenges, and the techniques for structure determination from powder diffraction data still lag behind those for structure determination from single crystal diffraction data in terms of the straightforward and routine manner in which they can be applied. Importantly, structure solution from powder diffraction data is not a black-box technique, and considerable care must be taken to ensure the correctness of the derived structural model. In spite of the advances made in recent years, there remains considerable scope for the future development and optimization of methodology for structure determination from powder diffraction data, both through the development of new and optimized procedures for searching $R(\Gamma)$ hypersurfaces in direct-space structure solution, and in terms of the development of new ways of defining the hypersurface such that global optimization may be achieved more efficiently. In addition, there is much scope for further advancements in the capabilities of indexing methodologies, particularly by tackling the indexing problem using a range of modern computational algorithms.

While the direct-space strategy for powder structure solution is particularly appropriate in the case of molecular solids, the opportunities for applying this strategy extend far beyond the molecular solid state, and the future application of direct-space techniques promises to reveal new insights into structural properties of a wide range of materials for which structural characterization by single crystal diffraction techniques is not possible.

## References

1  K. D. M. Harris and M. Tremayne, *Chem. Mater.*, 1996, **8**, 2554.
2  J. I. Langford and D. Louër, *Rep. Progr. Phys.*, 1996, **59**, 131.
3  D. M. Poojary and A. Clearfield, *Acc. Chem. Res.*, 1997, **30**, 414.
4  A. Meden, *Croat. Chem. Acta*, 1998, **71**, 615.

5 K. D. M. Harris, M. Tremayne and B. M. Kariuki, *Angew. Chem., Int. Ed.*, 2001, **40**, 1626.

6 K. D. M. Harris, M. Tremayne, P. Lightfoot and P. G. Bruce, *J. Am. Chem. Soc.*, 1994, **116**, 3543.

7 J. S. O. Evans and I. Radosavljević Evans, *Chem. Soc. Rev.*, 2004 (DOI: 10.1039/b316901b).

8 R. A. Young (Editor), *The Rietveld Method*, International Union of Crystallography, Oxford, 1993.

9 W. I. F. David, *J. Appl. Crystallogr.*, 1999, **32**, 654.

10 H. M. Rietveld, *J. Appl. Crystallogr.*, 1969, **2**, 65.

11 J. W. Visser, *J. Appl. Crystallogr.*, 1969, **2**, 89.

12 P.-E. Werner, L. Eriksson and M. Westdahl, *J. Appl. Crystallogr.*, 1985, **18**, 367.

13 A. Boultif and D. Louër, *J. Appl. Crystallogr.*, 1991, **24**, 987.

14 R. A. Shirley, *CRYSFIRE*, Suite of Programs for Indexing Powder Diffraction Patterns, University of Surrey.

15 G. S. Pawley, *J. Appl. Crystallogr.*, 1981, **14**, 357.

16 A. Le Bail, H. Duroy and J. L. Fourquet, *Mater. Res. Bull.*, 1988, **23**, 447.

17 D. Ramprasad, G. B. Pez, B. H. Toby, T. J. Markley and R. M. Pearlstein, *J. Am. Chem. Soc.*, 1995, **117**, 10694.

18 B. M. Kariuki, D. M. S. Zin, M. Tremayne and K. D. M. Harris, *Chem. Mater.*, 1996, **8**, 565.

19 M. Tremayne, B. M. Kariuki and K. D. M. Harris, *Angew. Chem., Int. Ed. Engl.*, 1997, **36**, 770.

20 C. M. Freeman, A. M. Gorman and J. M. Newsam, in *Computer Modelling in Inorganic Crystallography*, (Editor: C. R. A. Catlow), Academic Press, San Diego, 1997.

21 Y. G. Andreev, P. Lightfoot and P. G. Bruce, *J. Appl. Crystallogr.*, 1997, **30**, 294.

22 W. I. F. David, K. Shankland and N. Shankland, *Chem. Commun.*, 1998, 931.

23 G. E. Engel, S. Wilke, O. König, K. D. M. Harris and F. J. J. Leusen, *J. Appl. Crystallogr.*, 1999, **32**, 1169.

24 E. J. MacLean, M. Tremayne, B. M. Kariuki, K. D. M. Harris, A. F. M. Iqbal and Z. Hao, *J. Chem. Soc., Perkin Trans. 2*, 2000, 1513.

25 P. Miao, A. W. Robinson, R. E. Palmer, B. M. Kariuki and K. D. M. Harris, *J. Phys. Chem. B*, 2000, **104**, 1285.

26 S. Pagola, P. W. Stephens, D. S. Bohle, A. D. Kosar and S. K. Madsen, *Nature*, 2000, **404**, 307.

27 Y. Tanahashi, H. Nakamura, S. Yamazaki, Y. Kojima, H. Saito, T. Ida and H. Toraya, *Acta Crystallogr., Sect. B*, 2001, **57**, 184.

