Molecular envelopes derived from protein powder diffraction data
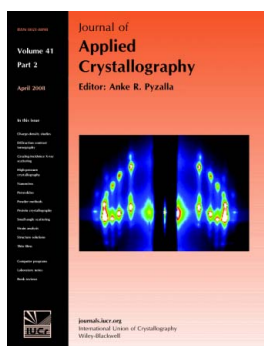
Jonathan P. Wright, Céline Besnard, Irene Margiolaki, Sebastian Basso, Fabrice Camus, Andrew N. Fitch, Gavin C. Fox, Philip Pattison and Marc Schiltz

Many research topics in condensed matter research, materials science and the life sciences make use of crystallographic methods to study crystalline and non-crystalline matter with neutrons, X-rays and electrons. Articles published in the *Journal of Applied Crystallography* focus on these methods and their use in identifying structural and diffusion-controlled phase transformations, structure–property relationships, structural changes of defects, interfaces and surfaces, *etc.* Developments of instrumentation and crystallographic apparatus, theory and interpretation, numerical analysis and other related subjects are also covered. The journal is the primary place where crystallographic computer program information is published.

Crystallography Journals Online is available from journals.iucr.org

Molecular envelopes derived from protein powder diffraction data

Jonathan P. Wright,[a]‡ Céline Besnard,[b]‡ Irene Margiolaki,[a] Sebastian Basso,[b] Fabrice Camus,[b] Andrew N. Fitch,[a] Gavin C. Fox,[a] Philip Pattison[b,c] and Marc Schiltz[b]*

[a]European Synchrotron Radiation Facility (ESRF), 6 Rue Jules Horowitz, F-38043 Grenoble Cedex 9, France, [b]École Polytechnique Fédérale de Lausanne (EPFL), Laboratoire de Cristallographie, CH-1015 Lausanne, Switzerland, and [c]Swiss–Norwegian Beamlines (SNBL) at the ESRF, 6 Rue Jules Horowitz, F-38043 Grenoble Cedex 9, France. Correspondence e-mail: marc.schiltz@epfl.ch

The preparation of single crystals suitable for X-ray analysis is frequently the most difficult step in structural studies of proteins. With the aid of two examples, it is shown that *de novo* solution of the crystallographic phase problem can be achieved at low resolution using microcrystalline powder samples *via* the single isomorphous replacement method. With synchrotron radiation and optimized instrumentation, high-quality powder patterns have been recorded, from which it was possible to generate phase information for structure factors up to 6 Å resolution. pH- and radiation-induced anisotropic lattice changes were exploited to reduce the problem of overlapping reflections, which is a major challenge in protein powder diffraction. The resulting data were of sufficient quality to compute molecular envelopes of the protein molecule and to map out the solvent channels in the crystals. The results show that protein powder diffraction can yield low-resolution data that are potentially useful for the characterization of microcrystalline proteins as novel micro- and mesoporous materials as well as for structural studies of biologically important macromolecules.

1. Introduction

X-ray diffraction is one of the most powerful techniques for investigating the molecular architecture of biologically important macromolecules. However, the growing of large single crystals, typically of average dimensions of the order of 100 µm, which are of sufficient quality for high-resolution diffraction studies, is currently one of the major bottlenecks in macromolecular crystallography. Crystal growth remains by and large a time-consuming trial-and-error process (McPherson, 1999). On specially dedicated microfocus beamlines at third-generation synchrotrons, crystals of sizes of only a few tens of micrometres have been studied successfully (Coulibaly *et al.*, 2007), but these measurements present considerable technical challenges. In particular, with samples of small size and with highly focused intense X-ray beams, radiation damage inside the crystal is a severe problem even if cryogenic cooling is applied. Frequently, diffraction data from several crystals have to be measured and merged in order to obtain a complete data set. Sliz *et al.* (2003) have estimated that, at 1 Å wavelength, the smallest crystal that can yield a full data set to 3.5 Å resolution must have a diameter of at least 20 µm (assuming a $100 \times 100 \times 100$ Å unit cell). There is thus little hope that the current size limit for single-crystal studies can be lowered in the future, since radiation damage will ultimately be the limiting factor.

On the other hand, microcrystalline powders can be obtained for many proteins by batch precipitation (Von Dreele, 2003), which is much less demanding than the growing of a large single crystal. The size of individual crystallites in such polycrystalline precipitates is usually far too small for single-crystal diffraction studies, but it is, in many cases, ideal for powder diffractometry. Indeed, in the protein powder diffraction studies reported so far, the estimated crystallite sizes were of the order of 0.1–1 µm, with very little microstrain broadening, thus giving rise to very sharp diffraction peaks. Since, in a powder diffraction experiment many thousands of crystallites are irradiated and contribute to the diffraction signal, the effects of radiation damage (for a given number of diffracted photons) are distributed over a much larger sample volume than would be the case for a single microcrystal (Margiolaki *et al.*, 2007).

Microcrystalline powders of proteins can also be considered as novel materials with properties potentially useful in technological applications such as biocatalysis, chiral separation (Margolin & Navia, 2001) and drug delivery (Basu *et al.*, 2004). These materials are not directly amenable to investigation by standard single-crystal diffraction techniques. X-ray powder diffraction may often provide the only technique able to deliver the essential structural data, which will allow the properties of these interesting biomaterials to be properly exploited.

Previous work by Von Dreele (2003) and others (Margiolaki & Wright, 2008; Margiolaki *et al.*, 2007) has demonstrated

‡ These authors contributed equally to this work.

that high-quality synchrotron diffraction data can be collected on polycrystalline protein powders, despite the very large unit cells of these materials (Von Dreele, 1999). During a diffraction experiment on a polycrystalline sample, an X-ray beam illuminates many randomly oriented crystallites. Thus, the three-dimensional diffraction data from each microcrystallite are also randomly oriented and superimpose upon each other. As a consequence, data are only recorded as a function of the scattering angle, and the loss of information owing to the overlap of Bragg peaks becomes progressively worse at higher angles. Notwithstanding this serious limitation with respect to single-crystal diffraction, the powder diffraction data recorded on a number of protein samples have been shown to be of sufficient quality to perform successfully a combined Rietveld and stereochemical restraint refinement of macromolecular structures (Von Dreele *et al.*, 2000; Von Dreele, 2007). It has also been established that the molecular replacement method (MR), which has been so successful in single-crystal protein crystallography, can be used with powder diffraction data (Margiolaki *et al.*, 2005). Finally, it has been demonstrated that powder diffraction can be used to detect the formation of protein–ligand complexes by difference Fourier techniques (Von Dreele, 2001, 2005). However, all of these applications essentially rely on the availability of a sufficiently good molecular model for the protein structure. The diffraction data are used either to refine the model or to place it in the crystal unit cell (MR) or to complete it for relatively small missing components (ligand detection). In these applications, the loss of information arising from the overlap of Bragg reflections in a powder spectrum is therefore manageable. The situation is considerably more complicated for *de novo* structure determination where no molecular model is available. In such cases, the standard experimental phasing methods used in single-crystal protein crystallography exploit intensity differences produced by labelling proteins with heavy atoms (method of isomorphous replacement) and/or through the phenomenon of anomalous X-ray scattering (Fourme *et al.*, 2000). Phase information is generated individually for each Bragg reflection through relatively small intensity modulations, so that the presence of overlapping peaks will seriously handicap these methods.

