feature articles

Structure determination from powder diffraction data

W. I. F. David* and K. Shankland

ISIS Facility, Rutherford Appleton Laboratory, Chilton, Oxon OX11 0QX, UK. Correspondence e-mail: bill.david@rl.ac.uk

Advances made over the past decade in structure determination from powder diffraction data are reviewed with particular emphasis on algorithmic developments and the successes and limitations of the technique. While global optimization methods have been successful in the solution of molecular crystal structures, new methods are required to make the solution of inorganic crystal structures more routine. The use of complementary techniques such as NMR to assist structure solution is discussed and the potential for the combined use of X-ray and neutron diffraction data for structure verification is explored. Structures that have proved difficult to solve from powder diffraction data are reviewed and the limitations of structure determination from powder diffraction data are discussed. Furthermore, the prospects of solving small protein crystal structures over the next decade are assessed.

1. Introduction

Over the past decade, structure determination from powder diffraction data (SDPD) has matured into a technique that, although not completely routine, is widely and successfully used in the context of organic, inorganic and organometallic compounds (David *et al.*, 2002). Although inorganic crystal structures generally have simpler chemical formulae and smaller unit-cell dimensions than organic materials, they are often more complicated to solve than their organic counterparts and many new inorganic structures are determined from powder diffraction data by analogy with chemically similar materials. However, for new and more complex systems, such as molecular sieves and a new generation of mixed metal oxides, chalcogenides and hydrides, crystal-structure solution from first principles can be a challenge. There are a number of reasons that this may occur. Firstly, the structural symmetry of inorganic materials is often significantly higher than their organic counterparts and thus the degree of complete reflection overlap is higher. Secondly, the topology of organic materials is generally straightforward to comprehend – isolated molecules of known connectivity pack together, leading to an easy parameterization for global optimization. Inorganic materials often consist of connected polyhedra and the topology of this connectivity is not generally known *a priori* and thus parameterization for global optimization is often less straightforward. Of course, these restrictions do not apply to direct methods and Patterson methods, which explains why they are still currently dominant in this field.

Although crystal structures have been solved from powder diffraction data from the earliest days of X-ray crystallography, an important marker for SDPD occurred a decade ago in 1998. Over the previous few years, global optimization methods for SDPD had begun to show significant potential (see, for example, Harris *et al.*, 1994; Ramprasad *et al.*, 1995) and a 'blind test' involving two unknown crystal structures (one inorganic, one organic) was organized by Le Bail and Cranswick in order to assess the maturity of the field (http://www.cristal.org/SDPDRR). Although the organizers allowed six weeks for the structure solution and there were 70 downloads of the diffraction data, only two successful solutions were reported for the molecular organic test structure, tetracycline hydrochloride, $C_{22}H_{24}N_2O_8 \cdot HCl$ (Fig. 1).

The organizers justifiably concluded that SDPD was not a routine procedure for the majority of researchers. However, the two successful solutions indicated that algorithms did exist to determine pharmaceutical structures from powder diffraction data apparently quite routinely. Interestingly, the two methods of solution were very different; one solution involved the use of traditional Patterson methods to first locate the Cl^- atom and subsequent cycles of Fourier synthesis to reveal the remaining atoms; the other employed a simulated-annealing global optimization technique (David *et al.*, 1998) and was
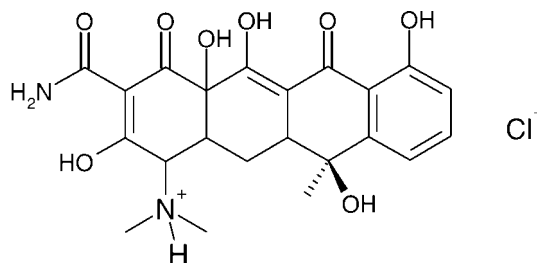

Figure 1
The molecular formula of tetracycline hydrochloride.

found to be the most accurate answer supplied with reference to a subsequently determined single-crystal structure.

Over the ten years since this round-robin challenge, it has been the latter strategy (and other similar strategies) that has proven to be the most effective in the generation of new crystal structures from powder data. The availability of easy-to-use computer programs, coupled with continual innovation in the area of algorithms, has meant that the many rather than the few can now take advantage of the power of SDPD. Fig. 2 underlines this, showing the steady rise in results obtained from global optimization methods, but also the regular stream of results from direct methods.

In this paper, we discuss a number of recent developments and present examples of the determination of organic, inorganic and biological crystal structures. Our main focus is, however, on molecular organic materials, as this is the area
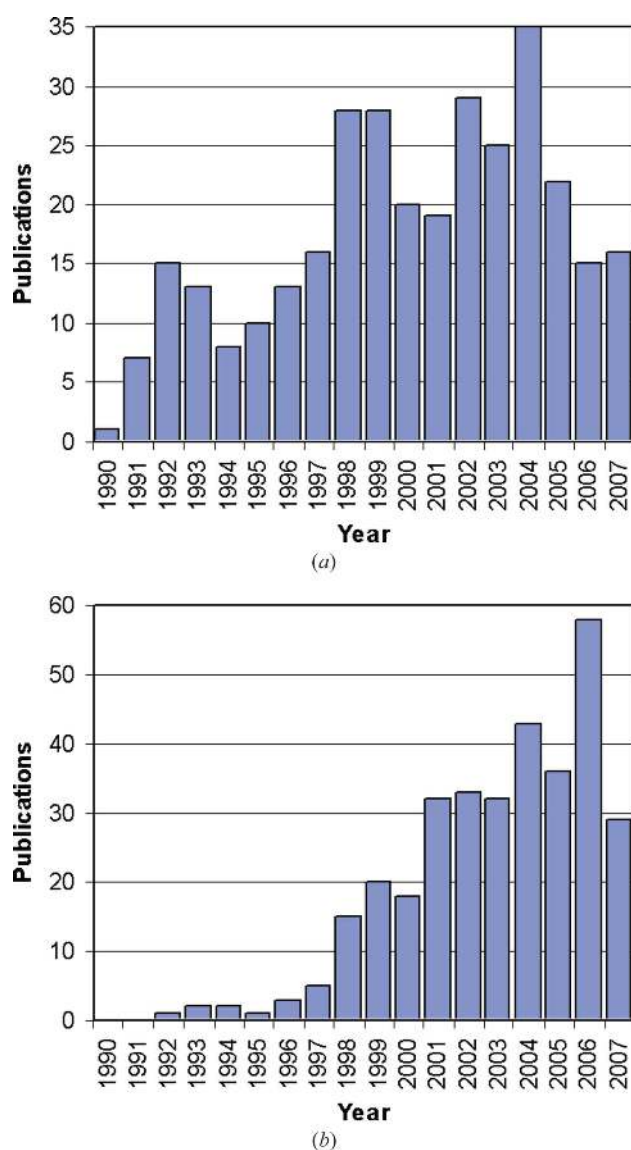
that has seen the most significant expansion in the last decade in terms of published crystal structures.

2. Algorithm developments

2.1. Introduction

SDPD is a sequential process with clearly defined stages and at each stage there can be problems that make it impossible to proceed (David *et al.*, 2002). The majority of problems are caused by the collapse of the three dimensions of reciprocal space to the single dimension of a powder diffraction pattern, with the resultant Bragg-peak overlap being particularly severe at shorter d spacings. This places some fundamental restrictions upon the amount of information that can be derived from the pattern; these restrictions are discussed in considerable detail in Appendix A and the reader's attention is drawn to this section, as an understanding of the restrictions is key to assessing future developments in SDPD.

It is apparent that, in such a sequential process, care has to be taken at all stages; even at the sample preparation stage, recrystallization to improve sample crystallinity or light grinding to improve powder averaging can lead to significantly better data. The traditional bottlenecks of indexing and structure solution turn out to be intimately linked; advances in structure-solution methods have stretched the capabilities of SDPD to the extent that the study of relatively large structures (with correspondingly large unit cells and sometimes very long axes) is now quite common. This size increase has exposed some of the limitations of well established powder-indexing programs that were written at a time when looking at structures (largely inorganic) with much smaller unit cells was in fact the norm. As such, it was to a large extent success in structure solution that suggested it was time to develop new algorithms and strategies for powder indexing.

2.2. Indexing

Unit-cell determination is an essential first step in structure solution. In most methods, peak positions are extracted and then trial unit cells are assessed in order to determine the correct lattice parameters. With high-resolution data, this process is often straightforward. However, with poorer data and particularly when the sample contains more than one crystalline phase, indexing can become a serious bottleneck. If the data appear reasonable and indexing does fail, then the most likely cause is the existence of more than one phase. Identification of known phases should be attempted and the corresponding peak positions removed from the indexing process. If, however, no known phases are identified, then allowance must be made for the possibility of 'impurity' phases such as starting materials – dividing peaks into groups of 'sharp' and 'broad' may help. If this fails, then two experimental approaches can also be used: (i) resynthesis (or recrystallization) of the material under different experimental conditions – for example, in the case of two polymorphic forms, this may lead to different proportions of the two polymorphs which can then aid in the identification of the two

Figure 2
Approximate numbers of publications involving direct methods and powder diffraction (upper graph) and global optimization and powder diffraction (lower graph) as a function of time since 1990. Source: Web of Knowledge search, September 2007.

feature articles

distinct sets of Bragg peak positions; and (ii) heating the sample – one phase may disappear at elevated temperatures. Differential scanning calorimetry is particularly useful in pre-screening for this effect.

