DALTON
PERSPECTIVE

Solving crystal structures of molecular solids without single crystals: a simulated annealing approach

Yuri G. Andreev and Peter G. Bruce

School of Chemistry, University of St. Andrews, St. Andrews, Fife, UK KY16 9ST

Received 13th July 1998, Accepted 6th October 1998

The *ab initio* determination of relatively complex crystal structures of flexible molecules without the need for single crystals is discussed. A method is described based on simulated annealing in which the powder diffraction patterns of randomly generated trial structures are calculated and compared with the observed powder diffraction pattern in order to identify the model which provides the best fit and therefore the true structure. By employing simulated annealing both downhill (improved fit) and uphill (reduced fit) moves are possible ensuring escape from local minima in order to find the global minimum in the goodness-of-fit, *i.e.* the true structure. Key to the successful solution of flexible molecules is the introduction of a geometrical description which specifies atomic positions within the unit cell in terms of bond lengths and angles. In this way only those random structures which are chemically plausible are generated, greatly reducing the number of trial structures and rendering tractable the otherwise impossible task of *ab initio* determination. It is shown that structures with 37 variable parameters can be solved from only a few milligrams of powder. The limits of structural complexity for this method should be similar to those for refinement using powder data, *i.e.* around 200 variables. The variables may be those of position, or orientation of the molecule(s) in the unit cell as well as bond lengths, bond angles or torsion angles.

1 Introduction

Single crystal X-ray diffraction remains the technique of choice for obtaining accurate structural data. In some quarters the absence of a single crystal is taken to imply that the compound is amorphous and not therefore amenable to crystallographic study, this is generally incorrect. The powdered form of the compound is usually composed of crystals with dimensions in the range 0.1–1 μm. Although too small for single crystal studies, crystal structures may be obtained by analysing the powder X-ray diffraction data. Developments within the last two years have led to robust methods by which crystal structures of relatively high complexity may be solved *ab initio* from powder diffraction data collected on a few mg of sample. Critically these structures may be of any type and include molecules that are completely flexible. The structural data that may be obtained is of high quality, significantly exceeding that of most other structural techniques available to chemists. We are seeing the dawn of a new and exciting age of crystallography without single crystals.

In this paper we describe the simulated annealing (SA) method of *ab initio* structure solution from powder diffraction data illustrated by examples involving the determination of molecular crystals with up to 37 variable structural parameters.

2 The challenge of *ab initio* structure solution from powder diffraction data

The same basic information is contained in powder and single crystal diffraction data sets. In both cases the intensities arise from reflection of X-rays by the different sets of lattice planes, however in the latter case the intensities are distributed in three dimensions around the crystal whereas in the former they are

Peter Bruce was awarded a PhD in chemistry following work with Professor A. R. West on lithium ion conducting solid electrolytes at the University of Aberdeen. After a period of postdoctoral study with Professor J. B. Goodenough then Head of the Inorganic Chemistry Laboratory at the University of Oxford and a period as Lecturer at Heriot-Watt University in Edinburgh he moved to the University of St. Andrews in 1991 where he now holds a Chair in Chemistry and is currently Head of Department. His research interests span the structure and properties of inorganic solids and polymer salt complexes. He has been particularly involved in the structural chemistry, largely using crystallographic techniques, of lithium ion conducting solids, intercalation compounds and polymer electrolytes. He has held research Fellowships from the Royal Society and the Royal Society of Edinburgh. In 1995 he was elected to a Fellowship of the Royal Society of Edinburgh.

Peter Bruce

Yuri Andreev

Yuri Andreev received his MSc in Applied Nuclear Physics in 1981 and PhD in Solid State Physics in 1986 from Moscow Engineering Physics Institute. He has worked as a guest scientist at the University of Uppsala with Professors Torsten Lundström and Josh Thomas. Since 1995, he has been a research fellow at the University of St. Andrews working for Professor Peter Bruce. His research interests extend throughout all aspects of powder diffraction data treatment.

compressed onto one dimension. A powder diffraction pattern contains data from all the randomly orientated crystals exposed to the X-ray beam and the intensities consist of a set of peaks separated only by the value of the *d*-spacings between the planes from which they arise.

Since the structural information is contained in the intensities of the reflections, the key problem facing structure determination from powder diffraction data is the presence of peak overlap. Although some peaks are isolated many are overlapped either partially or completely (the latter when the *d*-spacings of different sets of planes coincide). This results in a loss of access to the individual intensities which are critical to the structure solution. Modern high-resolution instruments have improved greatly the separation of peaks however significant numbers of reflections cannot be resolved by any improvement in instrumentation.

A number of structures have been determined from powder diffraction data over the last twenty years. These have employed mainly proven single crystal approaches, particularly Direct Methods[1] or Patterson synthesis.[2] Inevitably they rely on extracting the intensities of individual reflections or groups of reflections using equipartitioning or permutation estimates for grouped reflections, then treating these as a single crystal data set. The combination of a limited number of intensities and the absence of short *d*-spacing reflections due to the truncated nature of powder diffraction data compared with single crystals, can conspire to frustrate these single crystal methodologies. Nevertheless they have played a significant role in *ab initio* structure solution from powders and will continue to do so through for example the medium of programs such as EXPO.[1] The SA approach described in this paper, differs fundamentally from the established methods in that no attempt is made to extract individual intensities and treat them in a single crystal sense. Instead effort is concentrated on generating chemically plausible but random structural models which are tested against the whole powder profile. A brief survey of currently available methods for structure determination from powder diffraction data is given in ref. 3.

Rietveld recognised the inevitable limitation of any approach which relies on obtaining individual intensities and in response introduced whole-pattern fitting.[4] His method involves calculating the entire powder profile (not just the individual intensities) based on a model structure then varying the structural parameters (*e.g.* atomic positions) of the model by least-squares until a profile is generated which best fits the observed profile. The best fit is determined by minimising the figure-of-merit function, χ^2, which acts as a measure of the goodness-of-fit. This approach has been used successfully to *refine* the details of partially known structures in many hundreds of cases. The only barrier to *ab initio* structure determination, as opposed to refinement, with the Rietveld approach is the use of the least-squares method which necessitates a starting structural model that is close to the correct structure because the least-squares routine can only adjust structural parameters in the direction of reducing χ^2 (downhill move), *i.e.* increasing the goodness-of-fit between the calculated and observed profiles. In other words the Rietveld method can locate only the local minimum in the goodness-of-fit hence the necessity to start from a structural model for which the local minimum coincides with the global minimum. By definition in *ab initio* structure determination the starting structural model will bear little relationship to the correct structure. The probability is negligible that starting from such a random structural model, the true structure (corresponding to the global minimum in the goodness-of-fit) can be obtained.

Among the minimisation methods capable of finding the global minimum of the goodness-of-fit in the presence of multiple local minima are the methods of SA[5] and genetic algorithm.[6] Minimisation by the latter using full-profile fitting of a powder pattern was successfully applied to the solution of the crystal structure of *ortho*-thymotic acid (2-hydroxy-3-isopropyl-6-methylbenzoic acid), the molecule of which contains only two internal degrees of freedom.[7] It is also shown mathematically that the genetic algorithm method, although competitive with SA approaches, can be expected to be orders of magnitude less efficient.[8]

In the SA method a Monte Carlo procedure is used to generate random models for the structure. This is achieved by making stepwise increments, random in size and direction, of the structural parameters (*e.g.* atomic coordinates). The models may yield a better fit (downhill, *i.e.* lower χ^2) or worse fit (uphill, higher χ^2) between the calculated and observed profile. Critically the latter permits escape from local minima. As the minimisation progresses, tolerance for the uphill steps gradually decreases until steps in both directions are exhausted. At this point the set of adjustable atomic coordinates corresponds to the lowest possible value of the figure-of-merit function, *i.e.* the global minimum. The potential of SA in the realm of structure determination was first demonstrated by solving the previously known crystal structure of benzene using a modified Rietveld method.[9] In addition to a SA minimisation the authors included a rigid-body representation of the constituent benzene rings. This allowed significant reduction of the number of variable structural parameters. The crystallographic coordinates of all the constituent atoms used in calculating the powder pattern were, in the case of benzene, computed using only positional and orientation parameters of the rigid body as a whole.