In the present study, we address the problem of extracting *de novo* structural information from protein powder diffraction data. On two test samples, we exploit different methods to reduce the problem of overlapping reflections and show that the method of single isomorphous replacement (SIR) can be successfully applied to phase low-angle reflections in protein powder patterns, allowing the computation of molecular envelopes.

2. Experimental

2.1. Sample preparation

Two protein samples were selected for this study. The enzyme porcine pancreatic elastase (PPE) has a molecular weight of 25.9 kg mol^{-1} and is able to hydrolyse a wide variety of protein substrates, including native elastin, a substrate not attacked by other proteinases. Hen egg white lysozyme (HEWL) has a molecular weight of 14.6 kg mol^{-1} and catalyses the hydrolysis of specific kinds of polysaccharides comprising the cell walls of bacteria.

Lyophilized HEWL and PPE were purchased, respectively, from Sigma–Aldrich and SERVA Chemicals and were used without further purification. Microcrystalline slurries were prepared by batch precipitation following a published procedure (Von Dreele, 2003) and on the basis of the known crystallization conditions. For HEWL, gadolinium-containing heavy-atom derivatives were prepared by cocrystallizing the protein with the complex Gd-Hp-Do3A (Girard *et al.*, 2002) at concentrations of 10 and 50 m*M*. For both the native and the Gd-containing heavy-atom derivatives, samples were prepared at pH values of 3.5 and 4.5. Thus, a series of six samples was prepared, varying both the pH of the crystallization buffer and the concentration of the heavy-atom compound. For PPE, a uranium-containing heavy-atom derivative was obtained by soaking a native sample in a 5 m*M* uranyl nitrate solution for several days.

2.2. X-ray data collection

All samples were loaded into 1.5 mm-diameter glass capillaries and centrifuged in order to enhance the packing of the crystallites (Von Dreele, 2003). Excess mother liquor was removed and the capillaries were sealed with wax to prevent dehydration of the samples.

High-resolution powder diffraction data were collected at room temperature (291 K) on the undulator beamline ID31 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France (Fitch, 2004). The parallel beam geometry instrument is equipped with nine Si(111) analyser crystals and provides data with very high angular resolution, with a minimum instrumental contribution to the FWHM of around 0.003°. The analyser crystal removes the contribution to angular resolution that is due to the sample size and allows irradiation of larger sample volumes (of the order of 3 mm^3), which is important for obtaining good counting statistics from weakly scattering protein samples. The powder patterns obtained from proteins have very narrow diffraction lines, as low as 0.01° FWHM, which is essentially due to crystallite size and microstrain effects. Peak overlap, especially in the low-angle region, can therefore be significantly reduced. High-speed spinning of the sample was applied in order to ensure sufficient powder averaging. The nine detector modules (which are separated by approximately 2°) were scanned over a 20° range at a speed of 10° s^{-1}.

For each of the six HEWL samples, eight scans were collected at a wavelength of 1.54999 (3) Å, the capillary being translated in between scans in order to expose a fresh region of sample unaffected by radiation damage. These raw scans were added, leading to increased counting statistics. For PPE samples, data were collected at a wavelength of 1.25085 (3) Å. The samples were also translated, but at each sample position, four (for the native) and five (for the uranyl derivative) scans

Table 1
Variation of lattice parameters for HEWL as a function of pH and concentration of Gd.

Gd concen-tration (mM)	pH = 3.5		pH = 4.5	
	$a = b$ (Å)	c (Å)	$a = b$ (Å)	c (Å)
0 (native)	79.23045 (14)	37.95921 (11)	79.24040 (14)	37.93134 (11)
10	79.119566 (20)	38.03601 (19)	79.17331 (17)	38.03711 (15)
50	78.99279 (24)	38.45130 (21)	78.98786 (24)	38.44846 (21)

Table 2
Quality-of-fit of the multi-pattern Pawley refinements.

	PPE native	PPE uranyl derivative	HEWL native	HEWL Gd derivative (10 mM)	HEWL Gd derivative (50 mM)
R_{wp} (%)	8.34	3.33	5.55	5.96	6.92
R_{exp} (%)	5.78	2.79	4.63	5.06	4.95
R_p (%)	6.17	2.61	4.46	4.87	5.38
χ^2	1.44	1.19	1.43	1.39	1.95

were collected to exploit radiation-induced lattice changes. A total of six sample translations were used for both the native and the uranyl derivative of PPE. The raw scans from all sample positions corresponding to a given irradiation were added, leading to increased counting statistics. The final powder data consist of six different patterns for HEWL (two patterns corresponding to different pH values for the native and similarly for each one of the two Gd-Hp-Do3A derivatives) and nine different patterns for PPE (four patterns for native PPE and five patterns for the uranyl derivative, corresponding to different irradiation stages).[1]

2.3. Single-crystal reference data

Reference data for comparisons were collected on single crystals of native and uranylated PPE at the Swiss–Norwegian Beamline at the ESRF using an image-plate area detector. The data were processed with *MOSFLM* (Leslie, 1993) and reduced with software from the *CCP4* package (Collaborative Computational Project, Number 4, 1994).

3. Results

3.1. pH- and radiation-induced anisotropic lattice changes

The implementation of the SIR method in the context of single-crystal work consists of recording X-ray diffraction data on both the native and the heavy-atom derivative samples. For each Bragg reflection, the measured intensity difference is used to generate phase information *via* the so-called Harker (1956) construction. However, with powder diffraction data, the problem of overlapping peaks will preclude the success of this approach for many reflections, especially at higher Bragg angles. The situation is aggravated by the fact that there are often small differences in the cell parameters between the native sample and the heavy-atom derivatives. Thus, a reflection that is well resolved in the native pattern may be overlapping in the derivative data or *vice versa*. One way to reduce this problem consists of inducing small anisotropic variations of the lattice parameters (not affecting the crystal and lattice symmetries), which can lead to the improved separation of accidentally overlapping reflections (Shankland *et al.*, 1997). In small-molecule crystallography, anisotropic thermal

expansion has been exploited to evaluate the individual contributions of reflections that are overlapping at one temperature but are resolved at another (Shankland *et al.*, 1997; Brunelli *et al.*, 2003). Performing measurements at several temperatures is more difficult with protein powders since the range of stability of these samples is usually fairly limited and the intrinsic diffraction quality (peak widths) can be very sensitive to changes in temperature. There are, however, other ways to induce anisotropic lattice changes.

For HEWL, a series of six samples was used, varying both the pH of the crystallization buffer and the concentration of the heavy-atom compound. In the tetragonal polymorph of HEWL there is a systematic dependence of the unit-cell parameters on the pH of the crystallization buffer (Basso *et al.*, 2005), and we exploited this variation to improve the resolution of accidentally overlapping peaks. In previous work (Girard *et al.*, 2002), the variation of the Gd site occupancies as a function of Gd-Hp-Do3A concentration in HEWL was studied. In the light of these results we selected 10 and 50 mM as representative values and prepared microcrystalline slurries at pH 3.5 and 4.5 for both heavy-atom derivative concentrations, as well as for a native sample. Fig. 1 shows portions of the HEWL powder data where shifts in peak position and intensity between the different samples are evident. Table 1 reports the measured lattice parameters as a function of pH and concentration of Gd-Hp-Do3A.

