

Powder crystallography on macromolecules

I. Margiolaki* and J. P. Wright*

European Synchrotron Radiation Facility (ESRF), Grenoble, France. Correspondence e-mail: margiolaki@esrf.fr, wright@esrf.fr

Following the seminal work of Von Dreele, powder X-ray diffraction studies on proteins are being established as a valuable complementary technique to single-crystal measurements. A wide range of small proteins have been found to give synchrotron powder diffraction profiles where the peak widths are essentially limited only by the instrumental resolution. The rich information contained in these profiles, combined with developments in data analysis, has stimulated research and development to apply the powder technique to microcrystalline protein samples. In the present work, progress in using powder diffraction for macromolecular crystallography is reported.

© 2008 International Union of Crystallography
Printed in Singapore – all rights reserved

1. Introduction

In recent years, modern powder diffraction techniques have been applied to several microcrystalline proteins. The use of high-resolution synchrotron data, together with new analysis procedures, has stimulated exciting progress which is showing that powders can offer unique opportunities for the structural characterization of proteins and are complementary to existing methods. The developments in experimental methods and instrumentation have been absolutely essential, but they have been discussed recently in another article (Margiolaki, Wright, Fitch *et al.*, 2007) so that we have tried to avoid reproducing that information here.

Our own research backgrounds have been mainly in powder diffraction and it has been fascinating to move towards biological crystallography. Simple terminology like ‘high-resolution’ has a quite different meaning for the different communities. For powders, the term applies to reciprocal space, meaning ‘very sharp peaks’ but, for macromolecular crystallography, the term applies in direct space, meaning ‘atomic resolution’ electron-density maps. Similarly, a ‘phase’ in a powder diagram refers to a single component in a mixture while, for a single crystal, it refers to the phase angle needed to produce a map. The apparent divergence of these areas of research means there are now many opportunities to transfer ideas from one to the other.

In this article, we begin in §2 with the motivation for using powder methods and highlight some of the powder studies prior to Von Dreele’s application of the Rietveld method (Rietveld, 1969) to proteins. We then describe in §3 some of the structural refinements which have been carried out for powders. Recently, there has been an evolution from single to multiple pattern fitting, leading to improvements in the data and resulting models, and we describe these in §4. Complementary information from powder data, such as the microstructure and application to phase identification in mixtures,

are described in §5. Perhaps the most enticing goal for any crystallographer is to solve new structures, and §6 discusses the progress and prospects for powder data in this area.

2. Early powder diffraction studies of biological macromolecules

Powder methods were used during some of the earliest X-ray diffraction studies of macromolecules. Corey & Wyckoff (1936) investigated several different crystalline proteins using powders, and established that the structures contain long lattice spacings. Many of the plates they presented would now be termed as fibre diffraction, although for chymotrypsinogen ‘numerous powder rings can be seen’. Powder rings were also observed for precipitated tobacco mosaic virus proteins (Wyckoff & Corey, 1936) who wrote that ‘The patterns thus obtained, with many sharp reflections between 80 Å and 3 Å are exactly those to be expected from true crystals composed of large molecules’. Further studies of other plant virus proteins (Bernal & Fankuchen, 1941) allowed unit cells to be proposed in several cases.

Powder diffraction studies have also been reported by Amos *et al.* (1984) for two forms of the protein tubulin. Large hexagonal unit cells were proposed for tubulin crystals from sea urchin egg ($a = 371$, $c = 149$ Å) and bovine brain ($a = 373$, $c = 255$ Å). Powder methods were also used in a study of the crystalline inclusion bodies of cytoplasmic polyhedrosis virus from *Bombyx mori* (Xiaa *et al.*, 1991). Inclusion bodies of about 1–3 µm in linear size were compacted in a capillary tube while immersed in buffer and X-ray diffraction photographs showed powder rings, extending to 8.2 Å resolution. Anduleit *et al.* (2005) used powder diffraction to identify the crystalline lattice as a biological phenotype for insect viruses. They found a body-centred cubic lattice with almost identical unit cell for

polyhedrin proteins from different viral families without any significant amino-acid-sequence similarity.

Coulibaly *et al.* (2007) have shown that the polyhedrin of a nucleopolyhedrovirus (NPV) also assembles in this body-centred cubic lattice when they solved the structure using single-crystal microdiffraction. There appears to be a strong demand for crystal structures from ever smaller protein crystals. However, reductions in crystal size below $\sim 20\ \mu\text{m}$ for a $100\ \text{\AA}$ unit cell are not foreseen in the near future due to problems of radiation damage (Sliz *et al.*, 2003). A future challenge for crystallography will be to solve and refine structures from sub- μm sized protein crystals containing only a few unit cells. Powder methods offer a way to greatly increase the number of unit cells in the X-ray beam for microcrystalline samples, with the cost of overlapping Bragg reflections. Provided the overlap problem can somehow be overcome, powders offer a complementary route to structure analysis when there is a problem to obtain larger crystals. For example, the polyhedrin viral proteins crystallize *in vivo* and are then insoluble, making crystal growth a particular challenge.

3. Rietveld fits to high-resolution synchrotron powder data

In 1999, the first protein structure refinement using powder data was reported (Von Dreele, 1999) for metmyoglobin. High-angular resolution synchrotron data from beamline X3B1 ($\lambda = 1.14991\ \text{\AA}$) at the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory, USA, were used for the refinement (Fig. 1). A key characteristic of the powder profile was the observation of very sharp reflections associated with the highly crystalline protein sample. By combining a suite of stereochemical restraints with the measured diffraction profile, a successful refinement of the

atomic positions of the 1260-atom protein was achieved (pdb code 1f6h). In order to carry out this refinement, Von Dreele upgraded the *GSAS* (Larson & Von Dreele, 2004) software package in several ways. A suite of stereochemical restraints, as are normally used in low-resolution refinements with single-crystal data, were implemented with automatic recognition of atom and bond types for the standard amino acid residues. A novel restraint was introduced to describe the two-dimensional pseudo-potential surface of a Ramachandran plot (Ramachandran *et al.*, 1963). Also, a Babinet's principle solvent correction was used to account for the disordered solvent within the crystal structure. The excellent Rietveld fit obtained was a breakthrough, showing for the first time that a crystal structure refinement could be carried out for a protein molecule from powder data.

The following year, the structure of a new variant of the T_3R_3 human insulin-Zn complex, produced by mechanical grinding of a polycrystalline sample, was solved and refined from a powder (Von Dreele *et al.*, 2000). High-resolution synchrotron X-ray data (also from X3B1) were used to solve this crystal structure by a model-building approach. The strategy followed for structure solution consisted of the following steps.

1. Indexing of the powder diagram which indicated the *c*-axis doubling.
2. Development of a starting structural model from single-crystal coordinates for the T_3R_3 complex (Ciszak & Smith, 1994).
3. Positioning of two T_3R_3 complexes in the doubled *c*-axis unit cell, with one located nominally at the unit-cell origin and the other at $z = \frac{1}{2}$.
4. Initial rotations about the *c* axis of -15 and $+15^\circ$ for the two groups, respectively.
5. A three-parameter (two rotation angles and one translation) rigid-body Rietveld refinement.

A complete Rietveld refinement of the 1630-atom protein was then achieved by combining 7981 stereochemical restraints with the measured powder diffraction pattern ($d_{\text{min}} = 3.24\ \text{\AA}$, pdb code 1fu2). The fit is shown in Fig. 2, along with the fit for the previously known T_3R_3 variant and in Fig. 3 there is a view of the model projected down the *c* axis. It was determined that the reversible grinding-induced phase change is accompanied by 9.5 and 17.2° rotations of the two T_3R_3 complexes that comprise the crystal structure. These results were later born out by a single-crystal study (Smith *et al.*, 2001) at cryogenic temperatures (100 K). The mean C^α displacement between the powder and single-crystal structures was $1.36\ \text{\AA}$, which is insignificant given the relatively low resolution of the powder structure and that the powder data were collected at room temperature. Since there was no mechanical grinding prior to the single-crystal experiment, the origin of the phase change was proposed to be due to pressure arising from the contraction of the cryoprotecting drop.