The rigid-body approach has been further exploited in the structure solution of molecular structures that are marginally more complex than that of benzene.[3,10] However the authors did not utilise minimisation of the full-pattern goodness-of-fit function but instead generated trial rigid-body structures in a Monte Carlo fashion analysing subsequently all the moves to select several low minima. The implementation of such an approach involves multiple subsequent refinements in order to identify the structure that corresponds to the lowest value of the figure-of-merit function. While successful for relatively rigid structures, it is likely that this approach will represent a barrier to solving structures with larger numbers of intermolecular degrees of freedom.

If structure solution from powder diffraction data is to become of general utility it is vital to develop a robust approach capable of tackling flexible as well as rigid structures. Determination of flexible structures by SA is much more challenging than rigid bodies, since the number of possible structural permutations appears at first sight to be enormous. Indeed it has been estimated that the computation would take up to 10^9 years even for a relatively simple structure.[11] We have chosen to adopt the SA approach used successfully for benzene but have extended significantly its application to embrace the range of crystals containing highly flexible moieties. Flexibility can be in the bond lengths, bond angles or torsion angles and we shall show that in some cases it is essential to vary all of these in order to achieve a successful structure solution. The modified SA method is also capable of tackling the solution of polymeric structures in which a single molecule straddles more than one asymmetric unit. The problem here is that random models of the asymmetric unit are only valid if they generate chain continuity. Thus the inter-atomic connectivity at the junctions of neighbouring asymmetric units is determined merely by the relevant symmetry operator of the space group and cannot be maintained by random variation of the relevant bond length, bond angle and torsion angle.

Key to our approach is the development of a stereochemical description that permits the atomic positions of the structural model to be defined in terms of bond lengths, bond angles and torsion angles, rather than individual atomic coordinates.[12] This in turn permits us to restrict attention to chemically plausible structural models thus vastly reducing the number of trial structures and rendering tractable the otherwise impossible task

of solving the crystal structures of flexible molecules. The approach to a geometrical description of flexible molecules *via* stereochemical parameters is suitable for describing a molecular fragment of any kind.

3 Simulated annealing

The principles behind the SA approach and its distinctive features in comparison with other methods of continuous minimisation are best understood by analogy with the process of forming a solid by cooling from a melt. Let us assume that the solid phase can be either amorphous or crystalline. At temperatures above the melting point, atoms have a high mobility and are in chaotic motion, the total energy of the ensemble is also high. The minimum energy of this system corresponds to the crystalline solid. The amorphous phase is at an intermediate position on the energy scale. There are two extreme routes by which the melt may be solidified: slow cooling or quenching. During the latter process, a random atomic configuration is immediately frozen, forming a glass the total energy of which is somewhat higher than in the crystalline state. If the rate of temperature decrease is low enough such cooling corresponds to an annealing process in which the chaotic motion of free atoms in the melt is gradually reduced allowing the ensemble to explore fully the energy space and hence to adopt the most energetically favourable (crystalline) configuration. Applying this thermodynamic reasoning to crystal structure solution from powder data requires substitution of the atoms of the melt by variable structural parameters of the ensemble (*e.g.* the atomic coordinates or bond lengths) and energy by the value of a goodness-of-fit function (χ^2).

The most frequently used method of minimisation is 'conventional' gradient least-squares.[13] This is the analogue of quenching. It does not allow the necessary chaotic changes but as described above allows only downhill movement to the local minimum. SA does permit the essential *uphill* steps necessary to escape local minima in the search for the global minimum corresponding to the true structure. The number of uphill steps representing choatic behaviour of the figure-of-merit function should however be slowly decreased by introducing a varying attenuation factor so that minimisation towards the global minimum may occur. This procedure is analogous to that of slow cooling (annealing) with the attenuation factor acting as the temperature.

A convenient way of introducing random steps in the structural model which includes the possibility of uphill moves is known as the standard importance sampling algorithm.[14] Application of this procedure to the minimisation of a figure-of-merit function, $\chi^2(\mathbf{P})$, whose value is determined by a set of the crystallographic parameters (*e.g.* bond angles and torsion angles) $\mathbf{P}$, may be outlined as follows. A new set of parameter values $\mathbf{P^i}$ (*i.e.* a new crystal structure model) is accepted if either $\chi^2(\mathbf{P^i}) < \chi^2(\mathbf{P^{i-1}})$ or if $\exp\{-[\chi^2(\mathbf{P^i}) - \chi^2(\mathbf{P^{i-1}})]/\Delta\chi^2_{cur}\} > R$, where $\mathbf{P^{i-1}}$ is a previously accepted set of parameters, $\Delta\chi^2_{cur}$ is a current marginal value of the χ^2 variation serving as a temperature analogue, R is a random number in the range from 0 to 1. In the case of continuous minimisation each j-th component p_j^i of the $\mathbf{P^i}$ set is calculated *via* the p_j^{i-1} value of the $\mathbf{P^{i-1}}$ vector in a Monte Carlo fashion, where Δp_j is a

$$p_j^i = p_j^{i-1} + r_j \cdot \Delta p_j \qquad (1)$$

predefined maximum stepwidth and r_j is a random number in the range from −1 to 1. Once $\mathbf{P^i}$ is accepted then $\mathbf{P^{i-1}} = \mathbf{P^i}$ and the process reiterates.

An account of various types of the "temperature-reduction" procedure is given in ref. 15. A recently reported SA protocol[16] for structure solution incorporates features of genetic algorithm and, as claimed, facilitates the search for the global minimum. Here we mention a temperature-reduction scheme which was used successfully in the examples of the structure solutions presented in Section 5. At a given value of $\Delta\chi^2_{cur}$ the sampling algorithm reiterates for as long as the total number of rejected and accepted $\mathbf{P^i}$ sets of parameters (referred to as moves from here on) exceeds the pre-set value of N_{tot} or until the number of accepted moves becomes greater than $f_1 \cdot N_{tot}$, with the f_1 value also chosen in advance. As soon as this happens the value of $\Delta\chi^2_{cur}$ is reset to $(1 - f_2) \cdot \Delta\chi^2_{cur}$ with a predetermined value of f_2 and the whole procedure continues. Minimisation terminates when there are no downhill moves at a current value of $\Delta\chi^2_{cur}$.

4 Constraints and restraints

Reduction in the number of structures that must be explored is essential to the success of the minimisation process. This is particularly important in crystal structure solution when the original structural model is likely to be a poor approximation to the true crystal structure and where minimisation of the figure-of-merit function must be performed in an extremely time consuming Monte Carlo fashion. This can be accomplished by imposing rigorous or hard constraints and soft or slack constraints (restraints) on the number of parameters or on their values. By these means only chemically plausible molecules (*e.g.* no unrealistic bond lengths) need to be explored.

4.1 Non-structural constraints

The most obvious constraint to be used in the course of structure solution by a full-pattern fitting approach is fixing the values of those profile-defining parameters that are not directly related to the arrangement of atoms within the unit cell. The values of cell constants, peak-shape and half-width parameters, background, peak asymmetry *etc.* are readily determined with reasonable accuracy using full-pattern decomposition methods without reference to a structural model.[17,18] The only parameter that has no effect on the structure but cannot be fixed in advance is the scale factor for the calculated pattern. However, its value is easily computed for each new trial structural model using the linear least-squares method. The effect of introducing these constraints is twofold. First, there is the advantage of a reduction in the number of variables. Second, the disadvantage is that it is unreasonable to expect the fixed values to provide the best fit to the experimental powder pattern when the profile is calculated using the structural parameters instead of the integrated intensities of individual Bragg peaks used during the full-pattern decomposition. In this case the restrained structure solution terminates close to the 'true' global minimum rather than in the minimum itself. To reach the minimum the subsequent trivial task of refinement of all parameters using the Rietveld method is required.

4.2 Structural restraints

Structural restraints are already an integral part of modern software for structure refinement by the Rietveld method and are a means of taking into account our knowledge about the possible atomic arrangement.[19] In most cases the restraints are introduced in calculation of the figure-of-merit function such that models which violate the restraining limits produce a high χ^2. These restraints do not eliminate unreasonable structural models from the refinement process. Such an approach is rendered unattractive for structure solution because unrealistic models are still generated and this is computationally demanding in SA. An entirely different approach to the imposition of restraints on structures not only preserves all the attributes of the established procedure but, in addition, allows randomised minimisation by SA strictly within the pre-determined limits of the structural parameters. In other words, increments leading to parameter values outside the limits are automatically rejected by SA but they can still be accepted by a refinement procedure

if they produce a lower overall (pattern fit plus geometry-related terms) χ^2 value.