For PPE, we exploited X-ray-induced lattice changes that occur when the samples are irradiated by the high photon flux. The various effects caused by exposing samples to an intense X-ray beam are widely discussed topics in single-crystal protein crystallography, since they can seriously hamper structure determination (Garman & Nave, 2002). One common phenomenon is an irreversible radiation-induced lattice expansion, although the mechanism by which this occurs is not fully understood (Ravelli *et al.*, 2002). In many cases, the lattice changes are anisotropic and can thus be exploited to improve the resolution of overlapping reflections in a powder pattern (Von Dreele, 2007; Besnard *et al.*, 2007). For PPE, four native and five uranyl-derivative powder patterns were collected consecutively on the same sample volume. Each pattern therefore corresponds to a different exposure to X-rays. Fig. 2 shows portions of the low-angle region of the powder data where variations in peak position and intensity are apparent. Fig. 3 reports the variation of the unit-cell parameters as a function of X-ray irradiation.
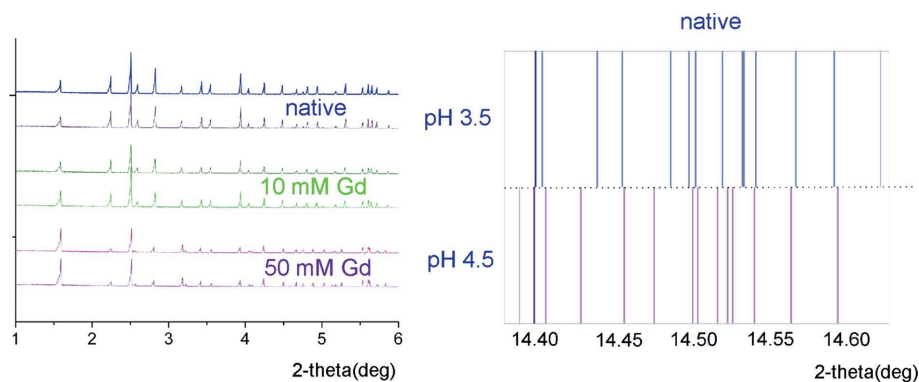
---

[1] The X-ray data have been deposited as supplementary material in the IUCr electronic archives. Reference: HX5065. Services for accessing these data are described at the back of the journal.

Figure 1
Effect of pH-induced anisotropic lattice changes on the position of diffraction from native HEWL and its Gd derivative. Left: Low-angle portion of the powder data recorded for the six samples of HEWL. In each case, the lower spectrum is for the sample crystallized at pH 4.5 and the upper for that at pH 3.5. Right: Effect of pH-induced anisotropic lattice changes on the position of diffraction peaks in the 2θ region 14.40–14.60° for native HEWL at the pH values of 3.5 and 4.5.
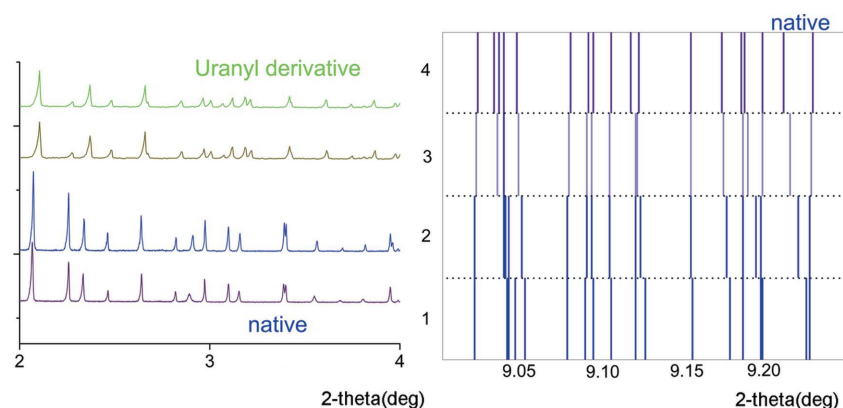

Figure 2
Effect of radiation-induced anisotropic lattice changes on the position of diffraction peaks from native PPE and its uranyl derivative. Left: Low-angle portion of the powder data recorded for native PPE and its uranyl derivative. In each case, the upper spectrum corresponds to a fresh sample and the lower spectrum corresponds to the same sample after extensive X-ray irradiation. Right: Effect of radiation-induced anisotropic lattice changes on the position of diffraction peaks in the 2θ region 9.00–9.25° for native PPE. The four different powder patterns correspond to sequential stages of X-ray irradiation.

3.2. Multi-pattern Pawley intensity extraction

Intensities were extracted by a multi-pattern Pawley (1981) refinement using the programs *PRODD* (Wright & Forsyth, 2000) for the HEWL data and *TOPAS-Academic* (Coelho, 2004) for the PPE data.[2] In this method, each diffraction pattern is calculated as a sum of overlapping reflections, the intensities of which are variables in a least-squares procedure. Thus, for each chemically distinct sample (native, heavy-atom derivative), a single set of intensities can be fitted to multiple data sets recorded for different values of pH (for HEWL) or at different stages of X-ray irradiation (for PPE). The statistics for the Pawley intensity extractions are presented in Table 2. This multi-pattern refinement allows one to exploit the partial separation of overlapping peaks that is brought about by

anisotropic lattice changes. Fig. 4 shows comparisons of structure factor amplitudes obtained from single-crystal measurements on PPE with data extracted from powder patterns. It is apparent that a substantial improvement in data quality is achieved if several powder patterns are combined in a multi-pattern Pawley extraction.

Further information about the data quality can be obtained from the least-squares matrix of the Pawley refinement (Sivia, 2000; Wright, 2004). The information content of a set of powder patterns is given by the eigenvalue spectrum of a reduced matrix constructed from the purely intensity-related elements of the full Pawley covariance matrix. In Fig. 5, we present the results of this kind of eigenvalue analysis for the data collected on the uranyl derivative of PPE. In the eigenvalue spectra, a clear dip is apparent at eigenvalues of about 10^{-3} in magnitude relative to the largest eigenvalue. The number of 'good pieces of intensity information' in a powder pattern (or in a set of powder patterns) is given by the number of eigenvalues of the reduced Pawley covariance matrix that have a significant magnitude, *i.e.* which are above the threshold of about 10^{-3} in our cases. These eigenvalues correspond to linear combinations of reflection intensities that are well determined independently of each other. It can be seen from Fig. 5 that the number of eigenvalues having a significant magnitude (*i.e.* the number of good pieces of intensity information) increases if several powder patterns are included in a multi-pattern Pawley extraction. A similar analysis for multi-pattern Pawley extraction with HEWL data has been presented by Basso *et al.* (2005).

Three sets of integrated intensities were computed for HEWL: one for the native and one for each Gd-Hp-Do3A derivative (at 10 and 50 mg ml^{-1} concentrations, respectively). For PPE, two sets of intensities were extracted: one for the native sample and one for the uranyl derivative. Values of the mean $I/\sigma(I)$ of the extracted intensities as a breakdown of Bragg spacing (resolution) are presented in Tables 3 and 4. The s.u. values on the intensities [$\sigma(I)$] correspond to the diagonal elements of the reduced error-covariance matrix obtained after convergence of the Pawley refinement. These values are thus reliable, but it must be kept in mind that, by only using s.u. values for individual reflections, the information

---

[2] The two data sets were processed in different laboratories by different groups of people, using different software packages. However, the results presented here do not critically depend on the use of one particular program or methods implementation.

about the error correlations is lost. This aspect will be further discussed in §4.1.

These intensity data were then treated as if they had been collected from a single crystal and no further modifications were made to account for the peak overlap problem.