The possibility of using powder diffraction data on microcrystalline protein samples for detection of ligands in protein-ligand complexes was illustrated by the studies of HEWL complexes with *N*-acetylglucosamine (NAG n , $n = 1-6$) oligo-

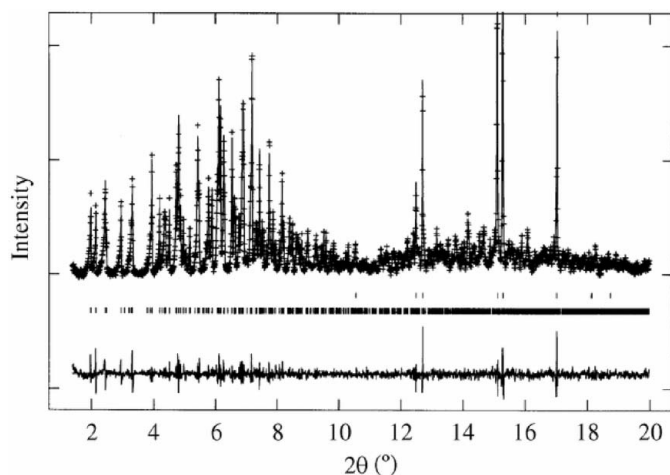


Figure 1

X-ray powder diffraction profile from the final Rietveld refinement of whale metmyoglobin. Observed intensities are shown as crosses (+), calculated and difference curves as lines, and reflection positions for $(\text{NH}_4)_2\text{SO}_4$ and protein are shown as tick marks (|). The background intensity found in the refinement has been subtracted from the observed and calculated intensities for clarity. [Reproduction of Fig. 4 from Von Dreele (1999).]

saccharides (Von Dreele, 2001, 2005). All of the protein powder diffraction patterns in these studies showed extremely sharp Bragg peaks consistent with crystallites greater than 1 μm in size, which were also devoid of line-broadening defects. The location of each NAG n could be found from difference Fourier maps generated from structure factors extracted during preliminary Rietveld refinements. Full NAG n -protein structures were subjected to combined Rietveld and stereochemical restraint refinements and revealed binding modes for NAG n that depended on the length of the NAG oligosaccharide (Fig. 4). The series of refined powder structures are available in the pdb database *via* access codes 1ja2, 1ja4, 1ja6, 1ja7, 1sf4, 1sf6, 1sf7, 1sfb and 1sfg. These studies showed that powder diagrams can give useful electron-density maps for protein samples and show unambiguously that a powder sample can be sufficient for structural analysis.

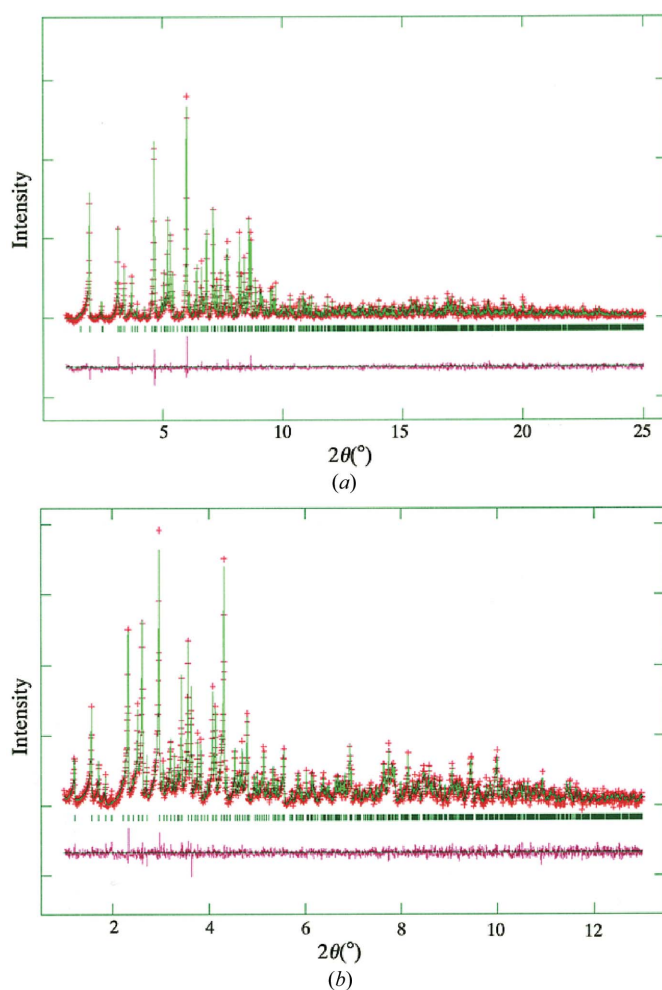


Figure 2

High-resolution X-ray powder diffraction profiles from the final Rietveld refinement of the doubled-unit-cell T_3R_3 complex [(a) with $\lambda = 1.401107 \text{ \AA}$] and also the normal T_3R_3 complex [(b) with $\lambda = 0.700233 \text{ \AA}$]. Observed intensities are shown in red (+), calculated and difference curves as green and magenta lines and the reflection positions are shown in black (|). The background intensity found in the refinement has been subtracted from the observed and calculated intensities for clarity. [Reproduction of Fig. 2 in Von Dreele *et al.* (2000).]

Following the tradition in the powder diffraction community, Von Dreele has made the *GSAS* software package freely available, including the most recent features for refinements of protein structures. We were able benefit from this generous philanthropy to use the package in our laboratory to carry out a Rietveld refinement of the structure of turkey egg white lysozyme (Margiolaki *et al.*, 2005). By carefully balancing the restraint weights with a highly resolved powder diagram collected at beamline ID31 at the ESRF, we were able to produce an excellent fit and refined model (pdb code: 1xft). In the course of refining this structure using *GSAS*, we also became convinced that powder methods can indeed be applied to such large structures.

4. Improvements in data quality *via* the use of multiple data sets

Powder data suffer from peak overlap. If this problem can somehow be circumvented, or at least reduced, we should be able to solve and refine larger structures more easily using powder methods. Changing the pattern of peak overlaps and collecting multiple data sets is one strategy to reduce the overlap problem. Varying the temperature of the sample for low-symmetry unit cells often leads to anisotropic thermal expansion, and therefore allows overlapped peaks to be untangled (Shankland *et al.*, 1997; Brunelli *et al.*, 2003). Similarly, if a sample can be prepared with a significant degree of preferred orientation, then overlapped peaks will have different intensities depending on the orientation of the sample (Dahms & Bunge, 1986; Bunge *et al.*, 1989; Cerny, 1996; Baerlocher *et al.*, 2004). Combining multiple data sets together, where either the cell parameters or the preferred orientation are different, gives some information about the way in which the overlapped peaks should be partitioned.

In order to benefit from anisotropic changes in unit-cell parameters for intensity extractions, the *PRODD* refinement program (Wright & Forsyth, 2000; Wright, 2004) was modified to allow a different set of unit-cell parameters to be used for each pattern in a multi-pattern Pawley fit (Pawley, 1981). A

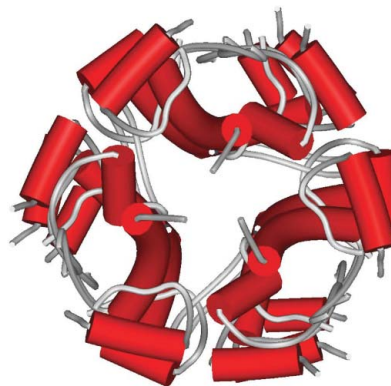


Figure 3

View of doubled-unit-cell T_3R_3 complex down the c axis where the 8° rotation of the complexes is visible. [Figure kindly provided by R. B. Von Dreele.]

single list of Bragg-peak intensities are fitted to the profiles while an overall isotropic thermal factor and scale factor for each profile accounts for the bulk of the differences between the patterns. A Pawley refinement of myoglobin using high-resolution synchrotron data ($\lambda = 0.413465 \text{ \AA}$, BM16, ESRF) and the *PRODD* software illustrated the applicability of the method in cases of significant complexity and severe peak overlap (Wright, 2004). The integrated intensities that are

extracted from the multi-pattern powder data are a considerable improvement over those that come from a single data set (Besnard *et al.*, 2007; Margiolaki, Wright, Wilmanns *et al.*, 2007) and are then useful for attempts at structure solution using single-crystal software packages. Care must be taken to ensure that the reflection intensities from the powder diagrams are corrected properly for Lorentz and multiplicity factors and also that sensible values for sigmas are supplied.