In the case of molecular crystals the connectivity of atoms within the molecule is generally known. By describing the positions of the atoms in terms of their bond lengths, bond angles, and torsion angles rather than independent atomic coordinates only chemically plausible structural models need be explored by restraining the values of these stereochemical parameters to be within certain limits. This greatly reduces the number of trial structures; however they may be reduced further by checking for unfavourably close proximity of non-bonding atoms and for the spatial continuity of an infinite, within the crystallite size, moiety (*e.g.* polymer chain) at the junction of neighbouring asymmetric units. Although it is the stereochemical parameters that are altered to generate each new chemically plausible model, the crystallographic coordinates for each model are still required in order to calculate the powder profile using the conventional mathematical formalism. However these may be readily obtained in terms of the stereochemical parameters (bond lengths $l_{j,j+1}$, bond angles $\varphi_{j,j+1,j+2}$, and torsion angles $\tau_{j,j+1,j+2,j+3}$) by expressing initially the atomic coordinates for each molecule in a local Cartesian frame following the approach proposed in ref. 20. This type of description requires representation of each molecule as a sequence of chains. In cases when bond angles are more readily constrained than torsion angles it is sometimes more convenient to calculate the coordinates of chosen atoms by rotating bonds around each other.[21] Transformation of the atomic coordinates from the local Cartesian to the crystallographic frame introduces a set of additional parameters which determine the position, through crystallographic coordinates of the reference atom positioned at the origin of the local frame, and orientation, through Eulerian angles Θ, Φ, Ψ, of the molecular fragment, as a whole, in the unit cell. A detailed mathematical account of the stereochemical description of molecules and of the frame transformations is given elsewhere.[12]

Once this procedure is introduced into a Rietveld-type algorithm in which the original least-squares procedure is substituted by the method of SA, the restraints are imposed in a straightforward manner by allowing the parameters to accept only reasonable values within predetermined limits instead of punishing the value of $\chi^2(\mathbf{P})$ when the limits are violated. The range of values for bond lengths and bond angles between atoms in most types of compounds is readily available, torsion angles vary between $-\pi$ and π, the coordinates of the reference atom are kept within the boundaries of the asymmetric unit. The limits on the Eulerian angles are $0 \leq \Theta$, $\Psi < 2\pi$ and $0 \leq \Phi < \pi$. Such a description allows the introduction of further constraints which reduce the total number of parameters to be varied. For example chemical knowledge can indicate that all like bond lengths or bond angles (*e.g.* all l_{C-C} and all φ_{C-C-C} in a benzene ring) can be treated as variable but equal to each other, or that a certain part of the molecule is flat implying that the corresponding torsion angles can be kept at fixed values 0 or π, or the whole molecular fragment is rigid in which case only the values of the Eulerian angles and of the reference-atom coordinates are to be varied. Although introduction of the above constraints is computationally beneficial, it must be used with caution because in certain cases, illustrated below, even a slight reduction of the molecular flexibility can mislead the structure solution.

5 Examples

The procedure for structure solution from powder data is best understood by considering some examples. The following structure solutions were performed using X-ray powder diffraction patterns collected in steps of 0.02° in transmission mode on a STOE STADI/P diffractometer with Cu-Kα_1 radiation.

The computer code implementing the full-profile-fitting procedure with minimisation by the SA method was written in C. The code for pattern calculation was adapted from the CPSR software package.[22] All the structures presented here are described in the monoclinic space group, $P2_1/c$. Details concerning indexing and space group determination, essential stages of all modern structure solution procedures, are given in refs. 12 and 23. Profile parameters and lattice constants were fixed at the values obtained from profile-fitting using the CPSR program suite. The background was subtracted manually. The set of variable parameters **P** used in the SA runs included the overall isotropic temperature factor, B. Hydrogen atoms were ignored during the structure solution and were added only at the refinement stage. Unless otherwise stated, the minimisation of the reduced $\chi^2(\mathbf{P})$ by SA was performed using a value of 5 for the initial 'temperature' parameter ($\Delta\chi^2_{cur}$), $N_{tot} = 5000$, and $f_1 = f_2 = 0.1$.

The program was mounted on a dual Pentium 100 MHz PC running under Windows NT. Final structure refinements were performed using the Rietveld procedure included in the GSAS program package.[24]

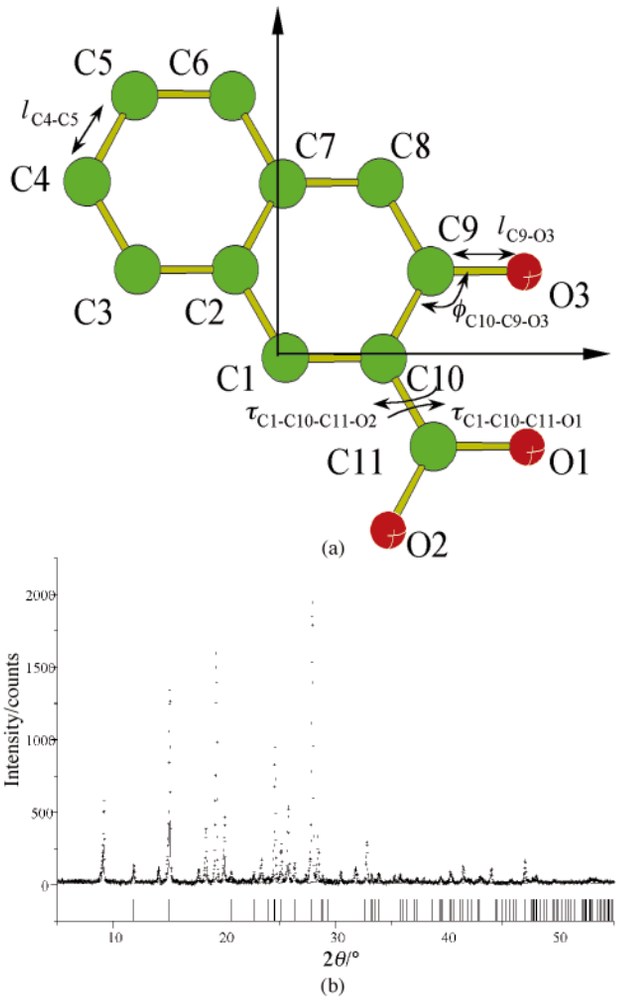

Fig. 1 (a) The 3-hydroxy-2-naphthoic acid molecule (hydrogen atoms are not shown) in a local Cartesian frame. (b) Experimental (+) and calculated (———) powder diffraction patterns, the latter from the randomly chosen initial structural model shown in Fig. 2(a). The background has been subtracted.

5.1 3-Hydroxy-2-naphthoic acid

The structure of 3-hydroxy-2-naphthoic acid $C_{10}H_6CO(OH)_2$ was solved originally using single crystal data.[25] In this section we discuss its solution using powder data as a test of the constrained SA approach.

The asymmetric unit consists of a single molecule which is

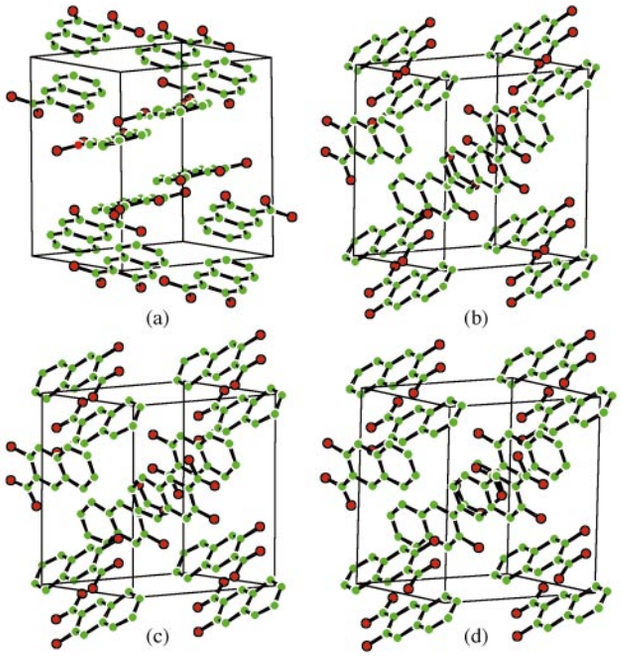

Fig. 2 Structural models of 3-hydroxy-2-naphthoic acid: (a) randomly chosen model as a starting point for the structure solution by the SA model; (b) after the SA run; (c) after structure refinement of (b) using the Rietveld method (hydrogen atoms are not shown); (d) as determined by the single-crystal-diffraction method (hydrogen atoms are not shown).