Perhaps the two most significant algorithmic developments in indexing over the past decade have been (i) the incorporation of the possibility of impurity phases into the exhaustive successive dichotomy algorithms of *DICVOL04* (Louer & Boultif, 2006) and (ii) the development of a singular-value-decomposition-based algorithm in *TOPAS* (Coelho, 2003). The *X-Cell* program (Neumann, 2003), which uses an extinction-specific dichotomy procedure to perform an exhaustive search of parameter space, is also capable of handling impurity phases and zero-point errors. Whilst it is difficult to assess the impact that these relatively recent developments have had upon the success or otherwise of indexing in general, it is certainly true to say that these programs, being of recent origin, are well suited to the indexing of materials with large unit cells and their increased use is likely to decrease the reliance on the older strategy of trying several distinct indexing programs in order to obtain a likely solution.

2.3. Space-group determination

Whereas the inability to determine the correct unit cell makes structure solution impossible, some uncertainty in the correct space group is not an intractable problem as each potential space group may be separately tested. Space groups are determined by examining diffraction patterns for systematically absent reflections. For example, in a monoclinic system, if all $0k0$ reflections (k odd) are absent, then it is probable that there is a 2_1 screw axis and that space group $P2_1$ is more likely than $P2$. Determining absences for long d-spacing reflections is normally straightforward, but at shorter d spacings reflection overlap makes it a much more subjective process. For example, a 010 reflection will typically be well separated from other reflections, making accurate determination of its intensity easy. In contrast, the 030/050/070 reflections are much more likely to lie in clumps of overlapped reflections, leading to difficulties in accurate intensity estimation. Accordingly, conclusions about contributing space-group-symmetry elements are generally drawn on the basis of a very small number of clear intensity observations. While observing lattice-centring extinctions is usually relatively easy, the determination of the correct space-group-symmetry elements is generally more challenging. Choosing the space group with the fewest number of contributing reflections, the process of parsimony, is a good pragmatic approach to resolving this problem and, in the case of molecular organic materials, considerable help in space-group selection comes from the well known frequency distribution of space groups, where some 80% of compounds crystallize in one of the following: $P2_1/c$, $P\bar{1}$, $P2_12_12_1$, $P2_1$ and $C2/c$. For relatively simple systems with a small number of atoms in the unit cell, the structure may be solved in $P1$ and the space group subsequently determined from a search for the appropriate

symmetry elements – this is a useful alternative strategy particularly for small inorganic structures. However, over the past decade, probabilistic approaches to space-group determination have been developed that remove the need for subjective judgements about the presence or absence of classes of reflections throughout the pattern (Markvardsen *et al.*, 2001; Altomare, Caliandro, Camalli, Cuocci, da Silva *et al.*, 2004). Following a model-independent fit (Pawley or Le Bail) to the diffraction data in the holosymmetric space group consistent with the observed crystal class, the user is presented with a list of possible extinction symbols ranked in order of probability. Armed with the most likely extinction symbol, it is usually a straightforward matter to pick the most likely space group (though see §3.2 for an example of a more difficult case). As such, these methods are particularly useful when dealing with systems of orthorhombic symmetry or higher, where the number of possible space groups (and settings) is relatively large.

2.4. Direct methods, Patterson methods and charge flipping

Direct methods of structure determination dominate the field of single-crystal diffraction because of their incredible success rate, range of applicability, speed, ease of use and reliability. However, in general, they expect to work on large numbers of well determined reflection intensities collected to atomic resolution. For a powder diffraction measurement, this situation is rarely the case; see the detailed discussion in Appendix *A* for further details. Even if 'good' data can be obtained to 1 Å (this is often the case with inorganic structures, particularly with neutron powder diffraction data), then, for all but the simplest materials, observed diffraction features can contain contributions from several overlapped reflections meaning that the condition of 'well determined' reflection intensities is not met. Experimental methods, such as texture analysis (Wessels *et al.*, 1999) and differential lattice expansion through multiple temperature measurements (Shankland, David & Sivia, 1997; Fernandes, 2006) can create almost single-crystal-like data sets which have been successfully used to solve crystal structures using direct methods. However, it is adaptations of traditional direct methods, specifically tailored to the analysis of powder diffraction data, which have been used most successfully in recent years. The best known package is *EXPO2004* (Altomare, Caliandro, Camalli, Cuocci, Giacovazzo *et al.*, 2004) which has evolved constantly and now incorporates indexing, space-group determination, structure solution and refinement capabilities (Altomare *et al.*, 2006). It can be applied to organic (Brunelli *et al.*, 2007; Altomare *et al.*, 2007), organometallic (Masciocchi & Sironi, 2005) and inorganic crystal structures (Fukuda *et al.*, 2007), but the majority of reported structures solved using *EXPO* (in its various versions) falls into the latter two groups, reflecting both the particular strengths of the package and its traditional user base.

As illustrated by one of the contributors to the first SDPD round robin, Patterson methods, even in a standard form, can be a powerful tool for structure determination in cases where

there are strongly scattering atoms. Indeed, some years ago it had been shown that they can also be applied where large fragments of known geometry are involved (Rius & Miravitlles, 1988). In an important series of papers (see, for example, Rius *et al.*, 2000, 2007), Rius and his colleagues have described modifications to traditional direct methods based around Patterson-function arguments that have permitted the solution of many very complex materials (see, for example, Corma *et al.*, 2003), whilst the usefulness of the Patterson function in decomposing overlapping Bragg peaks has also been shown (Estermann & David, 2002). Perhaps the most comprehensive approach is based upon hyperphase permutation algorithms (Bricogne, 1991) but their use remains to be fully exploited.

Recently, the charge-flipping method (Oszlányi & Süto, 2008) has been adapted to powder diffraction data with some very promising results (Baerlocher, Gramm *et al.*, 2007; Baerlocher, McCusker & Palatinus, 2007). As yet, it appears that it is still subject to the requirement for near atomic resolution data.

2.5. Global optimization methods

Global optimization methods of SDPD involve moving a molecular model of the molecule under study around a known unit cell, constantly adjusting its conformation, position and orientation until the best agreement with the observed diffraction data is obtained. Of course, there is no guarantee that the best minimum obtained for the agreement function will be the global one (corresponding to the correct crystal structure) but we are fortunate in diffraction that the value obtained can be compared with a value obtained for a corresponding Pawley- or Le Bail-type fit to the data, in order to inform us how close we in fact are to the 'best' fit obtainable.

That these methods have been so successful in the context of SDPD is almost entirely due to the fact that they incorporate a massive amount of prior chemical information in the form of the known molecular topology of the material under study; typically, all known bond lengths and angles for the molecule are fixed, leaving only its conformation, plus its position and orientation within the unit cell to be determined. It is this information that compensates for the reduced information content of the powder pattern.

2.5.1. Grid search. Grid-type searches represent the simplest form of the global optimization method available, where every parameter that defines the search space is explored on a systematic grid. Their one significant advantage is that, given a fine enough grid, one is guaranteed to find the global minimum. However, naïve grid-type searches are computationally intractable for problems of even relatively low complexity (Shankland & David, 2002) and are not at all competitive with other search methods of the type described below. That said, with some suitable modification, they can prove extremely useful – see, for example, Ivashevskaja *et al.* (2003) and Mazina *et al.* (2004).

2.5.2. Stochastic search algorithms. The most popular global optimization methods for the SDPD of organic compounds in recent years have been stochastic in nature. These methods can be described colloquially as a 'random walk through the good solutions'. Many different stochastic methods have been explored, including simple Monte Carlo (Harris *et al.*, 1994), genetic algorithms (Shankland, David & Csoka, 1997; Harris *et al.*, 2004; Feng & Dong, 2007), evolutionary strategies (Chong & Tremayne, 2006) and particle swarm (Csoka & David, 1999) but it is simulated annealing (Andreev *et al.*, 1997; Engel *et al.*, 1999; Putz *et al.*, 1999; Pagola & Stephens, 2000; Coelho, 2000; Favre-Nicolin & Cerny, 2004; David *et al.*, 2006) that is most widely used and that has had the largest impact. Its high level of success is undoubtedly attributable to the fact that it is an extremely effective algorithm that is easy to use (it has relatively few control variables, all of which can be set automatically) and, as such, is suitable for use by typical practitioners of powder diffraction. Various modifications to the basic annealing scheme, such as parallel tempering (Earl & Deem, 2005), have been implemented (Favre-Nicolin & Cerny, 2004) in order to further improve the ability of the algorithm to sample the search space efficiently.