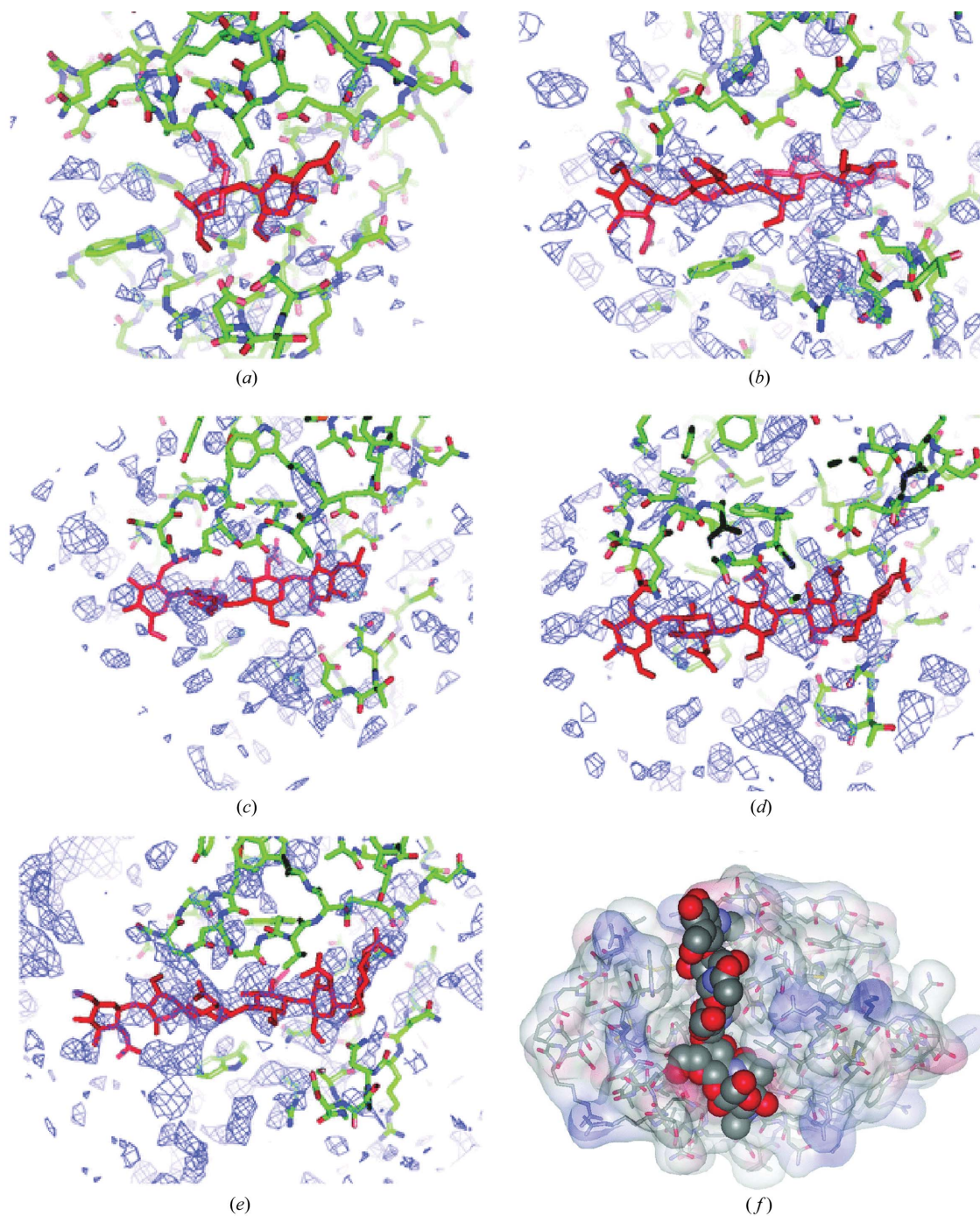


Figure 4

Views of difference electron density in the vicinity of the NAG_n -binding sites on HEWL. (a) NAG_2 on HEWL at $\sigma = 2$. (b) NAG_3 on HEWL at $\sigma = 1.5$. The end left and right saccharide rings are partially filled (35 and 65%, respectively). (c) NAG_4 on HEWL at $\sigma = 1.5$. (d) NAG_5 on HEWL at $\sigma = 1.5$. (e) NAG_6 on HEWL at $\sigma = 1.5$. (f) A view of the A–E binding of NAG_5 to HEWL. These figures were prepared with *PyMOL* (DeLano, 2004). [Reproduction of Figs. 1 and 7 in Von Dreele (2005).]

Perutz (1954) used perturbations of the unit-cell parameters with single-crystal data in an early attempt to solve the phase problem by oversampling of reciprocal space. This approach was apparently abandoned due to the development of other,

simpler, phasing methods. Thus far, we have not attempted to exploit the intensity differences which come with the unit-cell changes for phasing, although in theory the over-sampling method may still be applicable.