shown in Fig. 1(a). Aromaticity imposes planarity on the fused rings hence C1 to C11 and O3 are in a plane which is not necessarily the case for O1 and O2. These two features were taken into account when establishing the stereochemical description of the molecule in a local Cartesian frame with the origin located on the C1 atom [see Fig. 1(a)]. On the reasonable assumption that the hexagonal angles are ideal, the coordinates of the C2–C11 atoms are readily expressed *via* the C–C bond lengths ($l_{Ci,Cj}$) using trivial geometrical relationships. The coordinates of O3 are calculated *via* $l_{C9,O3}$ and $\varphi_{C10,C9,O3}$ as variable parameters with the values of $\tau_{C2,C1,C10,C9}$ and $\tau_{C1,C10,C9,O3}$ being fixed at 0 and π, respectively, to maintain the oxygen in the plane. The coordinates of O1 and O2 are expressed in a similar fashion with the only difference being that $\tau_{C2,C1,C10,C11}$ alone is fixed at π while the corresponding torsion angles, $\tau_{C1,C10,C11,O1}$ and $\tau_{C1,C10,C11,O2}$, are treated as variable parameters together with $l_{C11,O1}$, $\varphi_{C10,C11,O1}$ (in the case of O1) and $l_{C11,O2}$, $\varphi_{C10,C11,O2}$ (O2). The following additional assumptions were made to reduce the number of variables during the minimisation procedure $l_{C9,O3} = l_{C11,O1} = l_{C11,O2} \equiv l_{C-O}$, $\varphi_{C10,C9,O3} = \varphi_{C10,C11,O1} = \varphi_{C10,C11,O2} \equiv \varphi_{C-C-O}$, and all $l_{Ci,Cj} \equiv l_{C-C}$. Together with Θ, Φ, Ψ, x_{C1}, y_{C1}, z_{C1}, and B the total number of parameters which must be varied to generate random but chemically plausible structural models was 12.

The initial position and orientation of the molecule in the asymmetric unit was chosen at random giving the structure shown in Fig. 2(a) which offered a poor match to the experimental powder pattern [Fig. 1(b)]. After running the SA program during which the value of the figure-of-merit function calculated for each trial structure varied as shown in Fig. 3(a), the molecule 'froze' in the asymmetric unit at $\Delta\chi^2_{cur} = 0.035$ after over 140,000 trial structures and produced the structural model shown in Fig. 2(b) which yielded a significantly better fit to the experimental pattern [Fig. 3(b)].

Subsequent Rietveld refinement further improved the quality of the fit, $\chi^2 = 1.70$, [Fig. 3(c)] without changing significantly the appearance of the structure [Fig. 2(c)] which is in close agreement with that determined from the single-crystal data [Fig. 2(d)].

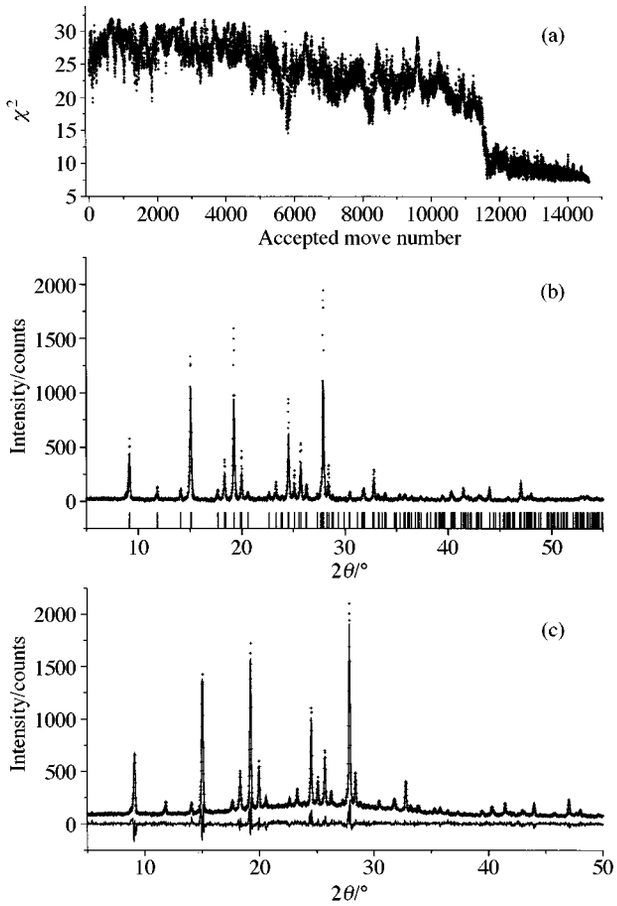

Fig. 3 (a) Variation of the figure-of-merit function during the structure solution of 3-hydroxy-2-naphthoic acid. (b) Observed (+) and calculated (———) powder diffraction patterns, based on the structural model shown in Fig. 2(b). The background has been subtracted. (c) Observed, calculated and difference powder diffraction patterns of 3-hydroxy-2-naphthoic acid after refinement by the Rietveld method (background included). Corresponds to the structural model shown in Fig. 2(c).

5.2 Poly(ethylene oxide)–salt complexes

Complexes of this type are composed of salts *e.g.* $LiCF_3SO_3$ dissolved in solid high-molecular-weight polymers, *e.g.* poly(ethylene oxide) (PEO). The polymer is a continuous linear chain with the repeat unit (CH_2–CH_2–O). Previous studies indicate that in complexes with ethylene oxide–salt ratios 3 : 1 and 4 : 1 the polymer chain adopts a helical conformation[26–28] while in the case of complexes with a 1 : 1 ratio the chain forms a stretched zigzag conformation.[29,30] The cations are coordinated by the oxygens of the chain, due to their strongly donating lone pairs, and the oxygens of the anion due to their partial negative charges.

5.2.1 $(PEO)_3$–$LiN(SO_2CF_3)_2$. Similarity between the lattice parameters of $(PEO)_3$–$LiN(SO_2CF_3)_2$ ($a = 12.034$ Å, $b = 8.660$ Å, $c = 19.139$ Å, $\beta = 128.5°$) and $(PEO)_3$–$LiCF_3SO_3$ ($a = 10.064$ Å, $b = 8.613$ Å, $c = 16.77$ Å, $\beta = 121.0°$)[27] suggested that the orientation and conformation of the PEO chain in $(PEO)_3$–$LiN(SO_2CF_3)_2$ may be similar to that found in $(PEO)_3$–$LiCF_3SO_3$, where the helical axis is parallel to the *b* axis and coincides with the 2_1 screw axis. Despite this, all attempts to refine the structure of $(PEO)_3$–$LiN(SO_2CF_3)_2$ on the basis of the known structure of $(PEO)_3$–$LiSO_3CF_3$, adjusted to the new dimensions of the unit cell, failed, as did attempts to solve the structure by approaches based on direct methods and difference-Fourier synthesis.