28 H.-P. Hsu, U. H. E. Hansmann and S. C. Lin, *Phys. Rev. E*, 2001, **64**, 56707.

29 E. D. L. Smith, R. B. Hammond, M. J. Jones, K. J. Roberts, J. B. O. Mitchell, S. L. Price, R. K. Harris, D. C. Apperley, J. C. Cherryman and R. Docherty, *J. Phys. Chem. B*, 2001, **105**, 5818.

30 K. Shankland, L. McBride, W. I. F. David, N. Shankland and G. Steele, *J. Appl. Crystallogr.*, 2002, **35**, 443.

31 V. Brodski, R. Peschar and H. Schenk, *J. Appl. Crystallogr.*, 2003, **36**, 239.

32 V. Favre-Nicolin and R. Cerny, *Mater. Sci. Forum*, 2004, **443**, 35.

33 B. M. Kariuki, H. Serrano-González, R. L. Johnston and K. D. M. Harris, *Chem. Phys. Lett.*, 1997, **280**, 189.

34 K. D. M. Harris, R. L. Johnston and B. M. Kariuki, *Acta Crystallogr., Sect. A*, 1998, **54**, 632.

35 B. M. Kariuki, P. Calcagno, K. D. M. Harris, D. Philp and R. L. Johnston, *Angew. Chem., Int. Ed.*, 1999, **38**, 831.

36 B. M. Kariuki, K. Psallidas, K. D. M. Harris, R. L. Johnston, R. W. Lancaster, S. E. Staniforth and S. M. Cooper, *Chem. Commun.*, 1999, 1677.

37 G. W. Turner, E. Tedesco, K. D. M. Harris, R. L. Johnston and B. M. Kariuki, *Chem. Phys. Lett.*, 2000, **321**, 183.

38 E. Tedesco, G. W. Turner, K. D. M. Harris, R. L. Johnston and B. M. Kariuki, *Angew. Chem., Int. Ed.*, 2000, **39**, 4488.

39 E. Tedesco, K. D. M. Harris, R. L. Johnston, G. W. Turner, K. M. P. Raja and P. Balaram, *Chem. Commun.*, 2001, 1460.

40 E. Y. Cheung, E. E. McCabe, K. D. M. Harris, R. L. Johnston, E. Tedesco, K. M. P. Raja and P. Balaram, *Angew. Chem., Int. Ed.*, 2002, **41**, 494.

41 S. Habershon, K. D. M. Harris and R. L. Johnston, *J. Comput. Chem.*, 2003, **24**, 1766.

42 E. Y. Cheung, S. J. Kitchin, K. D. M. Harris, Y. Imai, N. Tajima and R. Kuroda, *J. Am. Chem. Soc.*, 2003, **125**, 14658.

43 D. Albesa-Jové, B. M. Kariuki, S. J. Kitchin, L. Grice, E. Y. Cheung and K. D. M. Harris, *ChemPhysChem*, 2004, **5**, 414.

44 K. D. M. Harris, R. L. Johnston and S. Habershon, *Struct. Bonding*, 2004, **110**, 55.

45 K. Shankland, W. I. F. David and T. Csoka, *Z. Kristallogr.*, 1997, **212**, 550.

46 G. Reck, R.-G. Kretschmer, L. Kutschabsky and W. Pritzkow, *Acta Crystallogr., Sect. A*, 1988, **44**, 417.

47 N. Masciocchi, R. Bianchi, P. Cairati, G. Mezza, T. Pilati and A. Sironi, *J. Appl. Crystallogr.*, 1994, **27**, 426.

48 R. E. Dinnebier, P. W. Stephens, J. K. Carter, A. N. Lommen, P. A. Heiney, A. R. McGhie, L. Brard and A. B. Smith, III, *J. Appl. Crystallogr.*, 1995, **28**, 327.

49 R. B. Hammond, K. J. Roberts, R. Docherty and M. Edmondson, *J. Phys. Chem. B*, 1997, **101**, 6532.

50 V. V. Chernyshev and H. Schenk, *Z. Kristallogr.*, 1998, **213**, 1.

51 C. C. Seaton and M. Tremayne, *Chem. Commun.*, 2002, 880.

52 R. J. Cernik, A. K. Cheetham, C. K. Prout, D. J. Watkin, A. P. Wilkinson and B. T. M. Willis, *J. Appl. Crystallogr.*, 1991, **24**, 222.