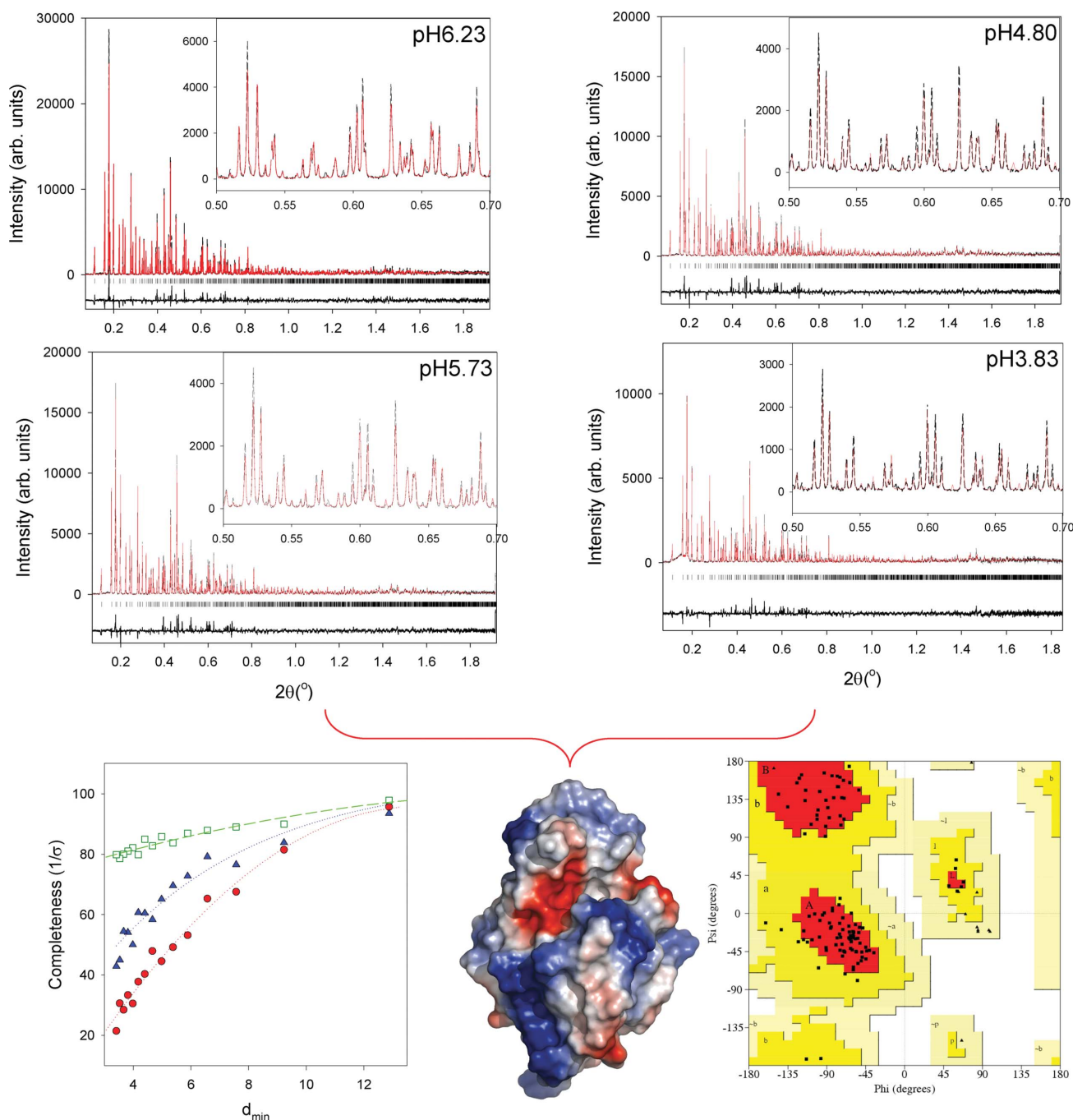


Figure 5

Upper and middle rows: The multi-data-set Rietveld refinement for HEWL samples at pH 6.23, pH 5.73, pH 4.80 and pH 3.83 and crystallized at 277 K. The dashed black, red and lower black lines represent the experimental, calculated and difference curves, respectively. The vertical bars mark Bragg-reflection positions and the insets show magnifications of a short region of the data. Background intensity has been subtracted for clarity. Lower left: Effective completeness for the powder diffraction data at a 1σ level (Basso *et al.*, 2005) where red circles represent a single powder pattern, blue triangles are the combined fit to four patterns and the green squares show the maximum completeness attainable in this case, which is less than 100% because of systematic overlaps due to the tetragonal symmetry (and ignoring Friedel pairs which are assumed to be equivalent). The dotted lines are guides to the eye. Lower middle: Electrostatic potential representation of the refined conformation of tetragonal HEWL, in total seven helices and three strands were observed, this figure was generated using *PYMOL* (DeLano, 2004). Lower right: Ramachandran plot for main-chain torsion angles (ϕ , ψ). Glycine residues are represented by triangles. Of the 113 non-proline, non-glycine residues in HEWL, 94 are in the most favoured regions (A, B, L, red); the rest (19) are in the additionally allowed regions (a, b, l, p, yellow).

Structure refinements can also be improved by the use of multiple data sets, provided there are not significant changes in structure factors between the different data sets. Two independent investigations of hen egg white lysozyme (HEWL) showed that the use of multiple patterns associated with slightly different lattice parameters leads to enhanced extracted intensities and more robust structure refinements (Basso *et al.*, 2005; Von Dreele, 2007) using a modified version of *GSAS* which could account for small lattice strains in different patterns (Larson & Von Dreele 2004). In our laboratory, we prepared a total of 44 different polycrystalline HEWL precipitates at 277 K and room temperature and in the pH range between 6.56 and 3.33 (Basso *et al.*, 2005). High-resolution powder diffraction data were collected at room temperature ($\lambda = 1.249826 \text{ \AA}$, ID31, ESRF). The anisotropic effect of pH of crystallization on the lattice dimensions of tetragonal HEWL shifts the peaks significantly and alleviates the peak-overlap problem. In order to try to quantify the improvements coming from the multiple data set approach, we selected four patterns (shown in Fig. 5 with the final Rietveld fit) across the range of pH for use in a multi-pattern Pawley refinement. Reflections with a reasonable signal-to-noise ratio could be observed visually in all data sets up to a maximum resolution of about 3.27 \AA . We tried to find a way to quantify the resolution limit of these multi-pattern data sets by following the methodology described by Sivia (2000). The least-squares matrix from a multi-pattern Pawley (1981) refinement characterizes the data set and the eigenvalue spectrum of this matrix is related to the effective error bar on the extracted intensities (or linear combinations of intensities when peaks are overlapped). To provide easier comparison with single-crystal data-processing statistics, we compute the effective completeness for the combined data set as the fraction of 'peaks' having $I/\sigma(I)$ greater than some threshold. Fig. 5 (lower left) shows the significant improvement of the effective completeness as a result of the use of multiple profiles during intensity extraction. Von Dreele (2007) has proposed a computationally simpler scheme which also attempts to quantify the improvements arising from the use of multiple data sets, and can also be applied to Rietveld refinements.

The same four histograms as used for the completeness analysis were employed in a multi-data-set Rietveld refinement (Rietveld, 1969) in order to extract an average structural model for tetragonal HEWL using the *GSAS* software (Larson & Von Dreele, 2004). Some differences between diffraction patterns could be accounted for by allowing the solvent-scattering coefficients to differ for each one of the four histograms following the observation that the largest observed differences in peak intensities were mainly at low angles. A more detailed description of this refinement can be found elsewhere (Basso *et al.*, 2005) and the model is deposited in the *pdb* database with access code 2a6u. The upper panels of Fig. 5 show the excellent fit that could be achieved. We found that this approach resulted in a much smoother and more robust refinement than previously experienced with single-pattern fits and produced a structural model with excellent stereo-

chemistry (the resulting Ramachandran plot is shown in the lower right panel of Fig. 5).

Von Dreele has also applied a multi-pattern approach to HEWL but using data sets obtained from an image plate and approaching single-crystal resolution limits ($d_{\min} \approx 2 \text{ \AA}$). The patterns differ due to changes in salt concentration and radiation-induced lattice strains. Anisotropic changes in the unit cell are taken into account in all of these refinements using a special profile function implemented in *GSAS* (No. 5). Using this function, only one set of unit-cell parameters is refined and those corresponding to the other patterns are related *via* a strain ($\Delta d/d$) of the reciprocal metric tensor elements. By selecting different combinations of data sets, Von Dreele refined three structural models and was able to add water molecules in each case (*pdb* codes 2hs7, 2hs9 and 2hso). Many of these water molecules are conserved between the three models and have also been found in single-crystal structures. The stereochemistries of the resulting multi-data-set models show significant improvements in comparison to previous single-pattern powder refinements.

In one of our most recent studies of the second SH3 domain of ponsin (Margiolaki, Wright, Wilmanns *et al.*, 2007), we exploited radiation-induced anisotropic lattice strains in a specially modified multi-pattern Pawley refinement taking into consideration likelihood criteria for partitioning overlapped peaks (Wright, Markvardsen & Margiolaki, 2007). The outcome of this analysis was a set of improved extracted intensities, in excellent agreement with single-crystal intensities later obtained by an independent experiment. The powder extracted data were sufficient for the structure solution of the domain *via* the molecular replacement method (see Fig. 8). Maximum-likelihood refinement improved the phases to a level where we could trace the main-chain alterations, build additional residues where needed and eventually place the correct side chains along the sequence; a substantial result from powder diffraction data. The protein conformation was refined in a multiple-data-set stereochemically restrained Rietveld analysis taking advantage of sample-induced anisotropic lattice strains. We further benefited from an approach combining multiple-data-set Rietveld analysis and periodical *OMIT* map (Bhat, 1988) computation, to reduce the bias of the final model, extending the resolution limits to levels comparable to single-crystal measurements, and even detecting several water molecules bound to the protein (Margiolaki, Wright, Wilmanns *et al.*, 2007). The benefits of the multi-pattern approach are perhaps most striking from the visual comparison of the single-pattern maps in Fig. 4 with the multi-pattern maps in Fig. 8.