Of the other stochastic algorithms, genetic algorithms in particular have been shown to be effective and capable of delivering solutions of considerable complexity (Albesa-Jove *et al.*, 2004; Pan *et al.*, 2006). Their usage has, to date, been limited by program availability though very recently a freely available program *GEST* has been released (Feng & Dong, 2007).

2.5.3. Deterministic algorithms. Whilst stochastic techniques have been shown to be effective global optimization methods, many other algorithms from different research areas remain to be evaluated in the context of SDPD and it is likely some of these will be more efficient and successful than the techniques currently in use. One example is the hybrid Monte Carlo (HMC) algorithm which combines the key components of Monte Carlo (MC) and molecular dynamics (MD) approaches into a single algorithm. An extensive discussion of HMC in the context of SDPD may be found in the paper of Johnston *et al.* (2002). In essence, HMC may be considered in terms of a particle that follows a trajectory determined by Hamilton's equations of motion in a hyperspace defined by a set of structural variables. The total energy of the particle at any point is equal to the sum of the kinetic energy and potential energy given by the goodness-of-fit target function. Whilst in theory the total energy is conserved, the use of finite time step sizes in the numerical evaluation of the equations of motion means that this is not the case. To control this effect, a Metropolis acceptance criterion is used to determine whether to accept or reject the configuration at the end of a given MD trajectory. The trajectory either continues from the end point if it is accepted or returns to the previous start point if it is rejected (Fig. 3). The effectiveness of HMC has been convincingly demonstrated with the structure determination of capsaicin which, with a total of 15 degrees of freedom, represents a moderate challenge for SDPD. When compared with the default implementation of simulated

annealing in *DASH*, HMC is a factor of two more successful in locating the global minimum over a series of 20 repeat runs. Significantly, the HMC algorithm required considerably fewer χ^2 evaluations than simulated annealing to achieve this level of success. Remarkably, given the discussion in Appendix *A*, the quickest solution required less than 20000 evaluations to locate the radius of convergence corresponding to 10^{-11} of the total parameter space.

2.5.4. Semi-global and local searches. The rate of convergence of stochastic algorithms can sometimes be improved by the incorporation of elements of local searching, such as steepest descent, into the overall minimization strategy. For example, the program *DASH* (David *et al.*, 2006) uses a semi-global simplex-type algorithm to further minimize the cost function at the end of each simulated annealing run; the program *Organa* (Brodski *et al.*, 2005) uses simple gradient minimization when appropriate and applies it to all of its cost functions; a sequential quadratic programming subroutine from the NAG library has been used to implement Larmackian-type evolution in the context of a genetic algorithm-based approach (Turner *et al.*, 2000). Some indication of the potential of the simplex algorithm is given by the fact that, in a simple test, of 500 simplex runs started from random points in *DASH*, one was actually successful in accurately solving the crystal structure of the form B polymorph of famotidine ($C_8H_{15}N_7O_2S_3$, $P2_1/c$, $Z' = 1$, 13 degrees of freedom).

2.5.5. Maximum-likelihood techniques. As mentioned at the start of §2.5, the principal reason for the success of global optimization methods in SDPD is the incorporation of the molecular geometry into the solution process. However, this strength is also the principal limitation of the technique – the complete (and correct!) molecular structure must be incorporated if the global least-squares minimum is to be reached. This limitation may, however, be relaxed if more generalized maximum-likelihood methods are used. This approach has

been widely adopted by the macromolecular crystallography community and has recently been applied successfully to solve structures from powder diffraction data (Markvardsen *et al.*, 2002; Favre-Nicolin & Cerny, 2004). Consider that the majority (but not all) of the structural contents is to be determined in the optimization process. This can occur, for example, when there are disordered solvent molecules present in the structure in addition to the main molecule of interest. It might also occur if structural fragments are omitted from the optimization process in order to decrease the complexity of the global search. In such cases, maximum-likelihood optimization still allows the majority of the structure to be correctly located. Use of this approach has been illustrated with the examples of the nitrate and acetate salts of the anticonvulsant agent remacemide. If the nitrate and acetate ions are excluded from a standard least-squares global optimization then the structures cannot be solved; the best solutions, whilst informative in that they show parts of the remacemide molecule located at the positions of the acetate and nitrate ions in an attempt to account for their scattering contribution, are not sufficiently close to the true structure to allow structure completion. In contrast to this, if the nitrate and acetate ions are not explicitly considered in the maximum-likelihood optimization, the remacemide ion is quickly and correctly located for both structures with a very high success rate. It is then a trivial matter to subsequently fix the remacemide ion within the unit cell and then locate and orient the nitrate and acetate molecules by global optimization. This approach relaxes the constraint that the correct molecular contents are included from the outset of the global optimization process – for materials such as hydrates and solvates or zeolites and molecular sieves with guest molecules, this is an important consideration.

2.5.6. Incorporating additional chemical information. Constraints form a fundamental part of most global optimization approaches, with bond lengths, bond angles and fixed dihedral angles in the material under investigation typically being held at known standard values during the optimization process. Note, however, that some practitioners advocate the use of 'loose restraints' (Favre-Nicolin & Cerny, 2004) in order to allow faster convergence to a minimum. That said, structural variables are typically restricted to the external molecular degrees of freedom plus those internal torsion angles whose values cannot be assigned *a priori*. Use of the Cambridge Structural Database (Allen, 2002) can help to provide likely bounds for torsion angles and the concept can be further extended to non-bonded contacts. While database mining can place bounds on likely torsion-angle values, the direct use of additional structural information from other techniques to determine torsion-angle values is more effective. For example, if the complete molecular conformation can be determined in advance of the diffraction experiment, global optimization is reduced to a problem of determining the position and orientation of a rigid molecule. Middleton and colleagues (Middleton *et al.*, 2002) outlined such a procedure in which a set of interatomic distances is measured by rotational-echodouble resonance (REDOR) SS-NMR. The
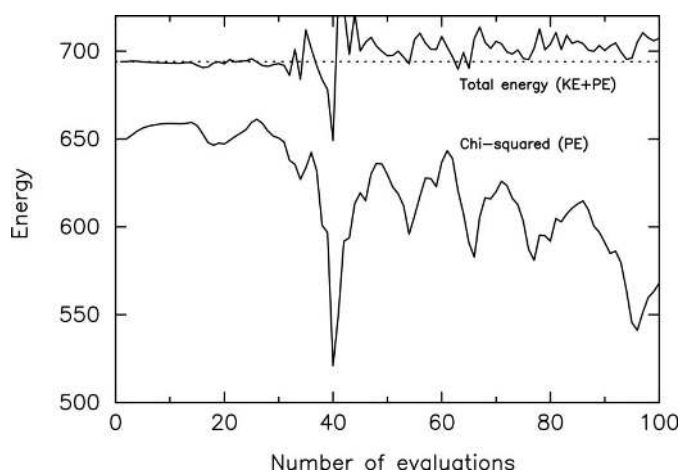

Figure 3
The potential energy (correlated integrated intensities χ^2) and total energy (kinetic energy plus potential energy) evaluated over a single MD trajectory during the crystal-structure solution of capsaicin. The initial total energy is shown as a dotted line in order to highlight the total energy fluctuations arising from the finite MD step size.

molecular conformation is then derived from a restrained molecular-dynamics optimization in which the use of high harmonic force constants ensures that all conformations in the simulation have interatomic distances that satisfy the input distances. The best conformation is then optimized against the X-ray powder diffraction data by global optimization. By way of example, the anti-ulcer drug cimetidine, in polymorphic form A, was solved from X-ray powder diffraction data using *DASH* with a MD optimized model derived from four SS-NMR-determined C—^{15}N distances. Each torsion angle in the MD-optimized starting model was allowed to vary $\pm 20°$ from its input value. In terms of structure-solution performance, this model delivers a speed and reliability approaching that of a rigid-body optimization. However, routine application will probably only be possible when the SS-NMR methodology develops to a stage where isotopic labelling is no longer a prerequisite and when the specialized SS-NMR instrumentation required is more commonly available.

An alternative way of biasing the search towards favourable molecular conformations and packing motifs is the incorporation of potential energy as an additional term in the overall cost function (Putz *et al.*, 1999; Coelho, 2000; Lanning *et al.*, 2000; Brodski *et al.*, 2005). The overhead in calculating such energies for simple van der Waals type interactions is small, though a suitable force field is required and a weighting factor is needed to balance the diffraction and energy contributions in the calculation of the overall cost function.