3.3. Heavy-atom detection

Subsequent crystallographic computations were carried out with programs from the *CCP4* software suite. The derivative data were scaled against the native, and conventional single-crystal techniques were employed in all subsequent computations. For HEWL, the heavy atoms in the high-concentration (50 mM) Gd-Hp-Do3A derivative could be located by direct methods as implemented in the software *SHELXD* (Schneider & Sheldrick, 2002). The program delivered a solution for two Gd sites, which correspond to the known

positions of these atoms in the reported HEWL–Gd-Hp-Do3A structure (Girard *et al.*, 2002). For PPE, isomorphous difference Patterson maps revealed clear interatomic peaks in the Harker sections (see Fig. 6). These peaks are in agreement with a single U site in the asymmetric unit of the cell, the position of which was subsequently verified by single-crystal measurements.

3.4. Heavy-atom refinement and phasing

Heavy-atom refinement and phasing were performed by maximum likelihood techniques as implemented in the program *SHARP* (Bricogne *et al.*, 2003; La Fortelle & Bricogne, 1997). Given the limited resolution of the data, the atomic displacement parameters of the heavy atoms were held fixed during the refinement. For HEWL, the two Gd-Hp-Do3A data sets were treated as different derivatives.
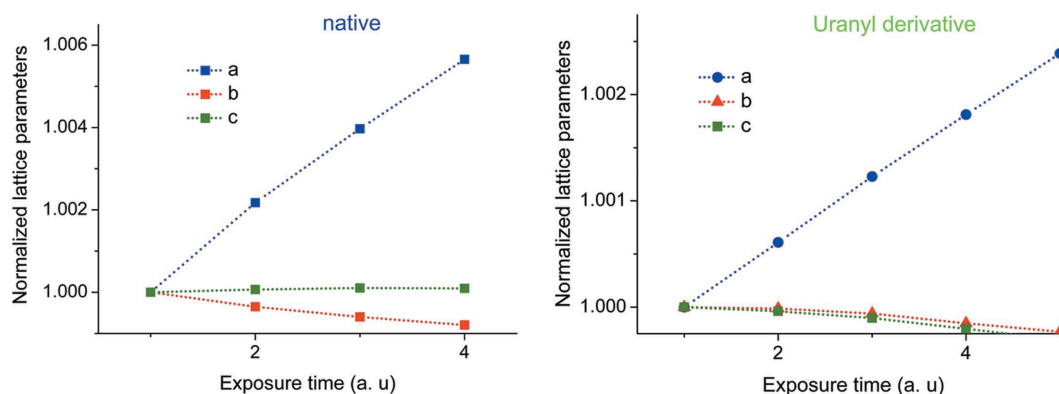

Figure 3
Radiation-induced anisotropic changes in the crystal lattices of native PPE (left) and of its uranyl derivative (right). Variation of lattice parameters as a function of X-ray irradiation for native PPE and its uranyl derivative. The absolute values for the lattice parameters at the origin (first data set) are $a = 51.865$, $b = 57.896$, $c = 75.308$ Å for the native data, and $a = 51.151$, $b = 57.979$, $c = 75.422$ Å for the uranyl derivative.
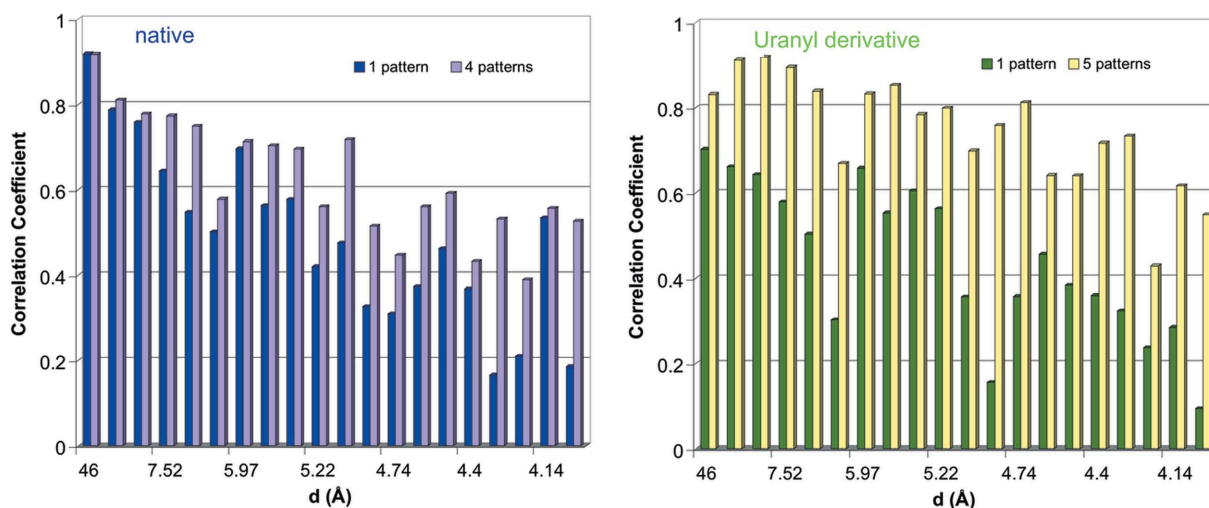

Figure 4
Effect of multi-pattern Pawley intensity extraction for native PPE (left) and its uranyl derivative (right). The graphs show, in shells of resolution, the correlation coefficients between structure factor amplitudes obtained from single-crystal measurements and those obtained from powder diffraction data by single- and multi-pattern Pawley extraction. For native PPE, the correlations reported are for the powder data extracted from one single pattern and for the series of four consecutive powder patterns pertaining to different radiation-induced lattice changes. For the uranyl derivative of PPE, the correlations reported are for the powder data extracted from one single pattern and for the series of five consecutive powder patterns pertaining to different radiation-induced lattice changes. A significant improvement in the powder data is observed when several patterns are used for the extraction of intensities.

Jonathan P. Wright *et al.* · Molecular envelopes

Table 3
Mean $I/\sigma(I)$ and number of reflections N_{ref} as a function of resolution of the extracted intensity data for native PPE and its uranyl derivative.

Resolution range (Å)		PPE native		PPE uranyl derivative	
		N_{ref}	$\langle I/\sigma(I) \rangle$	N_{ref}	$\langle I/\sigma(I) \rangle$
100.04	9.49	182	45.6	182	28.8
9.49	7.53	164	16.6	164	10.3
7.53	6.58	157	7.9	157	6.8
6.58	5.98	158	5.4	158	5.5
5.98	5.55	157	4.5	157	3.0
5.55	5.22	160	3.6	160	3.3
5.22	4.96	156	3.8	156	3.5
4.96	4.75	147	4.3	147	3.5
4.75	4.56	151	3.8	151	3.0
4.56	4.40	158	2.6	158	1.7
4.40	4.27	148	2.3	148	1.6
4.27	4.14	148	2.7	148	1.8
4.14	4.04	160	1.9	160	1.7
4.04	3.94	146	1.2	146	1.2
3.94	3.85	145	1.0	145	1.2
3.85	3.77	152	0.9	152	1.1
3.77	3.69	147	1.0	147	0.9
3.69	3.62	146	0.7	146	0.5
3.62	3.56	77	0.3	160	0.6
3.56	3.50	3	0.09	134	0.6

Table 4
Mean $I/\sigma(I)$ and number of reflections as a function of resolution of the extracted intensity data for native HEWL and its Gd derivatives.