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

54 M. Tremayne, B. M. Kariuki and K. D. M. Harris, *J. Mater. Chem.*, 1996, **6**, 1601.

55 S. Habershon, G. W. Turner, B. M. Kariuki, E. Y. Cheung, A. Hanson, E. Tedesco, D. Albesa-Jové, M. H. Chao, O. J. Lanning, R. L. Johnston and K. D. M. Harris, *EAGER*, Computer Program for Structure Solution from Powder Diffraction Data, Cardiff University and University of Birmingham.

56 S. Habershon, E. Y. Cheung, K. D. M. Harris and R. L. Johnston, *Chem. Phys. Lett.*, 2004, **390**, 394.

57 E. Y. Cheung, K. D. M. Harris and B. M. Foxman, *Cryst. Growth Des.*, 2003, **3**, 705.

58 R. E. Dinnebier, F. Olbrich, S. van Smaalen and P. W. Stephens, *Acta Crystallogr., Sect. B*, 1997, **53**, 153.

59 D. Seebach, J. L. Matthews, A. Meden, T. Wessels, C. Baerlocher and L. B. McCusker, *Helv. Chim. Acta*, 1997, **80**, 173.

60 G. Bähr, in *FIAT Review of WWII German Science*, 1939–1946. Inorganic Chemistry, Part II, Klemm, W., Ed., Dieterichsche Verlagsbuchhandlung: Wiesbaden, 1948, p 155.

61 R. Dinnebier and J. Muller, *Inorg. Chem.*, 2003, **42**, 1204.

62 L. V. Interrante, G. A. Sigel, M. Garbauskas, C. Hejna and G. A. Slack, *Inorg. Chem.*, 1989, **28**, 252.

63 W. Łasocha and R. Gryboś, *J. Mol. Struct.*, 2002, **641**, 153.

64 M. Brunelli, J. P. Wright, G. B. M. Vaughan, A. J. Nora and A. N. Fitch, *Angew. Chem., Int. Ed.*, 2003, **42**, 2029.

65 W. H. Zachariasen and F. H. Ellinger, *Acta Crystallogr.*, 1963, **16**, 369.

66 K. Shankland, W. I. F. David and D. S. Sivia, *J. Mater. Chem.*, 1997, **7**, 569.

67 M. Edgar, V. J. Carter, D. P. Tunstall, P. Grewal, V. Favre-Nicolin, P. A. Cox, P. Lightfoot and P. A. Wright, *Chem. Commun.*, 2002, 808.

68 G. M. J. Schmidt, *Pure Appl. Chem.*, 1971, **27**, 647.

69 J. M. Thomas, *Philos. Trans. R. Soc.*, 1974, **277**, 251.

70 R. Hoffmann, *Liebigs Ann. Chem.*, 1857, **102**, 1.

71 A. Kekulé, *Liebigs Ann. Chem.*, 1858, **105**, 288.

72 L. Elizabé, B. M. Kariuki, K. D. M. Harris, M. Tremayne, M. Epple and J. M. Thomas, *J. Phys. Chem. B*, 1997, **101**, 8827.

73 H. Ehrenberg, B. Hasse, K. Schwarz and M. Epple, *Acta Crystallogr., Sect. B*, 1999, **55**, 517.

74 K. D. M. Harris, *Am. Pharm. Rev.*, 2004, **7**, 86.

75 E. Y. Cheung, K. D. M. Harris, R. L. Johnston, S. J. Kitchin, K. L. Hadden and M. Zakrzewski, *J. Pharm. Sci.*, 2004, **93**, 667.

76 I. Csoregh, *Acta Crystallogr., Sect. C*, 1992, **48**, 1794.

77 H. J. Bruins-Slot, H. J. Behm and H. E. M. Kerkkamp, *Acta Crystallogr., Sect. B*, 1990, **46**, 842.

78 E. Y. Cheung, K. D. M. Harris, R. L. Johnston, K. L. Hadden and M. Zakrzewski, *J. Pharm. Sci.*, 2003, **92**, 2017.

79 J. Giovannini, L. Ter Minassian, R. Ceolin, S. Toscani, M. A. Perrin, D. Louër and F. Leveiller, *J. Phys. IV*, 2001, **11**, 123.

80 H. Nowell, J. P. Attfield, J. C. Cole, P. J. Cox, K. Shankland, S. J. Maginn and W. D. S. Motherwell, *New J. Chem.*, 2002, **26**, 469.

81 R. E. Dinnebier, P. Sieger, H. Nar, K. Shankland and W. I. F. David, *J. Pharm. Sci.*, 2000, **89**, 1465.

82 Y. H. Kiang, A. Huq, P. W. Stephens and W. Xu, *J. Pharm. Sci.*, 2003, **92**, 1844.

83 F. C. Chan, J. Anwar, R. Cernik, P. Barnes and R. M. Wilson, *J. Appl. Crystallogr.*, 1999, **32**, 436.