Another approach (Brenner *et al.*, 1997, 2002) utilizes a periodic nodal surface calculated from a few phased strong low-index reflections to divide the unit cell into regions of high and low electron density. In the case of molecular organic materials, the resultant 'structure envelope' can be used as a boundary within which to restrict the possible position/orientation/conformation of the molecule within the unit cell, leading to a significant reduction in the search space that needs to be explored.

2.5.7. Parallel computing. In the absence of a fine-grained grid search (a prohibitively slow method as mentioned earlier), none of the global optimization methods mentioned above can guarantee finding the global minimum in the relevant parameter space in a finite time frame. As such, it is prudent to perform multiple global optimization runs in order to improve the chances of locating the global minimum or some point sufficiently close to it to permit final refinement of the structure. Indeed, for complex structures with a large number of parameters, where the success rate in finding the global minimum can fall to a very small number (perhaps 1% or less), it is a *necessity* to perform multiple runs. This can turn SDPD into a highly CPU intensive process, where one might have to wait days for an answer, even when using highly efficient cost functions such as the method of correlated integrated intensities (David *et al.*, 1998). Fortunately, each run is independent of any other, and a simple and attractive option is to distribute the individual runs across a number of different computers/CPUs/cores in order to return the answer more quickly. Such a 'grid-type' computing approach to SDPD using both simulated annealing (as implemented in *DASH*)

and HMC has been described previously (Markvardsen *et al.*, 2005) and the speed gains to be had are, to a first approximation, proportional to the number of CPU cores contributing to the grid system. As such, speed gains of two orders of magnitude or more, over the already highly efficient execution speeds for *DASH* and HMC, can be expected from a modest grid of non-dedicated PCs. The importance of such speed gains is twofold: firstly, it allows results to be obtained on time scales that are more commensurate with the expectations of crystallographers for structure determination; secondly, it allows the parallel exploration of alternative strategies for solving the problem in hand, such as the use of multiple different starting models (*e.g. cis* and *trans* isomers), different diffraction data ranges, the inclusion of preferred-orientation corrections and the use of lower cooling rates in the annealing process.

Of course, there is nothing new in the parallel execution of large optimization problems, even in the context of SDPD – see, for example, Shankland, David & Csoka (1997) and Habershon *et al.* (2003). What is most significant about the work described above is the utilization of systems (such as *Condor* and *GridMP*) to harness non-dedicated PC resources; in this, crystallographers are following the trend set in other scientific areas such as protein–ligand docking, demonstrating that dedicated hardware (such as a Beowulf cluster) is not a pre-requisite to accessing massive computing power.
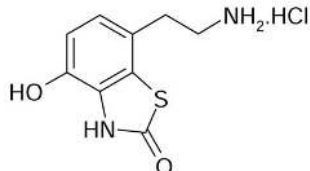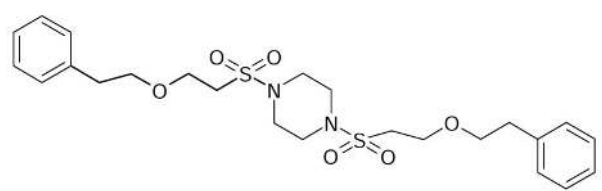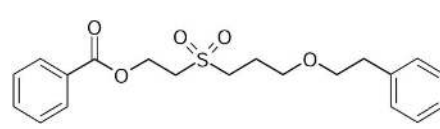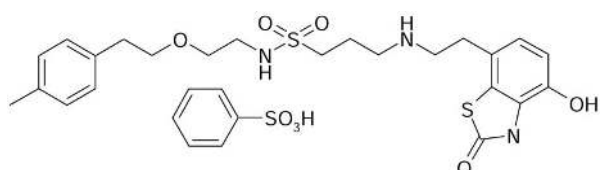
3. Examples

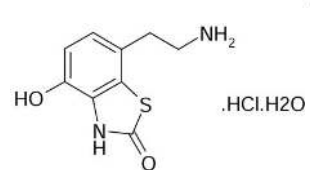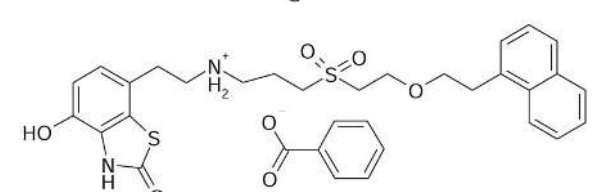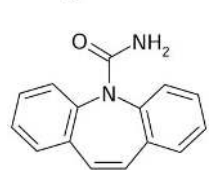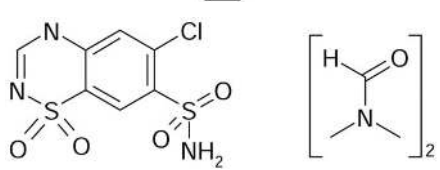### 3.1. Organic crystal structures

Some state-of-the-art results from SDPD from pharmaceuticals and organic crystal structures can be found in recent doctoral work (Docherty, 2004; Fernandes, 2006), where the structure determination (using *DASH*) and refinement (using *TOPAS*) of numerous compounds of pharmaceutical interest (see Table 1) from mainly laboratory X-ray powder diffraction is reported. As can be seen from the molecular formulas, degrees of freedom and number of independent fragments in the asymmetric unit, these compounds span a wide range of chemical and crystallographic complexity, yet all were solved relatively straightforwardly. The success rate in finding the global minimum fell to only a few percent for the most complex examples and this indicates that tackling still more complex examples will require further algorithmic developments. Of particular note are: (*a*) the benzoate structure, where the anion is in fact disordered and where the location of this disordered fragment was determined directly by global optimization of two 50% occupancy benzoates (Johnston *et al.*, 2004), (*b*) the γ form of carbamazepine, where $Z' = 4$ and there are 120 atoms in the asymmetric unit (Fernandes *et al.*, 2007) and (*c*) the dimethyl formamide solvate of chlorothiazide, where there are six independent fragments in the asymmetric unit and a total of 42 degrees of freedom (Fernandes *et al.*, 2007).

Other good examples include the crystal structures of a series of novel cyclic molecules (Terent'ev *et al.*, 2007) and of some mono-unsaturated triacylglycerols (van Mechelen *et al.*,

Table 1
Some complex pharmaceuticals solved from laboratory XRPD data.

N_{frag} = number of fragments in the molecular formula, SG = space group, Z' = number of formula units in the asymmetric unit, N_{tor} = number of flexible torsion angles in each formula unit, N_{dof} = total number of degrees of freedom.

Molecular formula	N_{frag}	SG	Z'	N_{tor}	N_{dof}
	2	$P2_1/a$	1	2	11
	1	$P2_1/a$	1	7	10
	1	$P2_1/n$	1	12	18
	2	$P\bar{1}$	1	14	26
	3	$Pca2_1$	2	2	28
	2	$P\bar{1}$	1	14	27
	1†	$P\bar{1}$	4	1	28
	3	$P2_1/c$	2	3	42

† Synchrotron data, ID31, ESRF.

2006a,b). The latter in particular highlight the contribution of prior chemical knowledge in deriving structures from relatively low quality diffraction data (Fig. 4).

Direct methods currently play a less important role in this area, although for organic materials with strongly scattering atoms present they continue to dominate – see, for example, Boufas et al. (2007). Nevertheless, the recent successful solution of phase II of bicyclo[3.3.1]nonane-2,6-dione (Brunelli et al., 2007) from very high quality synchrotron data, coupled with the continual developments in direct methods,

particularly in respect of density-map interpretation within the *EXPO* package (Altomare *et al.*, 2007) suggests that these still have a great deal to offer.

3.2. Inorganic crystal structures

There is no shortage of impressive examples of inorganic crystal structures solved from powder data; see, for example, recent work (Baerlocher, Gramm *et al.*, 2007) which shows that large zeolite structures can, with care, be determined. The IM-5 structure contains 24 crystallographically distinct Si atoms and was solved using the recently developed charge flipping algorithm along with structure envelope constraints and ancillary electron diffraction measurements.