Resolution range (Å)		HEWL native		HEWL Gd (10 mM)		HEWL Gd (50 mM)	
		N_{ref}	$\langle I/\sigma(I) \rangle$	N_{ref}	$\langle I/\sigma(I) \rangle$	N_{ref}	$\langle I/\sigma(I) \rangle$
100.05	8.41	146	41.8	147	33.3	149	32.9
8.41	6.67	131	15.6	132	11.5	130	12.2
6.67	5.83	128	12.7	126	9.1	126	8.6
5.83	5.30	124	8.4	126	6.4	127	5.5
5.30	4.92	120	8.6	119	6.4	120	5.8
4.92	4.63	121	6.5	122	5.4	127	4.9
4.63	4.39	121	7.6	119	6.5	118	6.0
4.39	4.20	118	8.7	119	7.0	123	6.2
4.20	4.04	131	6.8	131	5.1	126	5.0
4.04	3.90	118	4.8	117	3.4	119	3.0
3.90	3.78	108	3.8	110	3.2	111	2.4
3.78	3.67	129	3.2	128	2.7	121	2.4
3.67	3.57	118	2.5	119	2.4	119	2.1
3.57	3.49	109	2.0	109	1.8	116	1.4
3.49	3.41	124	2.2	124	2.1	123	2.0
3.41	3.33	119	2.2	121	2.0	116	1.9
3.33	3.27	117	1.9	115	1.7	114	1.7
3.27	3.21	110	1.7	108	1.7	117	1.6
3.21	3.15	113	1.6	117	1.4	96	1.4
3.15	3.10	7	1.2	100	1.3	14	1.4

However, the occupancy factors of the Gd atoms in the low-concentration (10 mM) data set refined to relatively low values, indicating that Gd binding was only weak in this derivative. This data set did not therefore substantially contribute to the generation of phase information. From the final heavy-atom models, phase probability distributions were computed for each reflection. The initial electron density map was computed by using as Fourier coefficients the centroids of the probability distribution of each structure factor. These maps are relatively noisy because of the bimodality of the phase probability distributions for acentric reflections (which represent the majority of reflections in noncentrosymmetric space groups) inherent in the SIR method. Correlation coefficients between these experimental maps and maps computed from the known molecular structures are reported in Table 5. Clearly, significant phase information has been generated by the SIR method for reflections up to about 8 Å resolution. Weaker, but nevertheless observable, phase information is generated for reflections up to about 6 Å resolution in the case of HEWL and to somewhat higher resolution for PPE.

3.5. Computation of solvent envelopes

In protein crystals, the macromolecules pack into an ordered crystalline structure leaving large channels that are filled by disordered solvent (between 30 and 80% of the total volume) (McPherson, 1999). The low-resolution structure factors contain the important information about the molecular envelopes, *i.e.* the boundaries between the protein molecules and the solvent. The local averages and fluctuations of the electron density are markedly higher in the regions occupied by the ordered protein molecules than in the solvent part. These properties can be exploited to delineate the molecular envelope, even in a noisy electron density map (Abrahams &

Leslie, 1996; Leslie, 1987; Wang, 1985). Such procedures are used in conventional single-crystal techniques to improve electron density maps by a cyclic procedure known as solvent
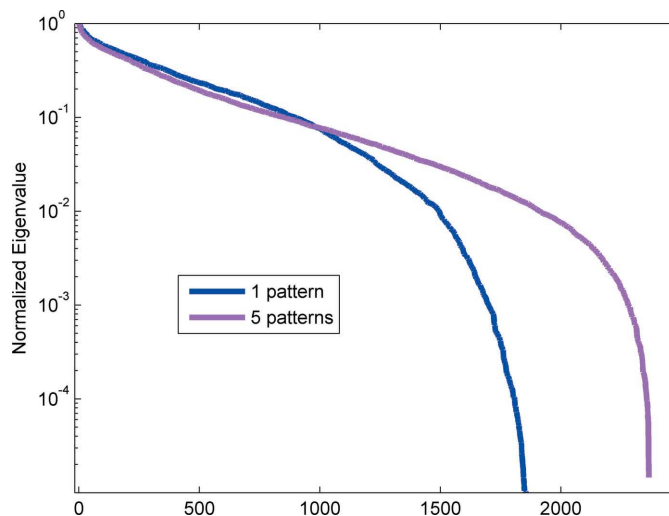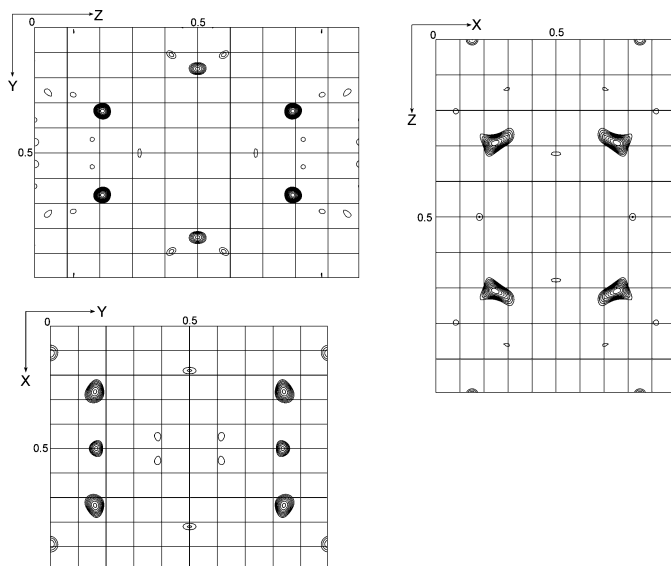

Figure 5
Effect of multi-pattern Pawley intensity extraction for the uranyl derivative of PPE. The figure shows the eigenvalue spectrum of a reduced matrix constructed from the purely intensity-related elements of the full Pawley covariance matrix. The eigenvalue spectra (normalized by the largest eigenvalue) are given for Pawley intensity extractions from one pattern (in blue) and from five patterns (in pink) pertaining to different radiation-induced lattice changes. The number of 'good pieces of intensity information' in a powder pattern (or in a set of powder patterns) is given by the number of eigenvalues that have a significant magnitude (Sivia, 2000; Wright, 2004). It can be seen from the graphs that there is a clear dip in the eigenvalue spectra of both data sets. However, for the data extracted from five patterns, the number of independent intensity data items is larger (about 2400 for a threshold set to 10^{-3}, as compared with about 1800 for the data extracted from one pattern).

Table 5
Phasing statistics of SIR phases computed from powder data for PPE and HEWL.

The table lists, in shells of resolution, the correlation coefficients (computed in reciprocal space) between maps computed with experimental SIR phases derived from powder data (using centroid structure factors) and maps computed from the known molecular structures of PPE (Protein Data Bank identifier 1lvy) and HEWL (Protein Data Bank identifier 6lyt).