Density measurements suggested the presence of one formula unit in the asymmetric unit of the cell. Initially a SA

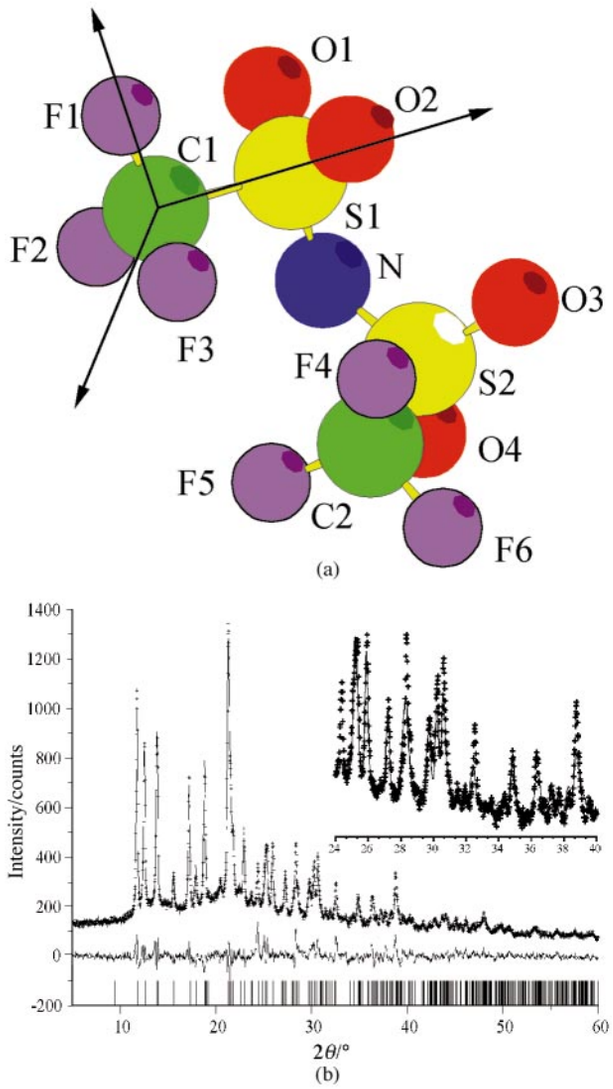

Fig. 4 (a) Imide anion in a local Cartesian frame. (b) Observed (+), calculated (———) and difference X-ray powder diffraction patterns for $(PEO)_3$–$LiN(SO_2CF_3)_2$ after refinement and following the SA run with fixed coordinates for the atoms belonging to the PEO chain. The insert shows an expansion of the region from 24 to 40° in 2θ.

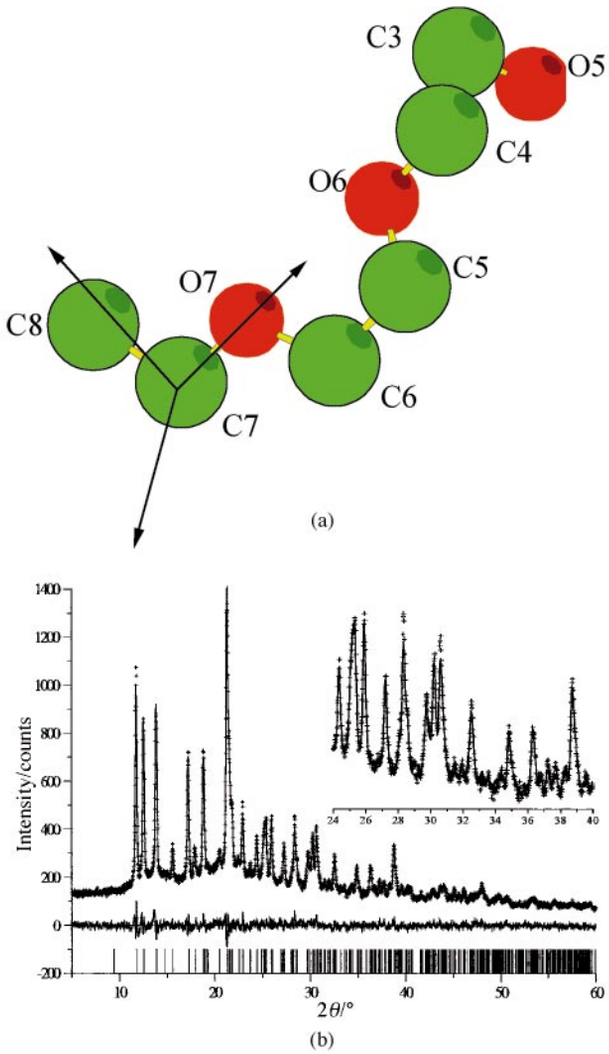

Fig. 5 (a) A fragment of the PEO chain in a local Cartesian frame. (b) Observed (+), calculated (———) and difference X-ray powder diffractions patterns for $(PEO)_3$–$LiN(SO_2CF_3)_2$ after refinement and following the SA run in which the position and conformation of the PEO chain were varied. The insert shows an expansion of the region from 24 to 40° in 2θ.

run was performed using fixed crystallographic coordinates for the atoms of the three EO units comprising the PEO chain taken from the $(PEO)_3$–$LiCF_3SO_3$ crystal structure but modified to account for the different unit cell dimensions. The Li^+ cation was placed inside the helix and its coordinates were allowed to vary. The coordinates of the atoms comprising the imide anion $N(SO_2CF_3)_2^-$ [Fig. 4(a)] were expressed in terms of stereochemical descriptors. The covalently bonded sequence F1–C1–S1–N–S2–C2–F4 was treated as a chain with the coordinates of the constituent atoms determined using the values of the consecutive bond lengths, bond angles, and torsion angles. Atoms F2, F3, O1, O2, O3, O4, F5, F6 were positioned by rotating C1–F1, S1–N, S2–C2, and C2–F4 bonds. The total number of structural parameters needed to define the structure was reduced from 55 to 24 by invoking the approximation that in the imide anion all bond lengths of a given bond type (*e.g.* all C–F or S–O bonds) are equal, all bond angles of a given type (*e.g.* all S–C–F or C–S–O angles) are equal, and all like torsion angles (N–S1–C1–F1 and N–S2–C2–F4) are equal. The set of parameters **P** used to calculate $\chi^2(\mathbf{P})$ included x_{Li}, y_{Li}, z_{Li}, x_{C1}, y_{C1}, z_{C1}, Θ_{imide}, Φ_{imide}, Ψ_{imide}, l_{C-F}, l_{S-C}, l_{S-O}, l_{S-N}, φ_{C-S-O}, φ_{S-C-F}, φ_{O-S-O}, φ_{F-C-F}, φ_{C-S-N}, φ_{N-S-O}, φ_{S-N-S}, $\tau_{N-S-C-F}$, $\tau_{C1,S1,N,S2}$, $\tau_{S1,N,S2,C2}$, and B. The SA run analysed ≈200,000 trial chemically plausible structural models. Approximately 15,000 of these were accepted while the rest were rejected either on the basis of the test for closest approach or by the Metropolis algorithm. The structural model, frozen at $\Delta\chi^2_{cur} = 0.02$, after subsequent refinement gave the fit to the experimental pattern shown in Fig. 4(b). Although the quality of fit is reasonably good, a noticeable mismatch is clearly seen in the inset of Fig. 4(b) indicating that the model is still inadequate.

To tackle the problem of the unsatisfactory fit the SA procedure was revisited, this time with extra flexibility added to the structural model. The polymer chain was allowed to vary its position and conformation in addition to the set of parameters involved in the first run [Fig. 5(a)]. Such a description added 12 parameters to the **P** set: x_{C7}, y_{C7}, z_{C7}, Θ_{PEO}, Φ_{PEO}, Ψ_{PEO}, $\tau_{O5,C3,C4,O6}$, $\tau_{C3,C4,O6,C5}$, $\tau_{C4,O6,C5,C6}$, $\tau_{O6,C5,C6,O7}$, $\tau_{C5,C6,O7,C7}$, $\tau_{C6,O7,C7,C8}$. The total number of rejected and accepted trial configuration at each 'temperature' was chosen to be $N_{tot} = 7000$ while the initial value of $\Delta\chi^2_{cur}$ was set to 0.5. Over 100,000 random structural models were generated with only 861 being accepted. Approximately 90% of the rejected trial models were discarded on the grounds of breaking the continuity of the PEO chain at the junctions of neighbouring asymmetric units. The best structural model was used in a new refinement which gave an excellent fit to the observed pattern [Fig. 5(b)]. Apart from a different chain conformation [Fig. 6(a)], the second SA run has revealed a different conformation for the SO_2CF_3

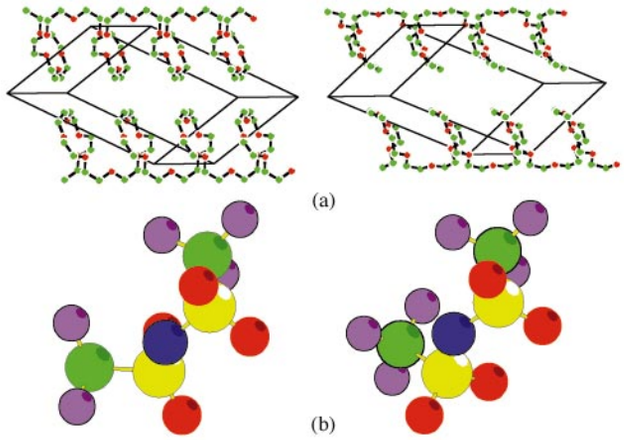

Fig. 6 (a) The PEO chain from the structural models corresponding to the fits shown in Fig. 4(b) (left) and Fig. 5(b) (right). (b) A single imide anion from the structural models corresponding to the fits shown in Fig. 4(b) (left) and Fig. 5(b) (right).