It is fair to say that the topological uncertainties inherent in the determination of inorganic structures causes complications, particularly in the use of global optimization methods. The difficulties associated with determining the crystal structures of apparently simple inorganic materials from powder diffraction data are illustrated here with two recent hydride examples, $Li_4(BH_4)(NH_2)_3$ (Chater *et al.*, 2006) and $Mg(BH_4)_2$ (Cerny *et al.*, 2007; Her *et al.*, 2007). With careful sample preparation, it is possible to prepare single-phase $Li_4(BH_4)(NH_2)_3$ which is trivial to index to a body-centred cubic lattice with $a = 10.66445$ (1) Å. Space-group determination shows that, apart from the body centring, there are no additional systematic absences and the extinction symbol is I--- which immediately creates complications by introducing a

sixfold space-group ambiguity; $Im3m$, $I\bar{4}3m$, $I432$, $Im3$, $I2_13$ and $I23$ all conform to I---. Chemical reasoning reduces this to just $I2_13$ and $I23$ if the material is presumed to be ordered and to contain BH_4^- tetrahedral anions. Through similarities with $LiNH_2$, it is probable that the BH_4^- and NH_2^- groups are based on a face-centred cubic arrangement. However, both $I2_13$ and $I23$ are consistent with this supposition, the only difference between them being the ordering of BH_4^- and NH_2^- groups. The fact that BH_4^- and NH_2^- are isoelectronic means that both space groups give good fits to the X-ray diffraction data. A complete Rietveld analysis gives a slight preference for $I2_13$ but strong confirmation of this is only easily obtained from additional neutron powder diffraction measurements. The difference in neutron scattering lengths between N and B is pronounced and enables a clear discrimination in favour of $I2_13$, whilst also returning accurate H-atom positions. Importantly, the neutron sample was not isotopically enriched; developments in high-intensity neutron powder diffractometers mean that accurate and reliable data from hydrogenous samples may be obtained in a few hours. This experimental advance is also important for organic and pharmaceutical structures where the combined use of X-ray and neutron powder diffraction will bring a greater certainty to correctness of the crystal structure. Independently of this work, the crystal structure of $Li_4(BH_4)(NH_2)_3$ was determined using X-ray diffraction measurements of a small single crystal (Filinchuk *et al.*, 2006). The level of agreement between the independently derived structures is excellent.

On first consideration, it is reasonable to presume that $Mg(BH_4)_2$ should adopt a simple crystal structure, similar to closely related compounds with similar stoichiometries, *e.g.* $Be(BH_4)_2$ or perhaps $Mg(AlH_4)_2$ which is based on a CdI$_2$-type structure. Database mining and density functional theory (DFT) calculations are now very important approaches to suggesting possible crystal structures. For $Mg(BH_4)_2$, DFT calculations (Cerny *et al.*, 2007) of 28 basic possible structure types suggest a structure similar to Cd(AlCl$_4$)$_2$. However, no structure matched the unexpectedly large hexagonal $P6_1$ unit cell [$a = 10.3182$ (1), $c = 36.9983$ (5) Å and $V = 3411.3$ (1) Å^3]. The structure was finally solved using a combination of X-ray and neutron powder diffraction data using the global optimization program *FOX*. There are five Mg^{2+} and ten $(BH_4)^-$ symmetry-independent isoelectronic entities in the unit cell (Fig. 5). There is, however, an additional twist to the structure of $Mg(BH_4)_2$. Independently of the work of Cerny *et al.*, Her and co-workers (Her *et al.*, 2007) not only determined the hexagonal phase but also the high-temperature orthorhombic phase, which is stable above 453 K. This structure adopts *Fddd* symmetry with $a = 37.072$ (1), $b = 18.6476$ (6), $c = 10.9123$ (1) Å and $V = 7543.8$ (5) Å^3 and there are two formula units in the asymmetric unit. Moreover, the orthorhombic phase was identified to have significant disorder through the formation of antiphase domain walls. From the SDPD viewpoint, both groups completed the crystal structure and performed their Rietveld analyses using the same *TOPAS-Academic* software package. It is noteworthy, however, to point out that Cerny used real-space global optimization
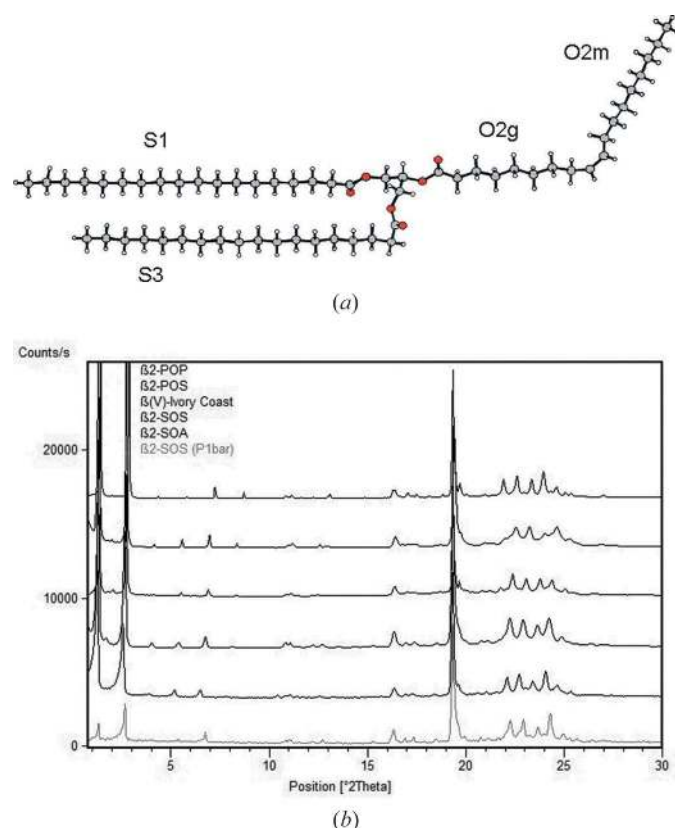

Figure 4
Triglyceride structure with representative XPRD data.

methods while Her employed a direct methods approach using the computer program *EXPO*. This underlines the fact that both real- and reciprocal-space methods are capable of tackling these very complex structures. The challenge is to make these algorithms more routine for structures of such unexpected complexity.

3.3. Biological crystal structures

Perhaps the ultimate challenge in structural solution from powders is in the area of macromolecular crystallography. One of the surprises over the past decade has been the quality of powder diffraction data that can be obtained at synchrotron sources. Following pioneering work (Von Dreele, 1999), there has been rapid progress in the development of this field. While the major emphasis in macromolecular powder diffraction is likely to be in the area of parametric studies of material properties at diffe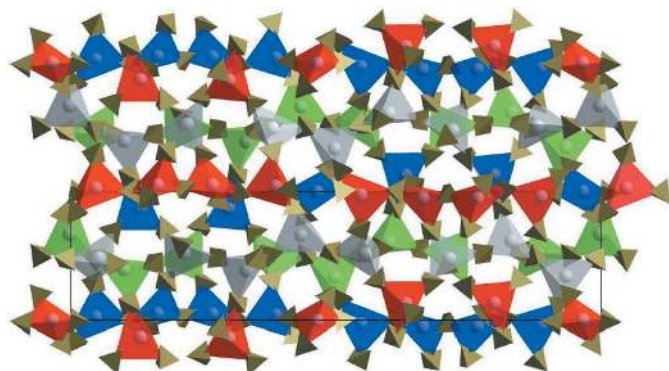rent temperatures and under different synthesis conditions, structure determination has been successfully attempted in a small number of cases. The first new protein crystal structure obtained from powder diffraction data was a study of a doubled-cell structure of insulin (Von Dreele *et al.*, 2000) which was subsequently confirmed by a single-crystal study (Smith *et al.*, 2001). More recently, variation in structure of hen egg white lysozyme has been investigated as a function of pH (Basso *et al.*, 2005); the structure was solved by molecular replacement and refined using multiple data sets which exploited the extra information available from differential lattice expansion. Perhaps the most significant success to date has been the successful determination of the second SH3 domain of ponsin from high-resolution synchrotron powder diffraction data (Margiolaki *et al.*, 2007), where the authors report the solution, model building and refinement of this 67-residue protein domain crystal structure which has a cell volume of 64 879 Å^3 (Fig. 6). This remarkable paper represents the most complex problem tackled to date using powder diffraction and suggests that, with improved algorithms and data-collection strategies, small protein structures may be regularly solved from powder diffraction data; the use of additional ancillary experimental information may further extend the power of this technique. A full discussion of powder diffraction studies in macromolecular research is to be found in the paper by Margiolaki & Wright (2008) in this issue.

Figure 5
Structure of the low-temperature $Mg(BH_4)_2$ phase in space group $P6_1$ viewed along the hexagonal *a* axis, showing two unit cells. The small opaque tetrahedra are BH_4 units; the larger (partially transparent) tetrahedra represent Mg and the four nearest B atoms. MgB_4 tetrahedra are coloured according to their projection along **a**; units centred near 0, 1/4, 1/2 and 3/4 are coloured red, green, blue and grey, respectively.

