# research papers

Acta Crystallographica Section D Biological Crystallography ISSN 0907-4449

Sebastian Basso,<sup>a</sup>\* Céline Besnard, $a,b$  Jonathan P. Wright, $c$ Irene Margiolaki,<sup>c</sup> Andrew Fitch, $c$  Philip Pattison<sup>a,d</sup> and Marc Schiltz<sup>a</sup>

<sup>a</sup>École Polytechnique Fédérale de Lausanne (EPFL), Laboratoire de Cristallographie, CH-1015 Lausanne, Switzerland, <sup>b</sup>Université de Genève (UNIGE), 24 Rue du Général-Dufour, CH-1211 Genève 4, Switzerland, <sup>c</sup>European Synchrotron Radiation Facility (ESRF), 6 Rue Jules Horowitz, F-38043 Grenoble CEDEX 9, France, and <sup>d</sup>Swiss–Norwegian Beamlines (SNBL) at the ESRF, 6 Rue Jules Horowitz, F-38043 Grenoble CEDEX 9, France

Correspondence e-mail: sebastian.basso@epfl.ch Features of the secondary structure of a protein molecule from powder diffraction data

Protein powder diffraction is shown to be suitable for obtaining de novo solutions to the phase problem at low resolution via phasing methods such as the isomorphous replacement method. Two heavy-atom derivatives (a gadolinium derivative and a holmium derivative) of the tetragonal form of hen egg-white lysozyme were crystallized at room temperature. Using synchrotron radiation, high-quality powder patterns were collected in which pH-induced anisotropic lattice-parameter changes were exploited in order to reduce the challenging and powder-specific problem of overlapping reflections. The phasing power of two heavyatom derivatives in a multiple isomorphous replacement analysis enabled molecular structural information to be obtained up to approximately  $5.3 \text{ Å}$  resolution. At such a resolution, features of the secondary structure of the lysozyme molecule can be accurately located using programs dedicated to that effect. In addition, the quoted resolution is sufficient to determine the correct hand of the heavy-atom substructure which leads to an electron-density map representing the protein molecule of proper chirality.

## 1. Introduction

In X-ray crystal diffraction studies of the three-dimensional molecular structure of biologically important macromolecules, the crystallization step is crucial and can prove to be a timeconsuming trial-and-error process. Single crystals are almost always the sought-after product of a crystallization procedure, given the success of the well established method of singlecrystal diffraction. Recently, however, interest has grown in the technique of powder diffraction, which makes use of the once-discarded microcrystalline precipitates. These microcrystalline powders can be obtained for many proteins by batch precipitation (Von Dreele, 2003; Margiolaki, Wright, Fitch et al., 2007) and the crystallites making up the powder are often far too small for single-crystal diffraction experiments. In addition, the powder technique is becoming a well recognized method in the field of structure refinement (Von Dreele, 1999; Margiolaki & Wright, 2008) and molecular replacement (Doebbler & Von Dreele, 2009; Margiolaki, Wright, Wilmanns et al., 2007) for relatively small biological macromolecules. Nevertheless, both of these methods require the availability and use of sufficiently similar protein structures as molecular models. In structure refinement the positions of the atoms in the model are refined, while in molecular replacement the model itself is placed within the unit cell in order to determine its correct position and orientation.

 $\odot$  2010 International Union of Crystallography Printed in Singapore – all rights reserved

Received 20 November 2009 Accepted 22 March 2010

#### Table 1

Data-acquisition parameters and lattice parameters for the three sets of HEWL samples: native (Nat-HEWL), gadolinium derivative (Gd-HEWL) and holmium derivative (Ho-HEWL).





Alternatively, de novo methods provide a means of obtaining structural information when 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 the protein molecules with heavy atoms (the isomorphous replacement method) and/or by the phenomenon of anomalous X-ray scattering (Fourme et al., 2000). Phase information is individually generated for each Bragg reflection through relatively small intensity modulations. Hence, the success of applying these methods to powder diffraction data will eventually be limited by the powderspecific problem of overlapping reflections, whether accidental or owing to symmetry.