Resolution (Å)		Map correlation coefficients of SIR phases	
from	to	PPE	HEWL
∞	12.36	0.6055	0.4842
12.36	9.84	0.4462	0.3988
9.84	8.61	0.3135	0.3740
8.61	7.82	0.3752	0.1466
7.82	7.26	0.4260	0.1845
7.26	6.84	0.4939	0.3012
6.84	6.50	0.4542	0.4218
6.50	6.21	0.3717	0.4251
6.21	5.98	0.2242	0.0401
5.98	5.77	0.3688	0.4308
5.77	5.59	0.5100	−0.0637
5.59	5.43	0.3636	0.4855
5.43	5.29	0.4014	0.1236
5.29	5.16	0.2881	−0.0995
5.16	5.04	0.2506	−0.1921
5.04	4.93	0.3737	0.0509
4.93	4.84	0.4233	−0.1892
4.84	4.74	0.2209	0.3677
4.74	4.66	0.1653	−0.2230
4.66	4.58	0.3051	−0.2444
4.58	4.51	0.3386	0.0142
4.51	4.44	0.1529	−0.2770
4.44	4.37	0.2020	0.1989
4.37	4.31	0.2901	−0.0067
4.31	4.25	0.2555	−0.0808
4.25	4.20	0.2309	−0.0262
4.20	4.15	0.1511	−0.0655
4.15	4.10	0.2533	0.1027
4.10	4.05	0.1553	0.2786
4.05	4.00	0.1065	0.3405

Figure 6
Isomorphous difference Patterson map of the uranyl derivative of PPE. The Harker sections $x = \frac{1}{2}$, $y = \frac{1}{2}$ and $z = \frac{1}{2}$ are shown. Contours are at intervals of 0.5σ, starting 3σ above the mean density (where σ is the r.m.s. of the map density). In all sections, clear Harker peaks, corresponding to the interatomic U–U vectors, can be seen.

flattening. The method amounts to applying a low-frequency filter to the electron density map, followed by the generation of a binary mask. We have used this method implemented in the *CCP4* programs *SOLOMON* and *DM* to generate solvent envelopes from the initial electron density maps. Figs. 7 and 8 show the superposition of the solvent masks with the known crystal structures for HEWL and PPE, indicating that the envelope obtained in this manner is essentially correct.

The cyclic density modification procedures implemented in these programs were also attempted but did not yield any significant improvement of the phases. This is not too surprising since, even in single-crystal studies and in the absence of noncrystallographic symmetry, the success of density modification techniques usually depends on the availability of at least some initial phase information at about 4–4.5 Å resolution.

4. Discussion

4.1. *De novo* phase determination from protein powder diffraction data: strengths and limitations

The development of *de novo* phasing methods for micro-crystalline protein samples offers new opportunities to obtain useful structural information for systems where the growing of large single crystals is problematic. We have shown that high-quality protein powder data can be collected and then processed as if they came from a single crystal. Two different schemes were exploited to induce anisotropic lattice changes in order to alleviate the problem of accidentally overlapping peaks. Detailed comparisons between the quality of the data obtained from single-pattern and multi-pattern Pawley extractions have demonstrated that the latter were crucial ingredients for obtaining the best possible data. Heavy-atom derivatives yielded SIR phase information up to about 6 Å resolution, which allowed the computation of molecular envelopes. While the experimental data extend to Bragg spacings of around 3.5 Å, the counting statistics and peak overlap were such that the phasing of the data at higher resolution is rather poor and the map is not of sufficient quality to reveal detailed structural features. Nevertheless, the information derived at lower resolution is essentially correct and indicates that, up to around 6 Å resolution, both the intensity extraction and the SIR phase determination have been successful. Our results show that a much wider range of samples are suitable for X-ray structure analysis than previously thought. In its present state, *de novo* structure determination from protein powder diffraction is a method that can be considered complementary to other low-resolution techniques that are commonly used in structural biology, such as electron microscopy and small-angle scattering.

While these results are promising, we must stress that the success of phase determination *via* the isomorphous replacement method is always limited by the intrinsic ambivalence regarding the correct chirality ('hand') of the structure (Harker, 1956). The true set of heavy-atom coordinates and a set where all coordinates are inverted will both yield equally

good solutions using Patterson or direct methods techniques. In macromolecular crystallography, there are two ways to resolve this twofold ambiguity. One is to exploit the anomalous scattering from the heavy atoms. However, since Friedel-related reflections exactly overlap in a powder diffraction pattern, this method is only applicable in single-crystal studies. The second method consists of simply determining the protein structure factor phases, starting in parallel from both heavy-atom substructure solutions. The electron density map computed with the phases derived from the correct heavy-atom substructure will show the true protein structure, whereas a map obtained from the inverted heavy-atom substructure will show a mirror image of the true protein structure where all of the α helices will be left-handed and all of the amino acids will be in the D rather than the L configuration. If a protein contains a β sheet, it will almost always have a distinct clockwise twist, which can also help to discriminate between the two alternative solutions. Thus, the
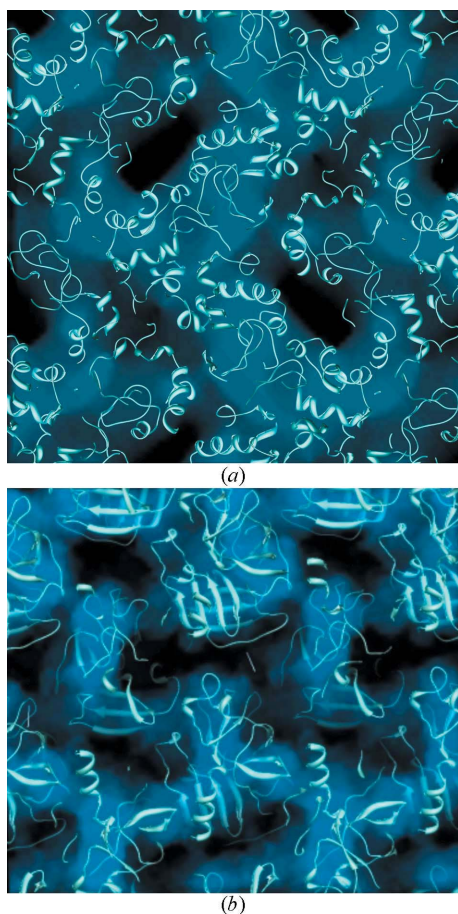
experimenter compares the two alternative maps and keeps the one with the correct protein configuration. However, in order to identify the handedness of helices, an electron density map at medium resolution (4.5 Å or higher) is required. This was not the case in the study reported here. Thus, although we have shown that heavy-atom detection works with powder data, we have voluntarily selected what we knew to be the correct hand of the heavy-atom substructure from the two possible choices. We therefore can only claim that we have
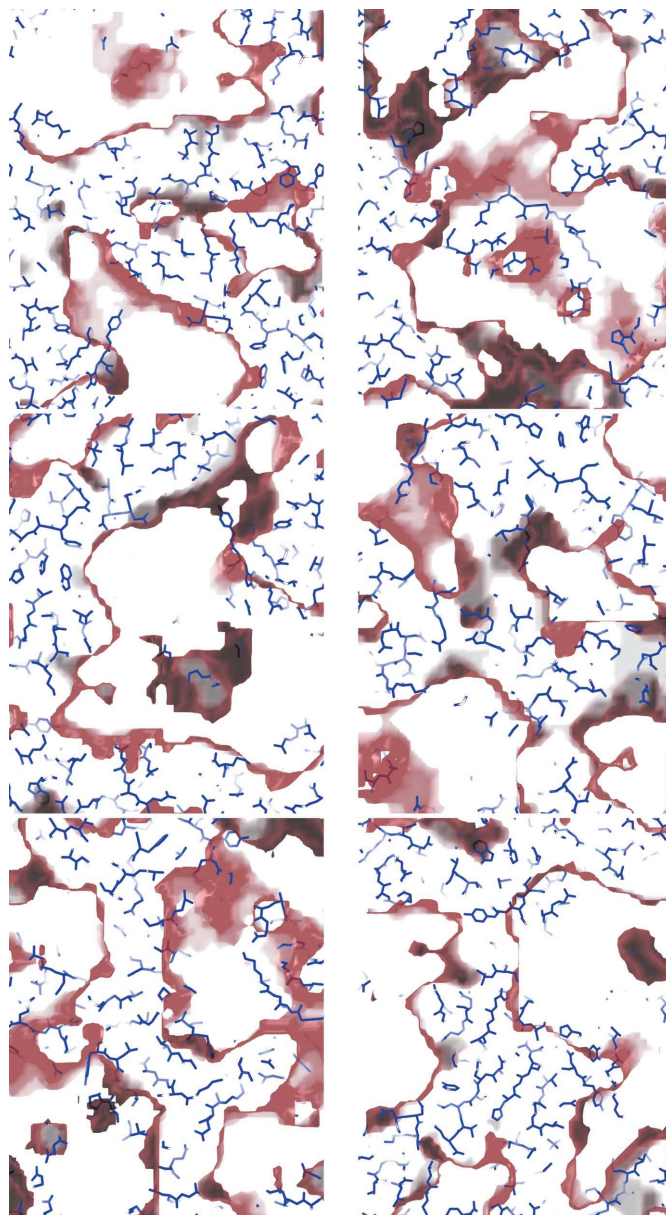