4. Conclusions

Structure determination from powder diffraction data has developed over the past decade from individual *tours de force* to a technique that is almost routine. Criticisms that were levelled a decade ago (*e.g.* that SDPD programs were not widely available and that methods only worked well with high-resolution synchrotron data) are no longer heard. The number of publications is increasing year on year, particularly in the area of global optimization methods as applied to organic
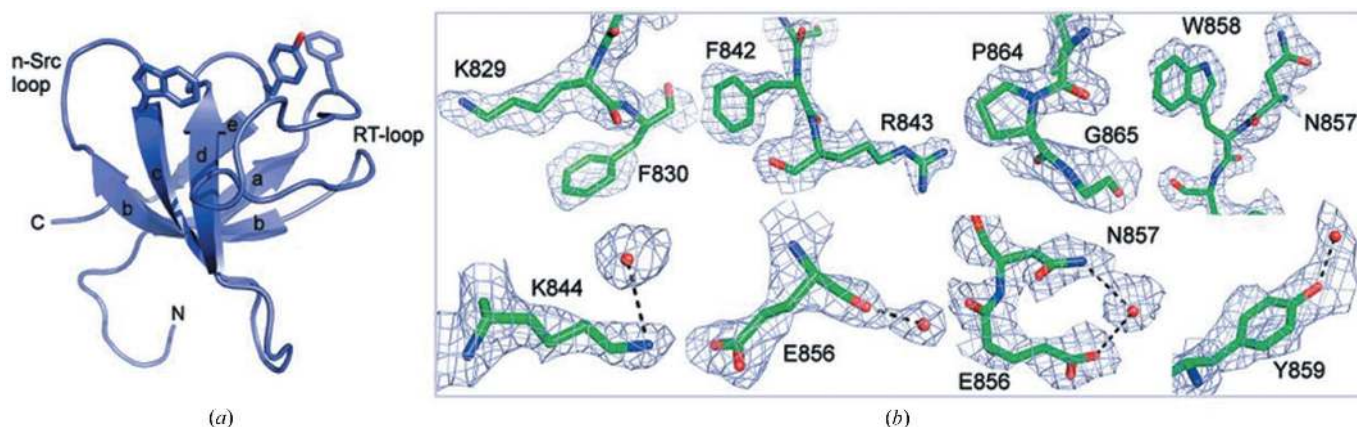

Figure 6
(*a*) Ribbon representation of SH3.2 indicating the secondary structure elements of the domain. The main hydrophobic residues of the binding interface as well as the positions of the n-Src and RT loops are indicated. (*b*) Selected regions of the final refined structural model in stick representation, and the corresponding total omit map contoured at 1 Å. This figure was generated using *PYMOL* (http://pymol.sourceforge.net/).

materials. Structure refinement programs, such as *Topas*, *FullProf* and *GSAS* have also developed to keep step with increasing structural complexity. Of course, with increased complexity must come increased vigilance, to ensure that published structures, especially those obtained by global optimization of conformationally complex fragments, meet the criteria of crystallographic and chemical sense.

A theoretical analysis (Appendix *A*) shows that there is an asymptotic limit to the number of independent groups of reflections that can be extracted from a powder diffraction pattern. With the reasonable assumption that the separation between individual independent observations follows a constant $R = \Delta d/d$ behaviour, then the maximum number of independent observations is simply $N_{max} \approx 1/3R$. With $R = 10^{-3}$ and 10^{-4} for 'standard' and 'best' diffractometer resolutions, this leads to respectively $N_{max} \approx 300$ and 3000. Assuming, perhaps conservatively, that an observation-to-parameter ratio of 10 is required for structure solution, then 30-parameter problems should be routinely tractable whilst 300-parameter problems represent the best that is likely to be attained. In practice, organic materials with 40–50 degrees of freedom are now being tackled. The ambitious target of more than 200 parameters brings the domain of small protein structures into consideration if rigid-body techniques are used. This represents a horizon that is still some way off.

To develop the full capability of SDPD, the correct handling of multiply overlapped reflections must be addressed. For a direct-methods approach, this will require full implementation of the correct statistical handling of reflection overlap. Global optimization methods, whether full-profile or correlated-integrated-intensities based, already tackle this overlap correctly – in this case, the issue that must be addressed is the searching for the global minimum in a hyperspace that grows exponentially with the number of parameters. The use of parallel computing and the development of new search algorithms (particularly deterministic ones) will help to realize the full potential of SDPD.

It is, as stated early in the book *Structure Determination from Powder Diffraction Data*, generally unwise to make predictions about how a particular research field will develop in the future. However, if the next decade is as productive as the last, there will be no shortage of pleasant surprises in store for the structural community.

APPENDIX *A*
On the loss of information in a powder diffraction pattern

The collapse of the three dimensions of reciprocal space to the single dimension of a powder diffraction results in a very substantial loss of information. For moderately sized crystal structures, even with the highest-resolution powder diffractometers, it is almost impossible to obtain individual integrated intensities at atomic resolution (<1.2 Å). Experimental techniques, such as the deliberate introduction of texture or the use of differential thermal expansion (or even differential

lattice expansion resulting from radiation damage), have been developed to help address this issue; a full set of uncorrelated integrated intensities may be obtained for structures with around 20–30 atoms in the asymmetric unit but larger structures prove much more intractable. In principle, Bragg-peak overlap may be treated as a hyperdimensional phase problem and results of simple crystal structures show great promise. However, this likelihood approach remains to be extensively developed.

To assess how severe Bragg overlap becomes at atomic resolution, consider the number of reflections, $\Delta N(d^*)$, within a shell of width Δd^* at a radius d^* in reciprocal space may be approximated to

$$\Delta N(d^*) \approx 2\pi V_A d^{*2} \Delta d^*, \tag{1}$$

where V_A is the volume of the asymmetric unit. In a single-crystal measurement, all these Bragg peaks will be resolved; for a powder diffraction experiment, they will be overlapped if the Bragg-peak widths, $W(d^*)$, from either instrumental or sample broadening are substantially larger than the shell width, Δd^*; a reasonable working assumption is to assume that Bragg peaks are separated if the peak separation, $\delta d^* > 0.2W(d^*) = w(d^*)$. For low-symmetry systems, the separation, δd^*, between neighbouring Bragg peaks is essentially random and can be shown to follow an exponential relationship[1]

$$\text{prob}(\delta d^*) = \Delta N(d^*) \exp[-\Delta N(d^*)\delta d^*]. \tag{2}$$

All neighbouring reflections with a separation, δd^*, larger than the resolvable separation $w(d^*)$ can be regarded as independent reflections. Given the exponential relationship, equation (2), the fraction of independent reflections is given by the equation $f(d^*) \approx \exp[-\Delta N(d^*)w(d^*)]$. The total number of independent reflection observations up to a resolution limit d^*_{max} in reciprocal space is thus given by

$$N_{ind}(d^*_{max}) \approx \int_0^{d^*_{max}} f(d^*)\Delta N(d^*)$$

$$= \int_0^{d^*_{max}} 2\pi V_A x^2 \exp[-2\pi V_A w(x)x^2]\,dx. \tag{3}$$

It is an excellent approximation for time-of-flight neutron powder diffraction and a reasonable one for X-ray synchrotron and laboratory diffractometers that Bragg peak widths will show a strain-like behaviour. If we choose this constant strain variation, $w(d^*) = Rd^*$, where $R = \Delta d^*/d^* = \Delta d/d$ is the Bragg-peak resolution, then equation (3) is easily integrated to give

$$N_{ind}(d^*_{max}) \approx [1 - \exp(-2\pi V_A R d^{*3}_{max})]/3R, \tag{4}$$

---

[1] Consider a peak at position d^*. If there are ΔN peaks per unit d^*, then the probability of finding a peak at $d^* + \delta d^*$ may be determined by dividing the separation δd^* into m equal segments ε ($\delta d^* = m\varepsilon$). The probability, $p(\delta d^*)\varepsilon$, of the peak being at $d^* + \delta d^*$ is the product of the probabilities of it not being at any of the intermediate segments and then being at $d^* + \delta d^*$. This is given by $p(\delta d^*)\varepsilon = (1 - \varepsilon\Delta N)m(\varepsilon\Delta N)$. Since $(1 - x/m)m \rightarrow \exp(-x)$ as $m \rightarrow \infty$, then $p(\delta d^*) = (\Delta N)(1 - \varepsilon\Delta N)m = (\Delta N)[1 - (\delta d^*\Delta N)/m]^m$ may be written as $p(\delta d^*) = (\Delta N)\exp(-\Delta N\delta d^*)$.

which is to be compared with the total number of independent reflections observed in a single-crystal measurement:

$$N_{tot}(d^*_{max}) \approx \int_0^{d^*_{max}} \Delta N(d^*) = \int_0^{d^*_{max}} 2\pi V_A x^2 \, dx = \tfrac{2}{3}\pi V_A d^{*3}_{max}. \quad (5)$$