5. Traditional strengths of the powder technique

Powder diffraction has been exploited as a technique for fingerprinting different crystalline substances and for rapid phase identification and quantification of mixtures (see *e.g.* Faber & Fawcett, 2002). High-resolution powder instrumentation is designed to obtain the sharpest possible peaks in reciprocal space and also to determine accurate peak

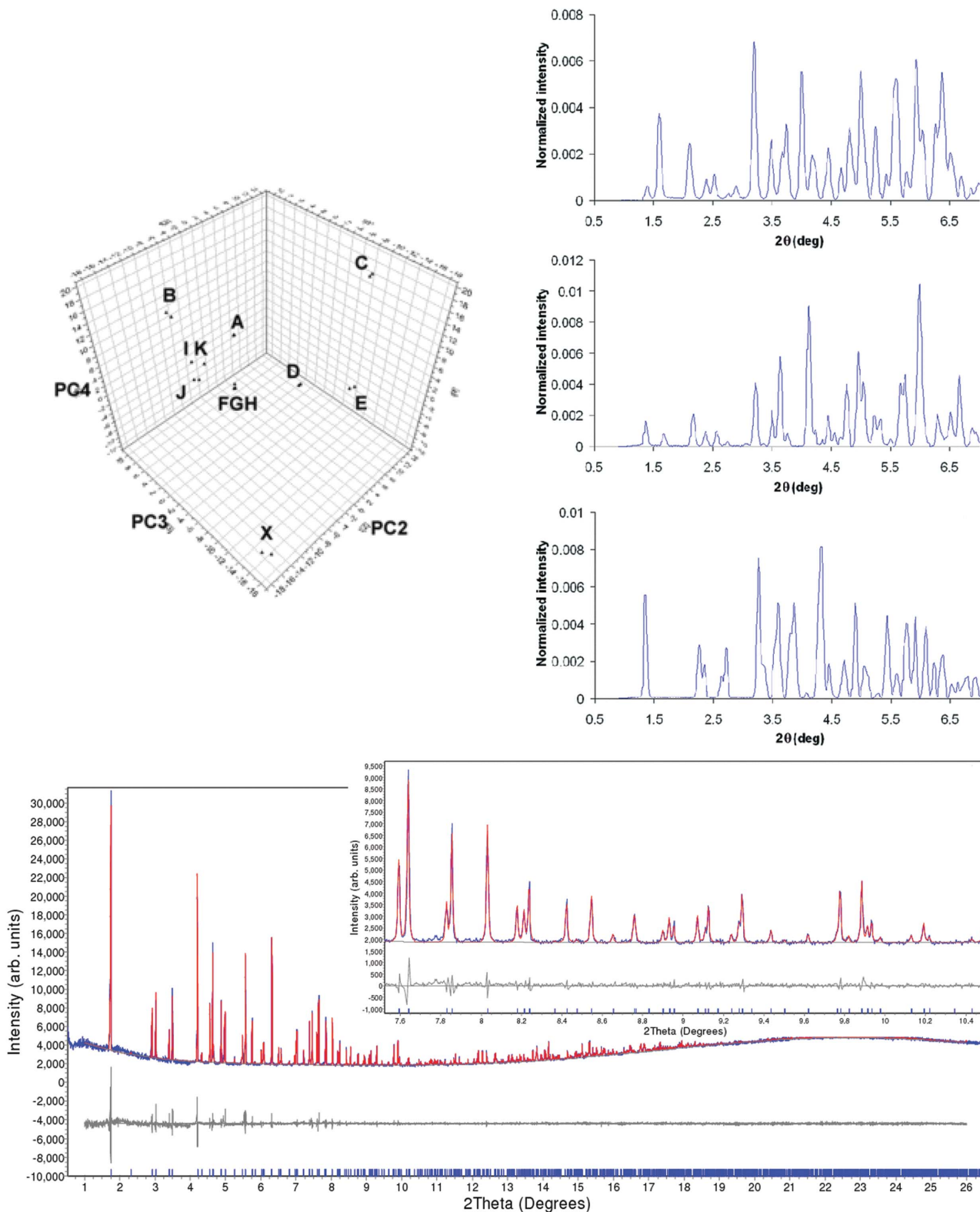


Figure 6

Upper left: Score plot from the PCA analysis of the insulin data sets. The three-dimensional plot is shown with PC2, PC3 and PC4 and shows the separation of different samples into nine separate clusters. [From Fig. 4 in Norrman *et al.* (2006).] Upper right: Normalized intensity versus 2θ for three types of rhombohedral crystals which have structural differences in the arrangement of the *N*-terminal residues of the insulin *B* chain (R_6 , $T_3R_3^1$ and T_6 conformation). [From Fig. 3 in Norrman *et al.* (2006).] Lower panel: Pawley refinement of ID31 data ($\lambda = 1.25$ Å, RT) of hexagonal insulin T_6 [space group $R3$, $a = 81.75313$ (3), $c = 34.0688$ (2) Å]. In total, 2265 intensities were extracted up to 2.8 Å resolution. The blue, red and lower grey lines represent the experimental and calculated patterns and the difference between experimental and calculated profiles, respectively. The vertical bars correspond to Bragg reflections compatible with the refined hexagonal structural model. The inset shows a magnification of a small 2θ region of the fit.

positions. These features are complementary to area-detector single-crystal experiments, where the aim is normally to measure diffraction peaks at the highest possible scattering angles, to improve the direct-space resolution in an electron-density map. By exploiting the excellent reciprocal-space resolution of powder instruments, a variety of complementary studies can be carried out which are typically more difficult to do with single crystals. Powder samples contain many millions of individual crystallites and give information about the bulk properties of a material; there is no question of whether a particular crystallite is representative of the whole sample. Wherever microcrystalline proteins come into use as biomaterials (Margolin & Navia, 2001) or medicines, there will be an increasingly important role for powder diffraction to play.

5.1. Crystal screening and principal component analysis (PCA)

Insulin injections are used daily by people who suffer from diabetes in order to regulate levels of glucose in their blood stream. The duration and effects of these injections depend on the precise details of the medical formulation. Microcrystalline forms of insulin tend to be slower acting as the crystallites must dissolve before acting. Crystallite size and polymorphism both affect dissolution rates and so they are important factors in understanding formulations. Because of the medical relevance, insulin has been intensively studied as a drug since the 1920's.

Recently, Norrman *et al.* (2006) have used powder diffraction to systematically explore a series of different polymorphs of crystalline insulin. Using a range of well characterized crystalline forms, they were able to show that the fingerprinting capabilities of a powder profile can be used with protein samples. Moreover, by using a principal component analysis (PCA) of the powder spectra, a rapid and reliable classification of the samples could be carried out. Fig. 6, top left, shows a score plot where different samples fall into distinct clusters corresponding to different crystalline polymorphs and molecular conformations.

The cluster marked 'X' on Fig. 6 corresponds to a new crystalline form of insulin discovered by Norrman *et al.* (2006) in the course of their study. Using their database of powder patterns, any new samples can quickly and reliably be characterized to see if they correspond to a known crystalline form or if they can be identified as 'new'. McCrone's (1965) statement that 'the number of polymorphs of a material is proportional to the time spent investigating' appears to apply to crystalline proteins as well. Further measurements at the synchrotron (Fig. 6, lower panel) confirm these results and furthermore allow more subtle differences between samples to be quantified.

5.2. Cryocooling of microcrystalline protein samples

In a powder experiment, we sample a large number of very small crystals and the specimen will still be a powder even if some of the crystallites are broken. This has made powder methods extremely useful for the study of phase transitions

and other phenomena which can ruin a fragile single crystal. If diffraction peaks are observed that are broader than the instrumental resolution, they can give some information about the microstructure of the sample. When there is a spread of cell parameters in the sample, a 'microstrain' broadening occurs and this has a larger effect at higher scattering angles. When measured at room temperature, many proteins give diffraction peaks where the widths are limited only by the instrumental resolution. This indicates that the samples are extremely homogeneous, with all crystallites having the same unit-cell parameters. The truncation of the crystal lattice by a very small crystallite size also leads to a peak broadening which is similar over all of reciprocal space and is normally significant for crystallite sizes below about 1 μm . For structural refinements, the microstrain peak broadening is more problematic as it increases at higher resolution (scattering angle) where the peak overlap is already more severe.

The suppression of radiation damage by cryocooling (Garman & Schneider, 1997) has meant that nowadays the majority of single-crystal structures are determined at low temperatures. For powder experiments, it is important to preserve sharp diffraction peaks and avoid any unnecessary peak overlaps, however, it would also be useful to suppress radiation damage by cryocooling. We have investigated the freezing processes occurring in turkey egg white lysozyme by collecting data while cooling and warming a powder sample (Margiolaki *et al.*, 2005). A phase transition in the protein structure is accompanied by freezing of the mother liquor and leads to a broadening of the powder peaks. Fig. 7 shows the powder data as a function of temperature, where the behaviour at the phase transition is clear. In that particular case, the 'microstrain'-related peak-broadening parameter was in agreement with the bulk overall contraction of the unit cell during freezing.