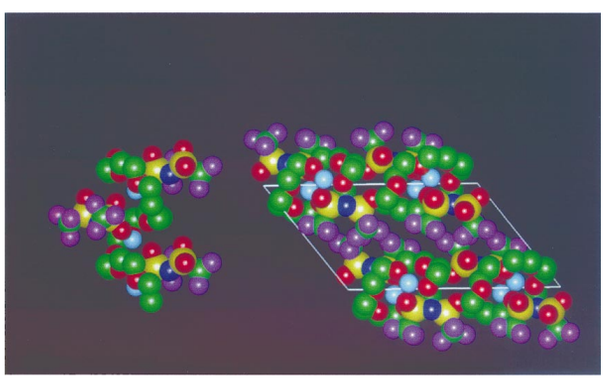

Fig. 7 Left, a portion of the PEO_3–$LiN(SO_2CF_3)_2$ structure showing a single polymer chain with associated ions. Right, view of the structure down the fibre axis. Light blue spheres, lithium; dark blue, nitrogen; yellow, sulfur; green, carbon; red, oxygen; purple, fluorine.

fragments of the imide group involving rotation about the S2–N bond [Fig. 6(b)], which did not appear during the first run with the chain fixed and could not be established in the course of the first refinement by the Rietveld method. The final structure of $(PEO)_3$–$LiN(SO_2CF_3)$ is shown in Fig. 7. Similar to other 3:1 complexes, one cation is located in each turn of the PEO helix and is coordinated by oxygen atoms. In this case three of the five oxygens coordinating Li^+ are from the chain while the other two belong to two imide anions. Each imide bridges neighbouring Li^+ ions along the chain by donating one oxygen to each Li^+. Note that both oxygens come from the same half of the imide, the other SO_2CF_3 moiety is not involved in coordination. Further discussion of the structure and details of the final refinement are given in ref. 31.

5.2.2 PEO–$NaCF_3SO_3$. Based on the observed density and unit cell volume, the asymmetric unit of PEO–$NaCF_3SO_3$ comprises a single EO unit, sodium cation, and triflate, $CF_3SO_3^-$, anion. The stereochemical description of the anion coincides with that of the NSO_2CF_3 moiety [see Fig. 4(a)] of the imide anion but with the nitrogen atom substituted by an oxygen.

The initial trial structure for the SA run [Fig. 8(a)] was chosen at random and, as can be seen from Fig. 8(b), did not provide a match between the calculated and observed diffraction patterns. During the minimisation 27 parameters were varied simultaneously with all bond lengths and bond angles associated with particular bond types in the triflate set to be equal. The full list of variable parameters included, for the triflate: x_S, y_S, z_S, Θ_{trif}, Φ_{trif}, Ψ_{trif}, l_{C-F}, l_{S-C}, l_{S-O}, φ_{C-S-O}, φ_{S-C-F},

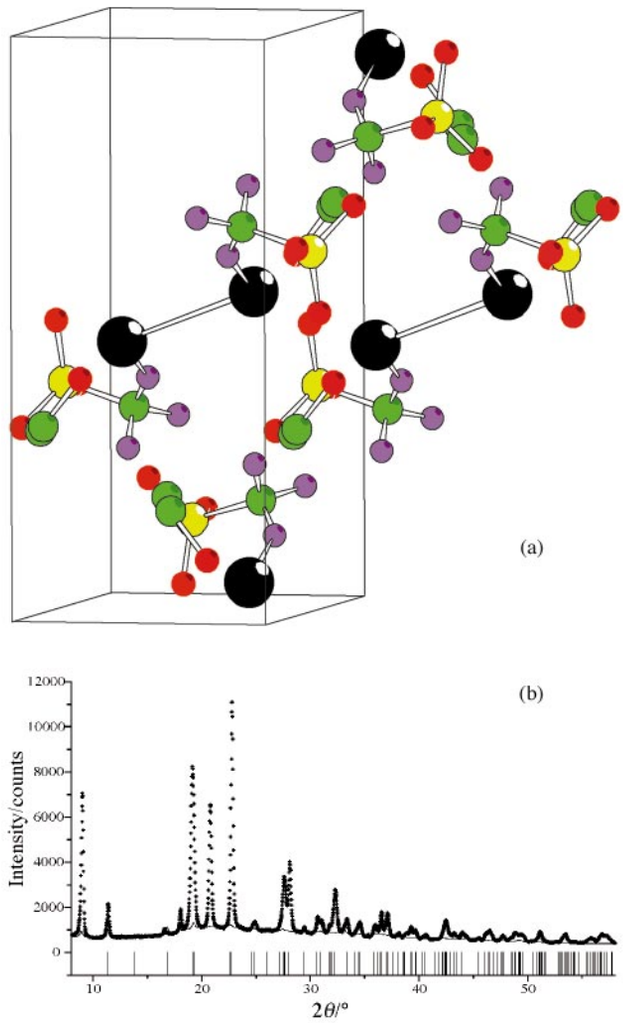

Fig. 8 (a) Randomly chosen initial trial structure of PEO–$NaCF_3SO_3$ used in the SA minimisation. Black spheres, sodium; the rest of the atom colours are the same as in Fig. 7. (b) Observed (+) and calculated (———) powder diffraction patterns of PEO–$NaCF_3SO_3$, the latter based on the above structural model.

φ_{O-S-O}, φ_{F-C-F}, $\tau_{O1,S,C1,F1}$; for the polymer chain: x_{C2}, y_{C2}, z_{C2}, Θ_{PEO}, Φ_{PEO}, Ψ_{PEO}, $l_{C2,C3}$, $l_{O4,C2}$, $\varphi_{O4,C2,C3}$; and for the sodium ion: x_{Na}, y_{Na}, z_{Na}; B. This constrained SA run produced a structural model [Fig. 9(a)] with a continuous PEO chain along the shortest cell axis giving a reasonable profile fit after subsequent refinement by the Rietveld method [Fig. 9(b)]. However all attempts to improve the fit further by refinement failed leaving the best χ^2 equal to 6 and a noticeable misfit in the 2θ range from 33 to 55° [see insert in Fig. 9(b)]. The refined model placed fluorines rather than the more negatively charged oxygens of the triflate anion adjacent to the Na^+ cation and did not ensure coordination of the sodiums by the chain oxygens [see Fig. 9(a)]. In addition, the separation of adjacent Na^+ ions was only 3.26 Å which is unlikely given the cation radius of 1.02 Å. Negative values of the thermal parameter B for some of the atoms provided further evidence indicating the inappropriateness of the structural model.

Successful structure determination was achieved only after removing the constraint that all like bond lengths and bond angles in the triflate were equal. A new SA minimisation was performed allowing all such lengths and the angles to vary independently. During this run 37 structural parameters were varied in a random fashion to generate the trial structures but with the imposition of chain continuity. The structural model obtained after further refinement [Fig. 10(a)] revealed sixfold coordination of the Na^+ ion by equidistant oxygens from the triflate and the chain and provided an excellent match between