The reduction in the number of independent reflections relative to the total number of reflections is illustrated in Fig. 7. Two different resolutions have been chosen, $R_1 = \Delta d^*/d^* = \Delta d/d = 10^{-4}$ and $R_2 = 10^{-3}$; R_1 represents the very best resolution achievable on the highest-resolution synchrotron and neutron instruments where sample broadening is minimal; R_2 represents a more typical value for good laboratory powder diffractometers where there may also be a small amount of sample broadening. Although around 3000 reflections may in principle be resolved down to 1 Å for small protein asymmetric unit volumes of 5000 Å^3 (Fig. 7a), this represents only a small fraction (<20%) of the total number of reflections (Fig. 7b). Given a volume of ~20 Å^3 for each atom,

the number of coordinates to be determined in a direct-methods approach is ~750, leading to a parameter/observation ratio of ~4/1. This seems perhaps tractable until it is realized that each of the independent observations corresponds on average to five overlapped reflections – with hyperphase determination methods, there is the exciting possibility that these problems may become tractable. However, these algorithms are not currently available for direct methods. Experimental approaches such as the deliberate introduction of texture and the use of differential thermal expansion can reduce the number of component reflections in a single independent observation; however, at atomic resolution even these methods cannot yield individual reflections which can be used in current single-crystal direct-methods programs. There is an additional pragmatic aspect of data collection which makes reaching atomic resolution difficult. For X-ray diffraction, the combination of Lorentz effects and form-factor fall-off means that Bragg peaks at 1 Å are approaching two orders of magnitude less intense than long d-spacing reflections.
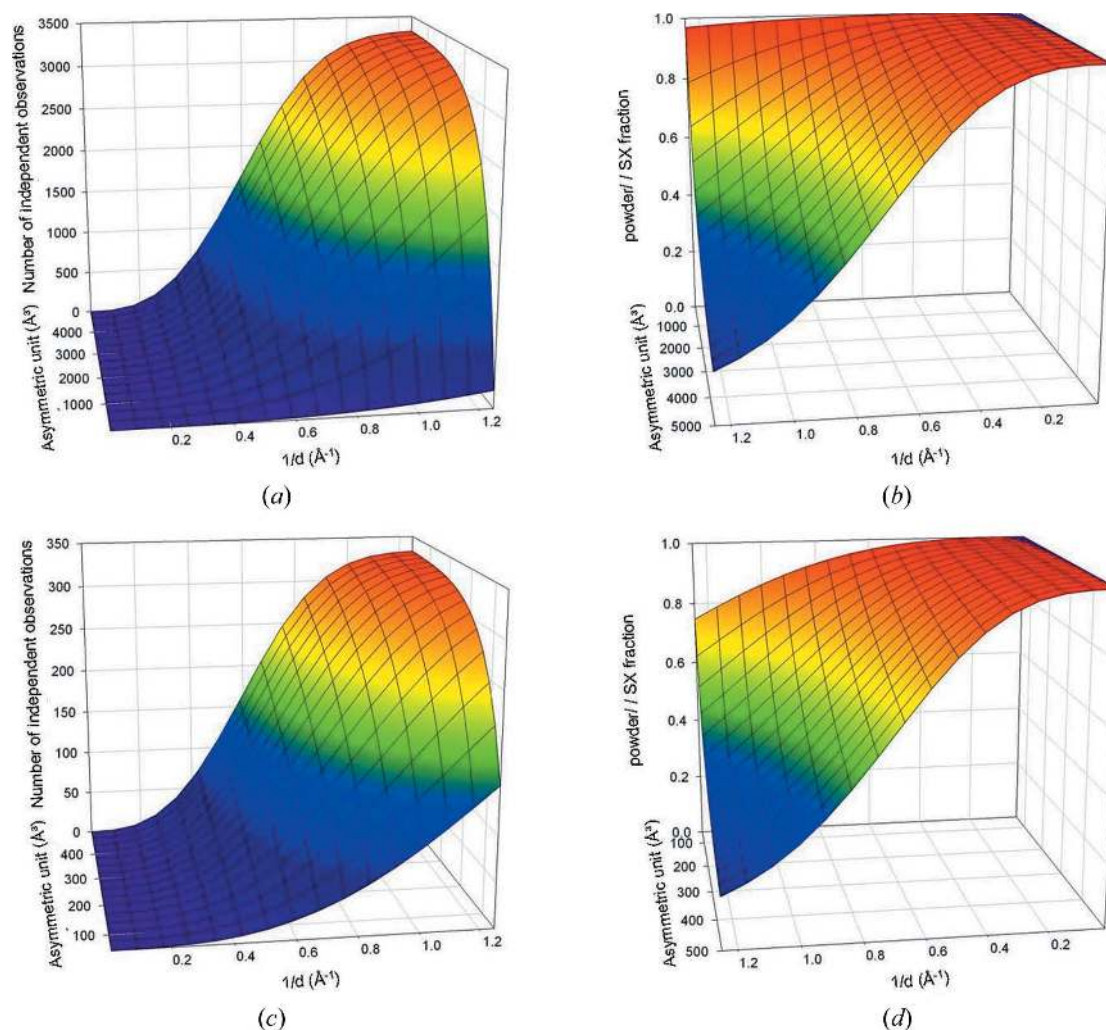

Figure 7
(a) The number of independent observations and (b) the ratio of powder diffraction *versus* single-crystal observations as a function of reciprocal d spacing (1/d) (Å^{-1}) and asymmetric unit-cell size (Å^3) for a $\Delta d/d$ resolution of 10^{-4}. (c) The number of independent observations and (d) the ratio of powder diffraction *versus* single-crystal observations as a function of reciprocal d spacing (1/d) (Å^{-1}) and asymmetric unit-cell size (Å^3) for a $\Delta d/d$ resolution of 10^{-3}. Note that an observation will generally consist of multiple Bragg reflections.

Larger solid angles and longer counting times are required to offset this intensity reduction. In practice, in the absence of significant sample broadening, 1 Å atomic resolution may be obtained for moderately sized asymmetric units (*ca* 700 Å^3); however, for small protein unit cells the best resolution obtainable is probably nearer 2 Å. From Fig. 7(*a*), this suggests that the maximum number of useful observations is around 1500–2000. Of course, global optimization methods that use full profile fitting or the equivalent correlated integrated intensities approach do not need to disentangle *a priori* the individual intensities. Working on the assumption that an observation/parameter ratio of 10 should enable structures to be solved if the appropriate algorithm exists, then, with appropriate parameterization, small protein structures may yet be solved from powder diffraction data. In many materials, however, the very highest resolutions are not attainable and $R_2 = \Delta d/d = 10^{-3}$ is more realistic. Fig. 7(*c*) shows the number of independent observations at this resolution. The highest values that can be expected at this resolution are around 300–350, suggesting that 30–35 parameters may be straightforwardly obtained. This is the experience of global optimization methods where the parameterization is in terms of external degrees of freedom and internal torsion angles. Fig. 7(*d*) again shows that the information loss in a powder measurement at atomic resolution is substantial compared to single-crystal measurements for moderately complex structures ($V_A > 300$ Å^3 corresponding to ~15 independent atoms in the asymmetric unit). This leads to the pragmatic experimental consideration that, if a small single crystal can be found, then it is best to perform a single-crystal experiment; an ancillary powder diffraction measurement is, of course, essential to verify that the single crystal is representative of the bulk.

References

Albesa-Jove, D., Kariuki, B. M., Kitchin, S. J., Grice, L., Cheung, E. Y. & Harris, K. D. M. (2004). *Chemphyschem*, **5**, 414–418.

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.

Altomare, A., Caliandro, R., Camalli, M., Cuocci, C., Giacovazzo, C., Moliterni, A. G. G. & Rizzi, R. (2004). *J. Appl. Cryst.* **37**, 1025–1028.

Altomare, A., Caliandro, R., Camalli, M., Cuocci, C., da Silva, I., Giacovazzo, C., Moliterni, A. G. G. & Spagna, R. (2004). *J. Appl. Cryst.* **37**, 957–966.

Altomare, A., Camalli, M., Cuocci, C., Giacovazzo, C., Moliterni, A. G. G. & Rizzi, R. (2007). *J. Appl. Cryst.* **40**, 344–348.

Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A. G. G. & Rizzi, R. (2006). *J. Appl. Cryst.* **39**, 145–150.

Andreev, Yu. G., Lightfoot, P. & Bruce, P. G. (1997). *J. Appl. Cryst.* **30**, 294–305.

Baerlocher, C., Gramm, F., Massuger, L., McCusker, L. B., He, Z. B., Hovmoller, S. & Zou, X. D. (2007). *Science.* **315**, 1113–1116.

Baerlocher, C., McCusker, L. B. & Palatinus, L. (2007). *Z. Kristallogr.* **222**, 47–53.

Basso, S., Fitch, A. N., Fox, G. C., Margiolaki, I. & Wright, J. P. (2005). *Acta Cryst.* D**61**, 1612–1625.

Boufas, S., Merazig, H., Moliterni, A. G. & Altomare, A. (2007). *Acta Cryst.* C**63**, m315–m317.

Brenner, S., McCusker, L. B. & Baerlocher, C. (1997). *J. Appl. Cryst.* **30**, 1167–1172.

Brenner, S., McCusker, L. B. & Baerlocher, C. (2002). *J. Appl. Cryst.* **35**, 243–252.

Bricogne, G. (1991). *Acta Cryst.* A**47**, 803–829.

Brodski, V., Peschar, R. & Schenk, H. (2005). *J. Appl. Cryst.* **38**, 688–693.