In order to find an optimized approach to cryocooling of powder samples, we have systematically varied the cryoprotectant type and concentration (Jenner *et al.*, 2007). With

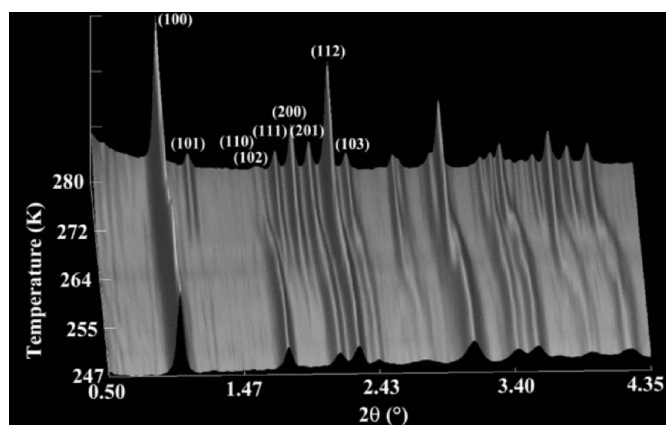


Figure 7 Diffraction profiles plotted as a function of temperature collected while warming a turkey egg white lysozyme sample through the freezing transition. The sample was initially cooled from 320 to 105 K at a rate of 80 K h⁻¹ and then warmed back to 320 K at a rate of 160 K h⁻¹. [From Margiolaki *et al.* (2005).]

HEWL, it was shown that the unit-cell parameters of the microcrystals at low temperature are strongly affected by the addition of cryoprotectants. The changes in cell parameters were anisotropic and depended on the type of cryoprotectant molecule used. Also, the amount of induced peak broadening was found to be proportional to the change in unit-cell parameters caused by freezing; when there is less peak broadening, the unit cell was closer to the room-temperature cell. In all of these cases, the powder diagrams could be well accounted for by considering only a ‘microstrain’-related peak-broadening term, meaning that the peak broadening arises mainly from having a distribution of unit-cell dimensions in the sample. Particle-size broadening effects were negligible in comparison. The increased sample size and use of a capillary to contain the powder sample means that the cooling rates are significantly lower for powders in comparison to single crystals in loops and some challenges remain in finding ways to optimally cryocool microcrystalline samples.

6. Obtaining new protein structures from powder data

If powder data contain enough information to refine the structure of a small protein, then it follows that the structure might also be solved from the same data. We are choosing an inclusive definition of ‘solving’ a protein structure in this context! The use of stereochemical restraints and introduction of prior knowledge, such as the amino acid sequence, will be

used to fill in missing information at high resolution. Powder data can certainly be used to uncover new structural information at low or even modest resolution, and we expect to see this becoming a more common approach in the future. As water molecules have been seen in at least two refinements already (Von Dreele, 2007; Margiolaki, Wright, Wilmanns *et al.*, 2007), we can be confident that powder data will be sufficient to unravel new structural information.

6.1. The molecular replacement method

In the molecular replacement method (see *e.g.* Rossmann, 1990), a model for a protein molecule is proposed based on other known structures of proteins which have similar amino acid sequences. This model is then oriented and positioned in the unit cell to match the experimental data for the unknown structure. In this era of structural genomics (Stevens *et al.*, 2001), the quality of the search models is expected to improve dramatically so that molecular replacement using a powder diagram may be sufficient to confirm that a new protein is indeed similar to one that is previously known. Proteins are often found in a variety of different organisms with very similar amino acid sequences, or are expressed with a few modified residues to investigate the effects these changes have on the protein.

There are only six degrees of freedom to be determined per molecule in solving a structure by molecular replacement. The first three give the orientation and the second three give the

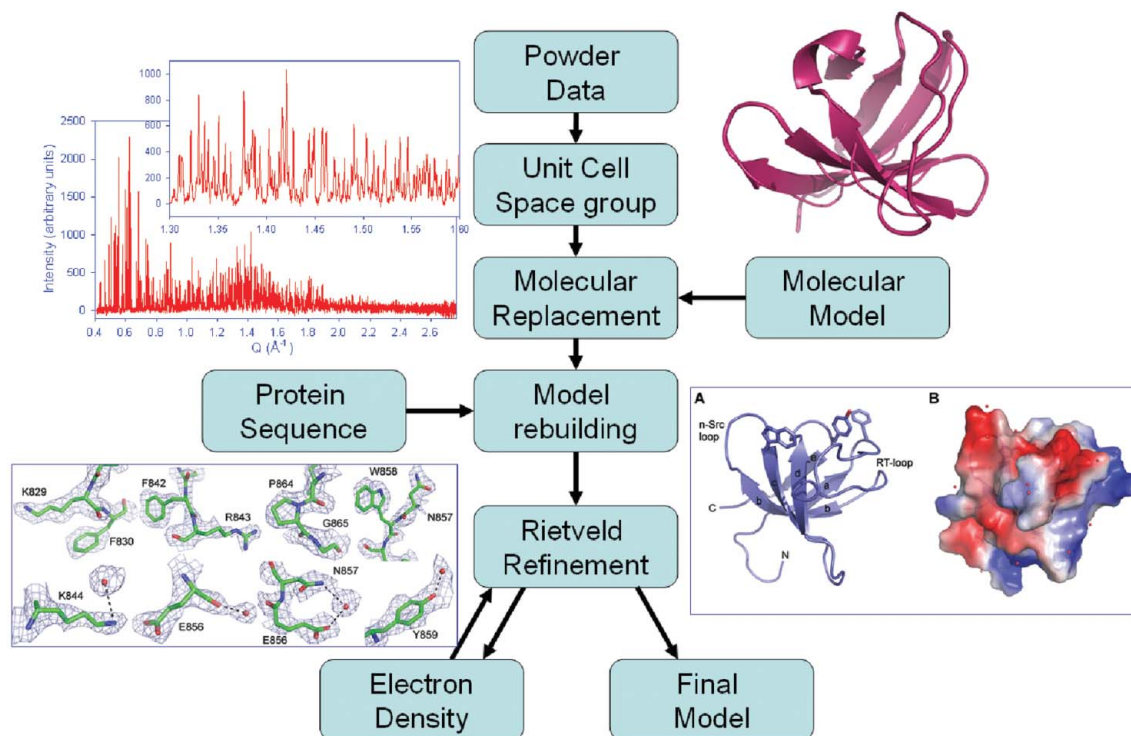


Figure 8

Powder diffraction data analysis procedure followed for structure solution *via* the molecular-replacement method, model building and structure refinement. The data and model shown correspond to the second SH3 domain of ponsin (Margiolaki, Wright, Wilmanns *et al.*, 2007) and final omit maps are shown on the lower left.

position of the molecule with respect to the symmetry elements of the space group. Relatively low resolution data can often be used for molecular replacement problems. A resolution of 6 Å can be sufficient and a cut-off of 3 Å is often applied. Typically, even with single-crystal data, the highest-resolution data are not so useful as the details of the model structure will not be correct at high resolution anyway.

The solution of the structure of T_3R_3 insulin by Von Dreele *et al.* (2000) is the first example of the molecular replacement method at work and is described above (§3). In our own laboratory, we have been able to successfully use the *MOLREP* software (Vagin & Teplyakov, 1997) together with powder data for a variety of small proteins. Integrated intensities are extracted (Wright, 2004) from the powder profile and then treated as if they came from a single crystal, ignoring the overlap problem. Fortunately, the molecular replacement method tolerates errors in the input data as if they were due to errors in the model for the structure. In the cases of turkey (Margiolaki *et al.*, 2005) and hen (Basso *et al.*, 2005) egg white lysozymes, there was a clear distinction of the correct structure which was proposed by *MOLREP* from the second best, and in both cases these were the correct positions and orientations for the molecule in the unit cell.

After determining the position and orientation of a model for the protein molecule in the unit cell, it is of critical importance to be able to go through the structure and identify the regions that are different in the new protein compared to the initial model. Fig. 8 gives an overview of the whole procedure as applied to the SH3 domain of ponsin (Margiolaki, Wright, Wilmanns *et al.*, 2007). As described in §4, the high-quality extracted intensities from the measured profiles led to the unique determination of an unknown 544-atom protein structure from powder data. The search model had a moderate similarity (38% sequence identity) and was sufficient for molecular replacement and model building based on electron-density maps.