Figure 7
The molecular envelopes derived from the experimental data reported in this work. The envelopes are represented as semi-transparent surfaces. (*a*) Solvent mask for HEWL, viewed down the crystal *c* axis. The protein crystal structure (Protein Data Bank identifier 6lyt), represented as a main-chain ribbon model, is superimposed on this map. (*b*) Solvent mask for PPE, viewed down the crystal *c* axis (*a* vertical in-plane and *b* horizontal in-plane). The protein crystal structure (Protein Data Bank identifier 1lvy), represented as a main-chain ribbon model, is superimposed on this map.

Figure 8
The molecular envelope of PPE derived from the experimental data reported in this work. The envelope is represented as a semi-transparent red surface. Each figure represents a 5 Å slab along the crystal *a* axis and covers a whole unit cell (with *y* along the horizontal direction and *z* along the vertical direction). The various sections represented here are spaced by 8 Å along the *a* axis, from *x* = 0 (upper left) to *x* = 0.84 (lower right). The protein crystal structure (Protein Data Bank identifier 1lvy) is superimposed on this map.

successfully managed to phase low-resolution protein structures up to an ambiguity regarding the correct chirality.

The question then arises as to whether it will be possible to extend the resolution range of phases that can be derived from powder diffraction data. We suggest two lines of attack for future improvements.

(1) The implementation of the multiple isomorphous replacement (MIR) method by preparing additional heavy-atom derivatives (Harker, 1956). A serious drawback of the SIR method is that the phase determination for an acentric reflection (which form a majority in a protein data set) is not unique, and leads, in general, to a bimodal phase probability distribution. The standard approach is then to use the centroid of that distribution for further computations. Though this is shown to be the best possible (*i.e.* the least damaging) strategy (Blow & Crick, 1959), these centroid phases are often significantly in error. This explains why the map correlation coefficients of our experimental SIR phases do not exceed 0.6, even in the lowest-resolution shell (Table 2). In the MIR method, when data from additional heavy-atom derivatives are included, the phase probability distributions for most reflections become strongly unimodal, so that the initial phases for many reflections will be more accurate than was the case here. We therefore expect this to be a more favourable situation to exploit fully solvent-flattening and other density-modification techniques for improving and extending the phases to higher resolution.

(2) A major limitation of the present approach is that current single-crystal phasing methods are based on the assumption that the errors are uncorrelated between different reflection intensities. The likelihood criterion which is optimized in these methods is thus expressed as a product of factors for various reflections (La Fortelle & Bricogne, 1997). With protein powder data, where there is considerable peak overlap in the high-angle region, such an assumption may not be justified. While it is possible to extract individual reflection intensities from a powder pattern, the error distributions of such data reveal a pattern of strong covariances. Though the Pawley procedure returns all necessary information about

these correlations in the form of the inverse of the error covariance matrix (David, 2004; Wright, 2004), this knowledge is discarded when passing over the data to single-crystal phasing programs, which use a diagonal approximation. These approximations are, however, not essential (Bricogne, 1991) and expressions for generalized multivariate likelihood functions, capable of accommodating arbitrary patterns of covariances between the various sources of errors in reflection data, have been devised (Bricogne, 2000). There is thus scope for further improvements and we anticipate that the implementation of these ideas will increase the resolution of the maps that can be derived by *de novo* phasing of protein powder data.

A further possibility to expand on the present studies would be to combine powder diffraction data with *ab-initio* methods for low-resolution phasing of macromolecules that have been developed by Lunin *et al.* (2002) and others over recent years. Protein powder diffraction is indeed an ideal technique for accurately recording low-angle reflections. For various reasons, the low-angle data are not always measured accurately (if at all) during a standard data collection on a protein single crystal, and these data are not always thought to be essential. With powder diffraction, however, the low-angle part of the diffraction spectrum is generally collected with a high degree of accuracy. In addition, the problem of overlapping reflections is less severe in the low-angle region since the density of reflections in a powder spectrum increases with the cube of the distance from the reciprocal space origin. Thus, the specific disadvantages of powder *versus* single-crystal diffraction tend to be marginal at low resolution.

4.2. Use of low-resolution envelopes for characterizing the mesoporous properties of crystalline protein materials

Protein microcrystals can be used for various purposes (Margolin & Navia, 2001), including catalysis and drug delivery. Protein crystals are porous solids with large solvent channels that traverse the body of a crystal and allow the transport of substrates and products in and out of the crystal. These materials thus represent a unique class of molecular sieves and can be considered analogous to zeolites (Vilenchik *et al.*, 1998). A distinct advantage of protein crystals over other porous materials is the inherently chiral nature of protein molecules, which can be exploited in the separation of enantiomers and in enantioselective catalysis (Lalonde *et al.*, 1995; Persichetti *et al.*, 1995; St Clair *et al.*, 2000; Vuolanto *et al.*, 2003; Zelinski & Waldmann, 1997). The width and topology of the pores in a protein crystal can range from 10 Å to sometimes more than 100 Å and depend in large part on the way the molecules pack into a crystal lattice. These properties can change completely for different crystal polymorphs of the same molecule. To characterize these features, it is not necessary to determine the complete high-resolution structure of the protein molecule itself (which usually does not change substantially between different polymorphs), but rather it is important to map out the solvent channels in the crystals and to determine the packing of the molecules in the crystal lattice.
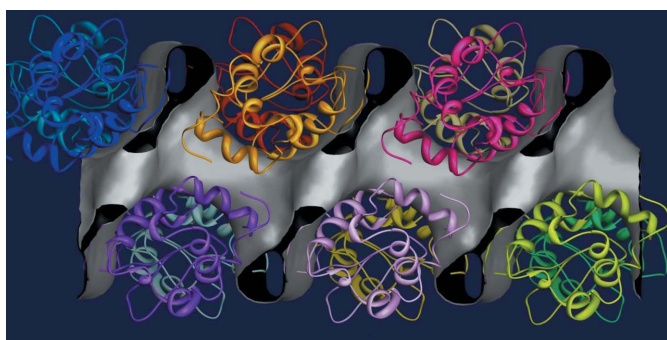

Figure 9
Solvent channels in HEWL crystals. The molecular envelope derived from the experimental data reported in this work is represented as a grey surface. The figure shows the linear solvent channel, which directly traverses the crystal parallel to the *c* axis (horizontal display direction). The protein crystal structure, represented as a main-chain ribbon model, is superimposed on this map.