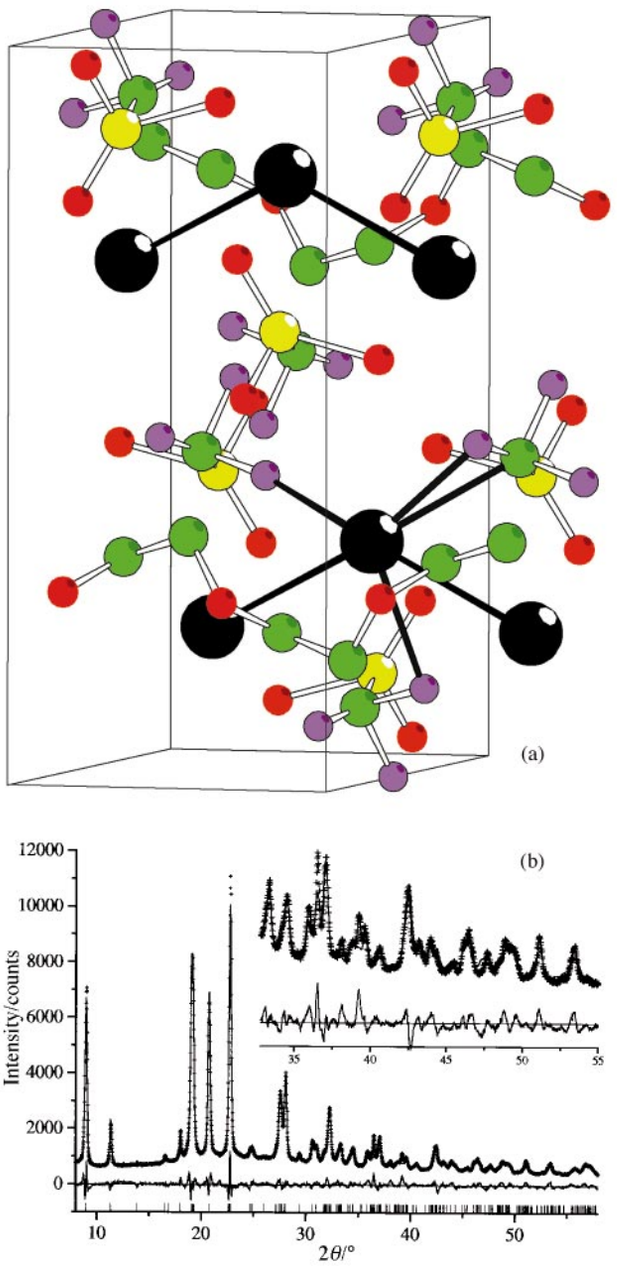

Fig. 9 (a) Refined structural model of PEO–$NaCF_3SO_3$ after the SA run with all like bond lengths and bond angles in the triflate ion treated as equal. Solid lines connect the Na^+ cation to its nearest neighbours. (b) Observed (+), calculated (———) and difference powder diffraction profiles for the above structural model of PEO–$NaCF_3SO_3$.

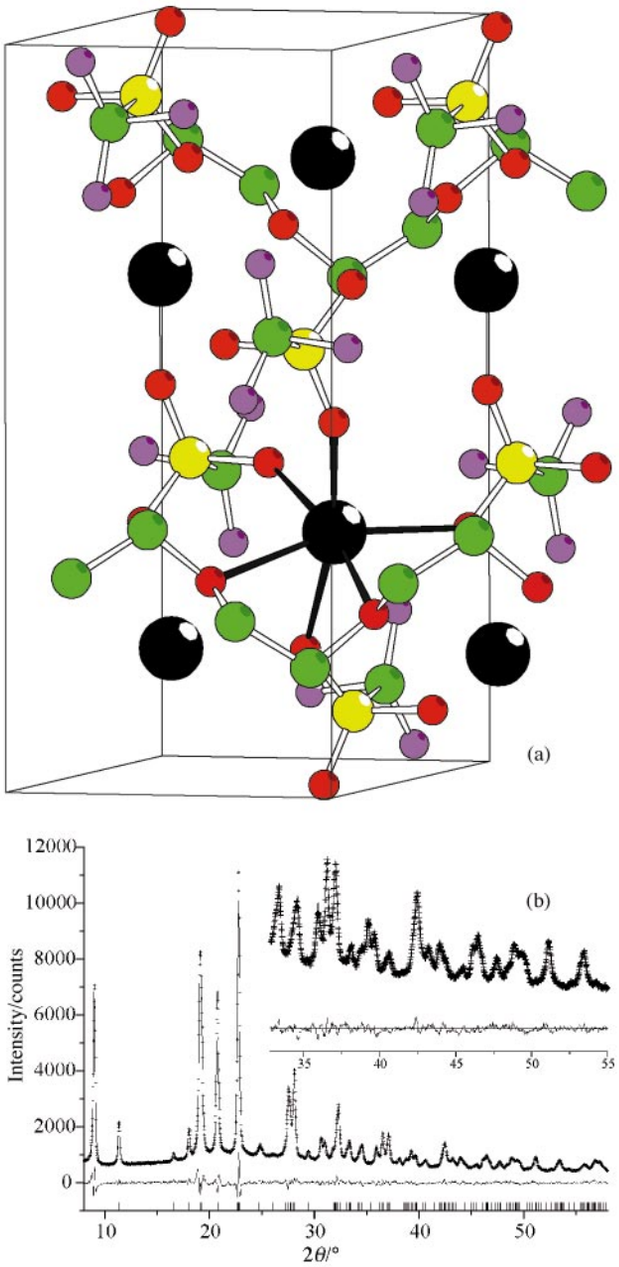

Fig. 10 (a) Refined structural model of PEO–$NaCF_3SO_3$ after the SA run in which all like bond lengths and bond angles in the triflate ion were varied independently. Solid lines connect the Na^+ cation to its nearest neighbours. (b) Observed (+), calculated (———) and difference powder diffraction profiles for the above structural model of PEO–$NaCF_3SO_3$.

the observed and calculated patterns with $\chi^2 = 1.1$ [Fig. 10(b)] and all B values positive.

The deleterious effect of averaging the bond lengths and angles of the triflate on the structure solution is dramatic and could not have been anticipated in advance since the same constraint did not negate location of the internal configuration and position of the imide ion (see Section 5.2.1) with almost twice as many like bond lengths and bond angles set to be equal. Nevertheless the distribution of the scattering power among the constituent atomic species in the case of PEO–$NaCF_3SO_3$ was such that a random search using the constrained model was biased from the start and could not yield the correct solution. As an illustration, Fig. 11 shows the significant change in the appearance of the calculated diffraction pattern after averaging the bond lengths and bond angles of the triflate in the final structural model. Further computational details and discussion of the structure may be found in ref. 15.

6 Conclusion

We have demonstrated that by combining the SA method with a geometrical description, which defines the atomic positions in terms of bond lengths and angles, it is possible to determine, *ab initio*, relatively complex crystal structures of flexible molecules without the need for single crystals.

In crystallography, the solution of structures containing different atoms of comparable scattering power (*e.g.* C, N, O) is often regarded as presenting the greatest challenge. This task is readily tackled for molecular structures by the SA approach.

Recently it has been suggested that instead of calculating a complete powder diffraction pattern for each random structure and fitting it directly to the observed pattern, this process could be divided into two.[16,32] Stage 1 involves fitting the observed pattern with a series of individual peaks thus yielding a set of integrated intensities. In stage 2 the randomly generated struc-

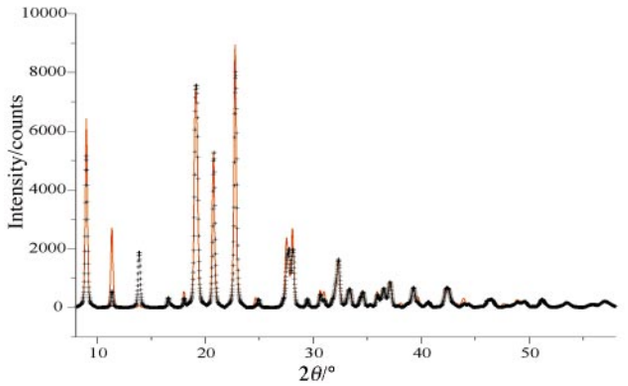

Fig. 11 Calculated powder diffraction pattern of PEO–$NaCF_3SO_3$ based on atomic coordinates obtained in the final refinement (———). Calculated powder diffraction pattern of the modified PEO–$NaCF_3SO_3$ structure by averaging all like bond lengths and bond angles in the triflate ion (+).

tures are tested against the set of intensities rather than the entire profile. Molecular structures, with up to 10 degrees of freedom (torsion angles), were solved using SA minimisation of the figure-of-merit function based on integrated intensities.[32] The advantage of this approach is that it offers a significant reduction of over 100 fold in the computational time. Typically the full profile SA method requires between 24 and 48 hours on a dual 100 MHz Pentium PC. As is frequently the case in modern crystallography, the computational efficiency is much greater than that of the other essential steps in the process of structure determination. Often the time taken to prepare the compound, collect high quality data, index the powder pattern and write the paper, considerably exceeds the computation time! In the case of relatively complex structures, peak overlap is likely to be severe and there has been considerable debate whether, in such circumstances, the fashion in which group-overlapped intensities are dealt with in the two-stage method is valid or leads to a loss of information compared with full-profile fitting. It is of course in the structure solution of such complex compounds that the increased computational efficiency would be particularly advantageous.