Another example of a molecular replacement style approach is the use of phase angle or structural information from another technique, combined with data collected using powder methods. Oka *et al.* (2006) have investigated the structure of purple membranes, which comprise two-dimensional crystals of bacteriorhodopsin trimers. They collected high-resolution synchrotron powder diffraction data using a 1 m path length Guinier-type camera at beamline BL40B2 at SPring-8. By combining extracted intensities up to a resolution of 4.2 Å with phases from cryoelectron microscopy, they were able to produce an electron-density map, which is reproduced here in Fig. 9. The data from the synchrotron improved the resolution for purple membrane compared to that attained previously with laboratory X-rays (intensities extracted up to ~7 Å) and opens up the possibility of studying these systems under a wider range of conditions.

6.2. The isomorphous replacement method

When there is no starting model available for a protein molecule, then we face the crystallographic phase problem.

The isomorphous replacement method (Perutz, 1956) determines the phases of X-ray reflections by comparing data from a series of crystals which contain different additive heavy atoms. This method was already in common use prior to the application to proteins for the determination of organic structures by comparing crystalline salts with differing cations or anions. The first application to proteins (Green *et al.*, 1954) was used to determine the signs of centric reflections in haemoglobin. In order to use the isomorphous replacement method, one needs to collect data for a protein with and without the added heavy atom and to be able to measure and interpret the differences in the data due to the added atoms. Difficulties arise when the addition of the heavy atom perturbs the structure leading to non-isomorphism and differences that can no longer be simply interpreted by a few atoms.

With powder data, we can determine whether or not a particular heavy atom has formed a derivative by comparing the unit-cell parameters, which are normally perturbed by the addition of a heavy atom. Von Dreele *et al.* (2006) have shown that an isolated Xe atom shows up in a difference Fourier map of HEWL under 8 bar of Xe pressure. This observation was a clear indicator that it should be possible to solve heavy-atom substructures and then go on to phase protein structures from powder data.

In the case of a gadolinium derivative of hen egg white lysozyme (Wright, Besnard *et al.*, 2007), we prepared derivatives using different concentrations of Gd and also different

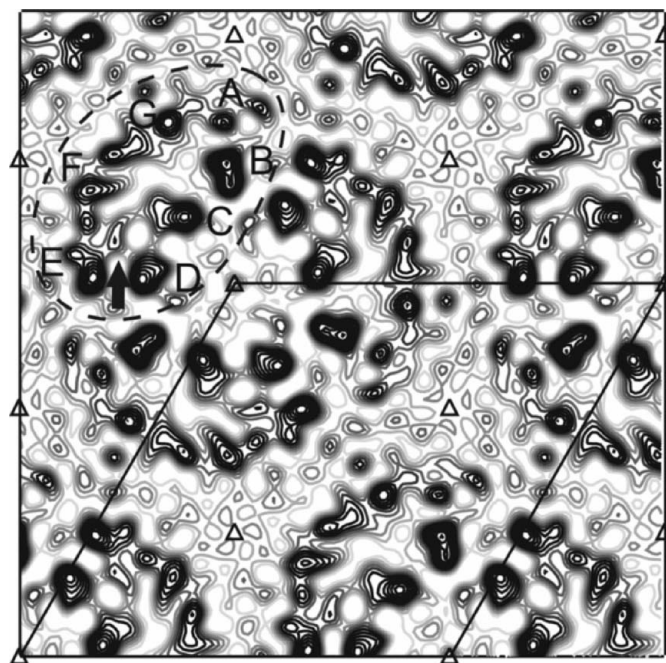


Figure 9

Electron-density map of purple membrane projected onto a membrane plane. It was calculated with diffraction amplitudes up to the (113) reflection, a resolution of 4.2 Å. The envelope shown by the dashed line shows the border of one *BR* monomer, and the characters *A–G* are the names of seven transmembrane α -helices. The triangle symbols denote the positions of threefold rotational symmetry points. [From Fig. 3 in Oka *et al.* (2006).]

pH values in order to alleviate the peak-overlap problem. Extracted intensities were then used to solve for the atomic positions of the Gd atoms using *SHELXD* (Uson & Sheldrick, 1999) and then refinement of the heavy-atom substructure and phasing were carried out using *Sharp* (La Fortelle & Bricogne, 1997). Fig. 10 shows the solvent channels in the protein structure determined during this phasing procedure and the positions of the molecules from the known structure (Wright, Besnard *et al.*, 2007).

Similarly, in the case of a uranium derivative of elastase (Besnard *et al.*, 2007), the peak-overlap problem was reduced by comparing data sets that had been exposed to the X-ray beam for different amounts of time. The gradual changes in unit-cell parameters due to radiation damage changed the pattern of peak overlaps sufficiently to deconvolve the overlapping peaks. The U atom is clearly visible in a difference Patterson map from native and derivative data sets and phasing using *Sharp* proceeds to give a solvent mask which clearly and correctly delineates the protein and solvent regions (Wright, Besnard *et al.*, 2007).

Intensive work is still in progress in this area of phase determination from protein powder data. The examples published so far clearly show that phasing is feasible, but there is still a struggle to cross the barrier between low-resolution solvent/protein mask information and high-resolution maps where the protein and side chains are visible.

6.3. Emerging powder phasing methods

With single-crystal samples, the multiple anomalous dispersion (MAD) method has revolutionized the process of solving novel protein structures. Unfortunately, in powder experiments, the Friedel pairs (h, k, l) and ($-h, -k, -l$) are always exactly overlapped, and so anomalous differences cannot be determined. Nevertheless, dispersive differences, due to the variation in f' can be measured. The implications of

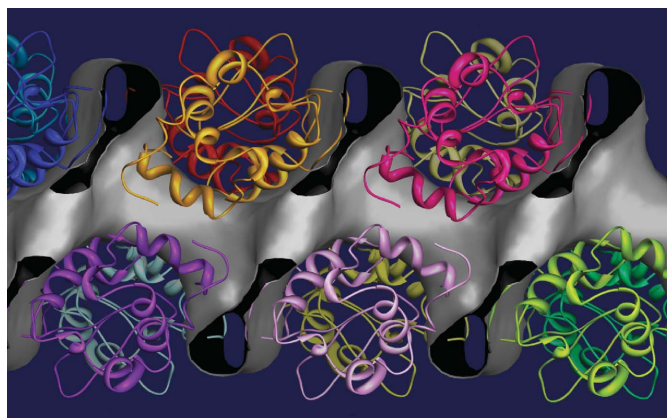


Figure 10

Solvent channels in hen egg white lysozyme crystals. The molecular envelope derived from single isomorphous replacement using a gadolinium derivative is represented as grey surface. The figure shows the linear solvent channel which traverses the crystal parallel to the c axis (horizontal display direction). The protein crystal structure, represented as a main-chain ribbon model, is superimposed on this map. [From Wright, Besnard *et al.* (2007).]

this systematic overlap problem have been discussed by Prandl (1990, 1994), who reached the conclusion that phases can be determined for centric reflections using a single anomalous scatterer but that, for acentric reflections, at least two anomalously scattering species are required. Plöhn & Büldt (1986) have also considered the powder phasing problem and also discuss the ways in which D_2O/H_2O substitution could be used for labelling with neutron data.

The dispersive difference phasing methods were demonstrated by application to $SrSO_4$ by Burger *et al.* (1998), where the centric structure was solved using data near the Sr edge. More recently, Helliwell *et al.* (2005) have collected data at the Ni edge for $Ni(SO_4) \cdot 6H_2O$ and determined optimal choices of wavelengths at the absorption edge. They indicate that the methods may also be applicable to microcrystalline protein samples. Work is in progress in our own laboratory, as well as in others, to exploit the dispersive differences that can be obtained by tuning the wavelength around an X-ray absorption edge for powder data.

7. Conclusions

In the last 10 years, the use of powder data with macromolecules has gone from being an ambitious suggestion to a respectable practice. A series of demonstration experiments and data analyses has been carried out which establishes the validity of the methodology. Although developments are still in progress for phasing methods and structure solution, there are already molecular-replacement and structure-refinement methods that are now mature enough for consumption by the more adventurous protein crystallographer. The solution of the structure of the SH3 domain of ponsin shows how powder diffraction is already moving 'beyond lysozyme' and is a technique ready to tackle genuine biological problems.