Brunelli, M., Neumann, M. A., Fitch, A. N. & Mora, A. J. (2007). *J. Appl. Cryst.* **40**, 702–709.

Cerny, R., Filinchuk, Y., Hagemann, H. & Yvon, K. (2007). *Angew. Chem. Int. Ed.* **46**, 5765–5767.

Chater, P. A., David, W. I. F., Johnson, S. R., Edwards, P. P. & Anderson, P. A. (2006). *Chem. Commun.* pp. 2439–2441.

Chong, S. Y. & Tremayne, M. (2006). *Chem. Commun.* pp. 4078–4080.

Coelho, A. A. (2000). *J. Appl. Cryst.* **33**, 899–908.

Coelho, A. A. (2003). *J. Appl. Cryst.* **36**, 86–95.

Corma, A., Rey, F., Valencia, S., Jorda, J. L. & Rius, J. (2003). *Nature Materials*, **2**, 493–497.

Csoka, T. & David, W. I. F. (1999) *Acta Cryst.* A**55**, Supplement, Abstract No. P08.03.012.

David, W. I. F., Shankland, K., McCusker, L. B. & Baerlocher, Ch. (2002). *Structure Determination from Powder Diffraction Data*, edited by W. I. F. David, K. Shankland, L. B. McCusker & Ch. Baerlocher, pp. 1–11. Oxford University Press.

David, W. I. F., Shankland, K. & Shankland, N. (1998). *Chem. Commun.* pp. 931–932.

David, W. I. F., Shankland, K., van de Streek, J., Pidcock, E., Motherwell, W. D. S. & Cole, J. C. (2006). *J. Appl. Cryst.* **39**, 910–915.

Docherty, A. (2004). PhD thesis, University of Strathclyde, Glasgow, Scotland.

Earl, D. J. & Deem, M. W. (2005). *Phys. Chem. Chem. Phys.* **7**, 3910–3916.

Engel, G. E., Wilke, S., König, O., Harris, K. D. M. & Leusen, F. J. J. (1999). *J. Appl. Cryst.* **32**, 1169–1179.

Estermann, M. A. & David, W. I. F. (2002). *Structure Determination from Powder Diffraction Data*, edited by W. I. F. David, K. Shankland, L. B. McCusker & Ch. Baerlocher, pp. 202–218. Oxford University Press.

Favre-Nicolin, V. & Cerny, R. (2004). *Z. Kristallogr.* **219**, 847–856.

Feng, Z. J. & Dong, C. (2007). *J. Appl. Cryst.* **40**, 583–588.

Fernandes, P. (2006). PhD thesis, University of Strathclyde, Glasgow, Scotland.

Fernandes, P., Shankland, K., Florence, A. J., Shankland, N. & Johnston, A. (2007). *J. Pharm. Sci.* **96**, 1192–1202.

Filinchuk, Y. E., Yvon, K., Meisner, G. P., Pinkerton, F. E. & Balogh, M. P. (2006). *Inorg. Chem.* **45**, 1433–1435.

Fukuda, K., Ito, M. & Iwata, T. (2007). *J. Solid State Chem.* **180**, 2305–2309.

Habershon, S., Harris, K. D. M. & Johnston, R. L. (2003). *J. Comput. Chem.* **24**, 1766–1774.

Harris, K. D. M., Habershon, S., Cheung, E. Y. & Johnston, R. L. (2004). *Z. Kristallogr.* **219**, 838–846.

feature articles

Harris, K. D. M., Tremayne, M., Lightfoot, P. & Bruce, P. G. (1994). *J. Am. Chem. Soc.* **116**, 3543–3547.

Her, J.-H., Stephens, P. W., Gao, Y., Soloveichik, G. L., Rijssenbeek, J., Andrus, M. & Zhao, J.-C. (2007). *Acta Cryst.* B**63**, 561–568.

Ivashevskaja, S. N., Aleshina, L. A., Andreev, V. P., Nizhnik, Y. P., Chernyshev, V. V. & Schenk, H. (2003). *Acta Cryst.* E**59**, o1006–o1008.

Johnston, A., Florence, A. J., Shankland, K., Markvardsen, A., Shankland, N., Steele, G. & Cosgrove, S. D. (2004). *Acta Cryst.* E**60**, o1751–o1753.

Johnston, J. C., David, W. I. F., Markvardsen, A. J. & Shankland, K. (2002). *Acta Cryst.* A**58**, 441–447.

Lanning, O. J., Habershon, S., Harris, K. D. M., Johnston, R. L., Kariuki, B. M., Tedesco, E. & Turner, G. W. (2000). *Chem. Phys. Lett.* **317**, 296–303.

Louer, D. & Boultif, A. (2006). *Z. Kristallogr.* Suppl. **23**, 225–230.

Markvardsen, A. J., David, W. I. F., Johnson, J. C. & Shankland, K. (2001). *Acta Cryst.* A**57**, 47–54.

Markvardsen, A. J., David, W. I. F. & Shankland, K. (2002). *Acta Cryst.* A**58**, 316–326.

Markvardsen, A. J., Shankland, K., David, W. I. F. & Didlick, G. (2005). *J. Appl. Cryst.* **38**, 107–111.

Margiolaki, I. & Wright, J. P. (2008). *Acta Cryst.* A**64**, 169–180.

Margiolaki, I., Wright, J. P., Fitch, A. N., Wilmanns, M. & Pinotsis, N. (2007). *J. Am. Chem. Soc.* **129**, 11865–11871.

Masciocchi, N. & Sironi, A. (2005). *C. R. Chim.* **8**, 1617–1630.

Mazina, O. S., Rybakov, V. B., Chernyshev, V. V., Babaev, E. V. & Aslanov, L. A. (2004). *Crystallogr. Rep.* **49**, 998–1009.

Mechelen, J. B. van, Peschar, R. & Schenk, H. (2006a). *Acta Cryst.* B**62**, 1121–1130.

Mechelen, J. B. van, Peschar, R. & Schenk, H. (2006b). *Acta Cryst.* B**62**, 1131–1138.

Middleton, D. A., Peng, X., Saunders, D., Shankland, K., David, W. I. F. & Markvardsen, A. J. (2002). *Chem. Commun.* pp. 1976–1977.

Neumann, M. A. (2003). *J. Appl. Cryst.* **36**, 356–365.

Oszlányi, G. & Sütő, A. (2008). *Acta Cryst.* A**64**, 123–134.

Pagola, S. & Stephens, P. W. (2000). *Mater. Sci. Forum*, **321–3**, 40–45.

Pan, Z. G., Xu, M. C., Cheung, E. Y., Harris, K. D. M., Constable, E. C. & Housecroft, C. E. (2006). *J. Phys. Chem. B*, **110**, 11620–11623.

Putz, H., Schön, J. C. & Jansen, M. (1999). *J. Appl. Cryst.* **32**, 864–870.

Ramprasad, D., Pez, G. P., Toby, B. H., Markley, T. J. & Pearlstein, R. M. (1995). *J. Am. Chem. Soc.* **117**, 10694–10701.

Rius, J., Crespi, A. & Torrelles, X. (2007). *Acta Cryst.* A**63**, 131–134.

Rius, J. & Miravitlles, C. (1988). *J. Appl. Cryst.* **21**, 224–227.

Rius, J., Torrelles, X., Miravitlles, C., Ochando, L. E., Reventós, M. M. & Amigó, J. M. (2000). *J. Appl. Cryst.* **33**, 1208–1211.

Shankland, K. & David, W. I. F. (2002). *Structure Determination from Powder Diffraction Data*, edited by W. I. F. David, K. Shankland, L. B. McCusker & Ch. Baerlocher, pp. 252–283. Oxford University Press.

Shankland, K., David, W. I. F. & Csoka, T. (1997). *Z. Kristallogr.* **212**, 550–552.

Shankland, K., David, W. I. F. & Sivia, D. S. (1997). *J. Mater. Chem.* **7**, 569–572.

Smith, G. D., Pangborn, W. & Blessing, R. H. (2001). *Acta Cryst.* D**57**, 1091–1100.

Terent'ev, A. O., Platonov, M. M., Sonneveld, E. J., Peschar, R., Chernyshev, V. V., Starikova, Z. A. & Nikishin, G. I. (2007). *J. Org. Chem.* **72**, 7237–7243.

Turner, G. W., Tedesco, E., Harris, K. D. M., Johnston, R. L. & Kariuki, B. M. (2000). *Chem. Phys. Lett.* **321**, 183–190.

Von Dreele, R. B. (1999). *J. Appl. Cryst.* **32**, 1084–1089.

Von Dreele, R. B., Stephens, P. W., Smith, G. D. & Blessing, R. H. (2000). *Acta Cryst.* D**56**, 1549–1553.

Wessels, T., Baerlocher, C. & McCusker, L. B. (1999). *Science*, **284**, 477–479.