Perhaps the most important question which arises in constrained SA is the degree of flexibility which must be introduced into a molecule in order to determine its crystal structure using either full profile fitting or the two stage process. If the structure is permitted to be fully flexible (*i.e.* all bond lengths, bond angles and torsion angles in addition to positional and orientational parameters of each fragment in the asymmetric unit are independent variables) then it is possible to reach the global minimum corresponding to the best possible fit to the data. When the asymmetric unit of the structure under determination consists of a single isolated molecule for which interatomic connectivity is well-established and the bond lengths and angles are particularly well defined, such as in the case of 3-hydroxy-2-naphthoic acid (Section 5.1) or the examples given in refs. 3, 10, 11, 16 and 32, then only a few variables are required in the SA minimisation in order to solve the structure. The required parameters are those defining the position and orientation of the molecule as well as the torsion angles; the bond lengths and angles may be fixed at typical values. However as clearly demonstrated in this paper, such an approach may be insufficient in many cases when the asymmetric unit consists of more than a single molecule. If the interactions between separate molecules in the asymmetric unit involve van der Waals force or ionic bonding, then they are largely non-directional and many relative positions of the fragments are possible for any given set of internal conformations for each fragment. As a result there may be many more local minima in the goodness-of-fit function which are sufficiently deep to be confused with the global minimum. In such circumstances a final discrimination between different fits of the calculated and observed data is essential before refinement can be expected to yield the correct structural model. This requires bond lengths and angles to be varied independently *in addition* to torsion angles. Variation of torsion angles alone will not suffice. The presence of ionic bonding between separate moieties in the structure may be particularly troublesome since such bonding is stronger than van der Waals' force and can perturb the internal dimensions of covalently bonded moieties compared with the case of two or more neutral molecules. The examples of structure determination we have presented here constitute a particularly severe test of the methodology in the context of these difficulties since the asymmetric unit comprises several independent moieties which interact both *via* van der Waals forces and ionic bonding. In the case of $(PEO)_3$–$LiN(SO_2CF_3)_2$ it was sufficient to set all similar bond lengths, bond angles and torsion angles in the imide anion as single variables while changing only the conformation, *i.e.* the torsion angles, of the polymer chain with fixed bond lengths and angles. A structural model which was obtained using similar constraints during the solution of PEO–$NaCF_3SO_3$ was very different from the true model despite giving a reasonably good fit to the experimental data. The correct model was found only when all stereochemical restraints were removed and all the parameters were varied in a random fashion.

Of particular interest to those wishing to exploit the constrained SA approach, is the level of structural complexity that may be tackled. The successful structure solution of compounds requiring the independent variation of 37 parameters and up to 25 symmetry-unrelated atoms has been demonstrated in this paper. All possible structural variables (position, orientation, bond lengths, bond angles and torsion angles) are included. The SA method does not discriminate between different types of structural variables. We could have selected a system in which all 37 parameters were torsion angles (had attention been confined to a compound with a single molecule in the asymmetric unit). In fact the upper limit for structure solution by whole pattern fitting using the SA minimisation combined with stereochemical description should be comparable to the limit of structural complexity found for Rietveld refinement which to date is about 60 non-symmetry related atoms. A better measure of complexity is the number of structural variables and the limit in this context should be around 200. Of course these numbers can only be taken as a rough guide since, as these limits are approached, the level of structural complexity that may be tackled will depend on the individual system.

SA has been applied previously in the case of energy minimisation. It is important to appreciate that no energy parameterisation of the system is involved here, *i.e.* there is no requirement for interatomic potentials. Minimisation is carried out with the use of the observed diffraction data alone.

6 Acknowledgements

The authors are grateful to Dr P. Lightfoot for his collaboration during the structure solution of $(PEO)_3$–$LiN(SO_2CF_3)_2$ and for useful discussions. Thanks are due to Mr G. S. MacGlashan and Mr L. J. M. Sawers. The financial support of this work by the EPSRC and The Leverhulme Trust is gratefully acknowledged.

7 References

1 A, Altomare, M. C. Burla, M. Camalli, G. Cascarano, C. Giacovazzo, A. Guagliardi, A. G. G. Moliterni, G. Polidori and R. Rizzi, *Mater. Sci. Forum*, 1998, **278**, 284.
2 M. J. Buerger, *Vector space*, Wiley, New York, 1959, pp. 167–168.
3 K. D. M. Harris and M. Tremayne, *Chem. Mater.*, 1996, **8**, 2554.
4 H. M. Rietveld, *J. Appl. Crystallogr.*, 1969, **2**, 65.
5 S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, *Science*, 1983, **220**, 671.

6 J. H. Holland, *Adaptation in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor, 1975.
7 B. M. Kariuki, H. Serrano-Gonzalez, R. L. Johnston and K. D. M. Harris, *Chem. Phys. Lett.*, 1997, **280**, 189.
8 L. Ingber and B. Rosen, *Math. Comput. Modell.*, 1992, **16**, 87.
9 J. M. Newsam, M. W. Deem and C. M. Freeman, *Accuracy in Powder Diffraction II*, NIST Special Publication No. 846, 1992, pp. 80–91.
10 K. D. M. Harris, M. Tremayne, P. Lightfoot and P. G. Bruce, *J. Am. Chem. Soc.*, 1994, **116**, 3543.
11 K. Shankland, W. I. F. David, T. Csoka and L. McBride, *Int. J. Pharm.*, 1998, **165**, 117.
12 Y. G. Andreev, P. Lightfoot and P. G. Bruce, *J. Appl. Crystallogr.*, 1997, **30**, 294.
13 *International Tables for Crystallography*, Kluwer Academic Publishers, Dordrecht, Boston, London, 1995, vol. C, pp. 594–608.
14 N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *J. Chem. Phys.*, 1953, **21**, 1087.
15 W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in C: the Art of Scientific Computing*, Cambridge University Press, 1992, pp. 451–455.
16 W. I. F. David, K. Shankland and N. Shankland, *Chem. Commun.*, 1998, 931.
17 G. S. Pawley, *J. Appl. Crystallogr.*, 1981, **14**, 357.
18 A. Le Bail, H. Duroy and J. L. Fourquet, *Mater. Res. Bull.*, 1988, **23**, 447.
19 C. Baerlocher, in *The Rietveld Method*, *IUCr Monographs on Crystallography 5*, ed. R. A. Young, Oxford University Press, 1993, pp. 186–196.
20 S. Arnott and A. J. Wonacott, *Polymer*, 1966, **7**, 157.
21 *International Tables for X-ray Crystallography*, Kynoch Press, Birmingham (Present distributor Kluwer Academic Publishers, Dordrecht), 1959, vol. 2, pp. 62–63.
22 Y. G. Andreev, T. Lundström and N. I. Sorokin, *Nucl. Instrum. Methods Phys. Res.*, *Sect. A*, 1995, **354**, 134.
23 Y. G. Andreev, G. S. MacGlashan and P. G. Bruce, *Phys. Rev. B*, 1997, **55**, 12011.
24 A. C. Larson and R. B. Von Dreele, Los Alamos National Laboratory Report Number LA-UR-86-748, 1987.
25 M. P. Gupta and B. P. Dutta, *Cryst. Struct. Commun.*, 1975, **4**, 37.
26 Y. Chatani and S. Okamura, *Polymer*, 1987, **28**, 1815.
27 P. Lightfoot, M. A. Mehta and P. G. Bruce, *Science*, 1993, **262**, 883.
28 P. Lightfoot, J. L. Nowinski and P. G. Bruce, *J. Am. Chem. Soc.*, 1994, **116**, 7469.
29 M. Yokoyama, H. Ishihara, R. Iwamoto and H. Tadokoro, *Macromolecules*, 1990, **2**, 184.
30 Y. Chatani, Y, Fujii, T. Takayanagi and A. Honma, *Polymer*, 1990, **31**, 2238.
31 Y. G. Andreev, P. Lightfoot and P. G. Bruce, *Chem. Commun.*, 1996, 2169.
32 K. Shankland, W. I. F. David and T. Csoka, *Z. Kristallogr.*, 1997, **220**, 550.

Paper 8/05437A