We are grateful to Dr A. N. Fitch for his advice and support during this project, and we also thank S. Basso, M. Jenner and L. Knight for their work with us on protein powders. We would like to thank Drs R. B. Von Dreele, G. Fox, M. Norrman, G. Schluckebier, K. Miura, M. Wilmanns and N. Pinotsis, and also Dr M. Schiltz and his group, for their help, collaboration and permission to reproduce figures. Finally, we thank ESRF for provision of synchrotron beam time.

References

- Amos, L. A., Jubb, J. S., Henderson, R. & Vigers, G. (1984). *J. Mol. Biol.* **178**, 711–729.
- Anduleit, K., Sutton, G., Diprose, J. M., Mertens, P. P. C., Grimes, J. M. & Stuart, D. I. (2005). *Protein Sci.* **14**, 2741–2743.
- Baerlocher, Ch., McCusker, L. B., Prokic, S. & Wessels, T. (2004). *Z. Kristallogr.* **219**, 803–812.
- Basso, S., Fitch, A. N., Fox, G. C., Margiolaki, I. & Wright, J. P. (2005). *Acta Cryst.* **D61**, 1612–1625.
- Bernal, J. D. & Fankuchen, I. (1941). *J. Gen. Physiol.* **25**, 111–165.
- Besnard, C., Camus, F., Fleurant, M., Dahlström, A., Wright, J. P., Margiolaki, I., Pattison, P. & Schiltz, M. (2007). *Z. Kristallogr.* In the press.
- Bhat, T. N. (1988). *J. Appl. Cryst.* **21**, 279–281.

- Brunelli, M., Wright, J. P., Vaughan, G. B. M., Mora, A. J. & Fitch, A. N. (2003). *Angew. Chem. Int. Ed.* **42**, 2029–2032.
- Bunge, H. J., Dahms, M. & Brokmeier, H. G. (1989). *Crystallogr. Rev.* **2**, 67–86.
- Burger, K., Cox, D., Papoular, R. & Prandl, W. (1998). *J. Appl. Cryst.* **31**, 789–797.
- Cerny, R. (1996). Proc. of the 45th Annual Conf. on Applications of X-ray Analysis, Denver, 1996. *Adv. X-ray Anal.* **40**, 433–438.
- Ciszak, E. & Smith, G. D. (1994). *Biochemistry*, **33**, 1512–1517.
- Corey, R. B. & Wyckoff, R. W. G. (1936). *J. Biol. Chem.* **114**, 407–414.
- Coulibaly, F., Chiu, E., Ikeda, K., Gutmann, S., Haebel, P. W., Schulze-Briese, C., Mori, H. & Metcalf, P. (2007). *Nature (London)*, **446**, 97–101.
- Dahms, M. & Bunge, H. J. (1986). *Textures Microstruct.* **6**, 167–179.
- DeLano, W. L. (2004). *The PyMOL Molecular Graphics System*. DeLano Scientific LLC, San Carlos, CA, USA, <http://www.pymol.org/>.
- Faber, J. & Fawcett, T. (2002). *Acta Cryst.* **B58**, 325–332.
- Garman, E. F. & Schneider, T. R. (1997). *J. Appl. Cryst.* **30**, 211–237.
- Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *Proc. R. Soc. London Ser. A*, **225**, 287.
- Helliwell, J. R., Helliwell, M. & Jones, R. H. (2005). *Acta Cryst.* **A61**, 568–574.
- Jenner, M. J., Wright, J. P., Margiolaki, I. & Fitch, A. N. (2007). *J. Appl. Cryst.* **40**, 121–124.
- La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
- Larson, A. C. & Von Dreele, R. B. (2004). *General Structure Analysis System (GSAS)*, Los Alamos National Laboratory Report LAUR 86-748, Los Alamos, USA.
- McCrone, W. C. (1965). *Polymorphism in Physics and Chemistry of the Organic Solid State*, Vol. 2, edited by D. Fox, M. M. Labes & A. Weissberger, pp. 725–767. New York: Wiley Interscience.
- Margiolaki, I., Wright, J. P., Fitch, A. N., Fox, G. C., Labrador, A., Von Dreele, R. B., Miura, K., Gozzo, F., Schiltz, M., Besnard, C., Camus, F., Pattison, P., Beckers, D. & Degen, T. (2007). *Z. Kristallogr.* In the press.
- Margiolaki, I., Wright, J. P., Fitch, A. N., Fox, G. C. & Von Dreele, R. B. (2005). *Acta Cryst.* **D61**, 423–432.
- Margiolaki, I., Wright, J. P., Wilmanns, M., Fitch, A. N. & Pinotsis, N. (2007). *J. Am. Chem. Soc.* **129**, 11865–11871.
- Margolin, A. L. & Navia, M. A. (2001). *Angew. Chem. Int. Ed.* **40**, 2204–2222.
- Norrman, M., Ståhl, K., Schluckebier, G. & Al-Karadaghi, S. (2006). *J. Appl. Cryst.* **39**, 391–400.
- Oka, T., Miura, K., Inoue, K., Ueki, T. & Yagi, N. (2006). *J. Synchrotron Rad.* **13**, 281–284.
- Pawley, G. S. (1981). *J. Appl. Cryst.* **14**, 357–361.
- Perutz, M. F. (1954). *Proc. R. Soc. London Ser. A*, **225**, 264–286.
- Perutz, M. F. (1956). *Acta Cryst.* **9**, 867–873.
- Plöhn, H.-J. & Büldt, G. (1986). *J. Appl. Cryst.* **19**, 255–261.
- Prandl, W. (1990). *Acta Cryst.* **A46**, 988–992.
- Prandl, W. (1994). *Acta Cryst.* **A50**, 52–55.
- Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.
- Rietveld, H. M. (1969). *J. Appl. Cryst.* **2**, 65–71.
- Rossmann, M. G. (1990). *Acta Cryst.* **A46**, 73–82.
- Shankland, K., David, W. I. F. & Sivia, D. S. (1997). *J. Mater. Chem.* **7**, 569–572.
- Sivia, D. S. (2000). *J. Appl. Cryst.* **33**, 1295–1301.
- Sliz, P., Harrison, S. C. & Rosenbaum, G. (2003). *Structure*, **11**, 13–19.
- Smith, G. D., Pangborn, W. & Blessing, R. H. (2001). *Acta Cryst.* **D57**, 1091–1100.
- Stevens, R. C., Yokoyama, S. & Wilson, I. A. (2001). *Science*, **294**, 89–92.
- Uson, I. & Sheldrick, G. M. (1999). *Curr. Opin. Struct. Biol.* **9**, 643–648.
- Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
- Von Dreele, R. B. (1999). *J. Appl. Cryst.* **32**, 1084–1089.
- Von Dreele, R. B. (2001). *Acta Cryst.* **D57**, 1836–1842.
- Von Dreele, R. B. (2005). *Acta Cryst.* **D61**, 22–32.
- Von Dreele, R. B. (2007). *J. Appl. Cryst.* **40**, 133–143.
- Von Dreele, R. B., Lee, P. L. & Zhang, Y. (2006). *Z. Kristallogr. Suppl.* **23**, 3–8.
- Von Dreele, R. B., Stephens, P. W., Smith, G. D. & Blessing, R. H. (2000). *Acta Cryst.* **D56**, 1549–1553.
- Wright, J. P. (2004). *Z. Kristallogr.* **219**, 791–802.
- Wright, J. P., Besnard, C., Margiolaki, I., Basso, S., Camus, F., Fitch, A. N., Fox, G. C., Pattison, P. & Schiltz, M. (2007). *J. Appl. Cryst.* In the press.
- Wright, J. P. & Forsyth, J. B. (2000). *PRODD, Profile Refinement of Diffraction Data using the Cambridge Crystallographic Subroutine Library*. Rutherford Appleton Laboratory Report RAL-TR-2000-012, Didcot, Oxon, UK.
- Wright, J. P., Markvardsen, A. J. & Margiolaki, I. (2007). *Z. Kristallogr.* In the press.
- Wyckoff, R. W. G. & Corey, R. B. (1936). *J. Biol. Chem.* **116**, 51–55.
- Xiaa, D., Yu-Kunb, S., McCraec, M. A. & Rossmann, M. G. (1991). *Virology*, **180**, 153–158.