The X-ray powder diffraction methods developed in the present study are thus likely to emerge as important analytical tools for investigating polycrystalline protein materials on the mesoscopic scale. From the experimentally derived molecular envelopes computed here, it can be seen that, for HEWL, linear solvent channels of approximately 15 Å diameter directly traverse the crystal parallel to the c axis (see Fig. 9), whereas the packing is relatively dense along the other directions, allowing only for narrow and tortuous channels. In PPE, there is no linear solvent channel that directly traverses the crystal. However, there are continuous spiral channels along all three crystallographic axes with smallest diameters ranging from about 8 to 14 Å. The shape and size of the pores are accurately represented by the experimentally derived molecular envelopes. It is interesting to note that the above-mentioned ambiguity regarding the correct chirality of a structure derived by SIR phasing does not prejudice this type of analysis since the sizes of the pores and solvent channels remain identical in a mirror-inverted image of the structure. Thus, the techniques described in this study may be ideally suited to study these systems and to deliver the essential structural data which will allow the properties of these interesting materials to be properly exploited.

5. Conclusion

The present study demonstrates that *de novo* phasing from protein powder diffraction data is feasible and can yield low-resolution structural information. With the development of these methods, microcrystalline precipitates, which were not considered useful up to now, can actually be used for structural investigations of proteins. These methods are likely to find important applications in the characterization of micro- and mesoporous biomaterials and also in structural biology.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Basso, S., Fitch, A. N., Fox, G. C., Margiolaki, I. & Wright, J. P. (2005). *Acta Cryst.* D**61**, 1612–1625.
Basu, S. K., Govardhan, C. P., Jung, C. W. & Margolin, A. L. (2004). *Expert Opin. Biol. Ther.* **4**, 301–317.
Besnard, C., Camus, F., Fleurant, M., Dahlström, A., Wright, J. P., Margiolaki, I. & Schiltz, M. (2007). *Z. Kristallogr. Suppl.* **26**, 39–44.
Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
Bricogne, G. (1991). *Acta Cryst.* A**47**, 803–829.
Bricogne, G. (2000). *Proceedings of the Workshop on Advanced Special Functions and Applications, Melfi (PZ), Italy, 9–12 May 1999*, edited by D. Cocolicchio, G. Dattoli & H. M. Srivastava, pp. 315–321. Rome: Arcane Editrice.
Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). *Acta Cryst.* D**59**, 2023–2030.
Brunelli, M., Wright, J. P., Vaughan, G. B. M., Mora, A. J. & Fitch, A. N. (2003). *Angew. Chem. Int. Ed.* **42**, 2029–2032.
Coelho, A. A. (2004). *TOPAS-Academic*. http://members.optusnet.com.au/alancoelho.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Coulibaly, F., Chiu, E., Ikeda, K., Gutmann, S., Haebel, P. W., Schulze-Briese, C., Mori, H. & Metcalf, P. (2007). *Nature (London)*, **446**, 35–37.
David, W. I. F. (2004). *J. Appl. Cryst.* **37**, 621–628.
Fitch, A. N. (2004). *J. Res. Natl Inst. Stand. Technol.* **109**, 133–142.
Fourme, R., Shepard, W., Schiltz, M., Ramin, M. & Kahn, R. (2000). *Structure and Dynamics of Biomolecules: Neutron and Synchrotron Radiation for Condensed Matter Studies*, edited by E. Fanchon, E. Geissler, J.-L. Hodeau, J.-R. Regnard & P. A. Timmins, pp. 36–78. Oxford University Press.
Garman, E. & Nave, C. (2002). *J. Synchrotron Rad.* **9**, 327–328.
Girard, É., Chantalat, L., Vicat, J. & Kahn, R. (2002). *Acta Cryst.* D**58**, 1–9.
Harker, D. (1956). *Acta Cryst.* **9**, 1–9.
La Fortelle, E. D. & Bricogne, G. (1997). *Methods in Enzymology*, Vol. 276, *Macromolecular Crystallography*, edited by R. M. Sweet & C. W. Carter Jr, pp. 472–494. New York: Academic Press.
Lalonde, J. J., Govardhan, C., Khalaf, N., Martinez, A. G., Visuri, K. & Margolin, A. L. (1995). *J. Am. Chem. Soc.* **117**, 6845–6852.
Leslie, A. G. W. (1987). *Acta Cryst.* A**43**, 134–136.
Leslie, A. G. W. (1993). *Proceedings of CCP4 Study Weekend on Data Collection and Processing*, pp. 44–51. Daresbury: SERC Daresbury Laboratory.
Lunin, V. Y., Lunina, N., Podjarny, A., Bockmayr, A. & Urzhumtsev, A. (2002). *Z. Kristallogr.* **217**, 668–685.
Margiolaki, I. & Wright, J. P. (2008). *Acta Cryst.* A**64**, 169–180.
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. Suppl.* **26**, 1–13.
Margiolaki, I., Wright, J. P., Fitch, A. N., Fox, G. C. & Von Dreele, R. B. (2005). *Acta Cryst.* D**61**, 423–432.
Margolin, A. L. & Navia, M. A. (2001). *Angew. Chem. Int. Ed.* **40**, 2205–2222.
McPherson, A. (1999). *Crystallization of Biological Macromolecules*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
Pawley, G. S. (1981). *J. Appl. Cryst.* **14**, 357–361.
Persichetti, R. A., St Clair, N. L., Griffith, J. P., Navia, M. A. & Margolin, A. L. (1995). *J. Am. Chem. Soc.* **117**, 2732–2737.
Ravelli, R. B. G., Theveneau, P., McSweeney, S. & Caffrey, M. (2002). *J. Synchrotron Rad.* **9**, 355–360.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
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.
St Clair, N., Wang, Y.-F. & Margolin, A. L. (2000). *Angew. Chem. Int. Ed.* **39**, 380–383.

Vilenchik, L. Z., Griffith, J. P., St Clair, N., Navia, M. A. & Margolin, A. L. (1998). *J. Am. Chem. Soc.* **120**, 4290–4294.

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

Von Dreele, R. B. (2001). *Acta Cryst.* D**57**, 1836–1842.

Von Dreele, B. (2003). *Methods in Enzymology*, Vol. 368, *Macromolecular Crystallography*, Part C, edited by R. M. Sweet & C. W. Carter Jr, pp. 254–267. San Diego: Academic Press.

Von Dreele, R. B. (2005). *Acta Cryst.* D**61**, 22–32.

Von Dreele, R. B. (2007). *J. Appl. Cryst.* **40**, 133–143.

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

Vuolanto, A., Kiviharju, K., Nevanen, T. K., Leisola, M. & Jokela, J. (2003). *Crystal Growth Des.* **3**, 777–782.

Wang, B. C. (1985). *Methods in Enzymology*, Vol. 115, *Diffraction Methods for Biological Macromolecules*, edited by H. Wyckoff, C. H. W. Hirs & S. N. Timasheff, pp. 90–112. New York: Academic Press.

Wright, J. P. (2004). *Z. Kristallogr.* **219**, 791–802.

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.

Zelinski, T. & Waldmann, H. (1997). *Angew. Chem. Int. Ed. Engl.* **36**, 722–724.