In this article the need to assess the quality of the structural information that can be extracted from protein powder diffraction data is addressed. This assessment can only be achieved if no prior knowledge of the system under study is taken into account through the use of molecular models. Consequently, the de novo method of isomorphous replacement was employed. In a previous study (Wright et al., 2008), it was shown that the single isomorphous replacement (SIR) method can successfully be applied to protein powder data and that the resulting phases are of sufficient quality to enable the computation of a molecular envelope of the protein molecule. The results obtained on adding a second heavyatom derivative to the analysis, switching to multiple isomorphous replacement (MIR), are presented here. Owing to the greater phasing power of the MIR data, the accuracy of the generated phase information increased, enabling the electrondensity map representing a molecular envelope to develop into one in which elements of the secondary structure of the protein molecule can be located.

# 2. Experimental procedures

Two heavy-atom derivatives of hen egg-white lysozyme (HEWL) were crystallized for this study. The HEWL protein molecule has a molecular weight of  $14.6 \text{ kg mol}^{-1}$ . The first

derivative contained Gd atoms bound to the protein molecule, while the second contained Ho atoms. A complete experimental description for the native and gadolinium-containing HEWL samples, including X-ray data collection on beamline ID31 (Fitch, 2004) at the European Synchrotron Radiation Facility (ESRF), has previously been published in Wright et al. (2008).

#### 2.1. Crystallization

HEWL was purchased from Sigma– Aldrich in lyophilized form and was used without further purification. All polycrystalline samples for each of the derivatives were synthesized by cocrys-

tallization using the standard salting-out procedure in batch mode, i.e. protein solution was mixed with heavy-atom complex/salt-containing solution and left to crystallize. The gadolinium-derivative samples were prepared at two different pH values using the complex Gd-HPDO3A (Girard et al., 2002) at a concentration of 50 mM. Native samples were also crystallized at both pH values. For the holmium derivative, the heavy-atom-containing compound used was HoCl<sub>3</sub>. It was diluted to a concentration of 400 mM in 0.05 M sodium acetate buffer containing 1.2 M NaCl and subsequently added to the protein solution. The samples were crystallized at pH values of 4.5 and 5.5. In this pH range HEWL crystallizes in the tetragonal space group  $P4<sub>3</sub>2<sub>1</sub>2$  (Basso *et al.*, 2005).

#### 2.2. X-ray data collection

High-resolution powder diffraction data were collected at room temperature (291 K) from a total of six HEWL polycrystalline samples: one native and two derivatives (Gd and Ho), each at two different pH values. The samples were loaded into a 1.5 mm diameter glass capillary and centrifuged so as to enhance the crystallite packing (Von Dreele, 2003). Most of the excess mother liquor was removed and the capillaries were sealed with wax to prevent dehydration.

The holmium derivatives were measured on the Swiss– Norwegian Beamline (SNBL) BM01B at the ESRF at a wavelength of  $0.8000(1)$  Å. Situated on a bending magnet, BM01B has a considerably lower photon flux compared with beamline ID31, which is located on an insertion device. As a result, radiation damage on BM01B is not significant enough to justify translating the capillary as was necessary for the native and Gd-derivative samples measured on ID31 (Wright et al., 2008). Scans that were unaffected by radiation damage for a given sample were merged together in order to improve the counting statistics. Each beamline is equipped with analyser crystals: six on BM01B and nine on ID31 (Margiolaki, Wright, Fitch et al., 2007). Consequently, they both provided data with a very high angular resolution with a diffraction line width as low as  $0.01^{\circ}$  FWHM, the instrumental contribution to which is roughly  $0.003^{\circ}$ . A summary of the experimental parameters and the lattice parameters for the six samples used in this study can be found in Table 1.

### 3. Results

#### 3.1. Intensity extraction

For this project, the intensities of the peaks in the powder diagrams, once extracted using the Pawley method (Pawley, 1981), were treated as though they originated from a singlecrystal experiment. This enables the use of well established biomacromolecular single-crystal software to perform the isomorphous replacement analysis. As a result, the process of intensity extraction is an important one. However, it is one that is hampered by the powder-specific problem of peak overlap. To partially overcome this problem, the method of multi-pattern Pawley intensity extraction can be employed. This method exploits the anisotropic variation in the lattice parameters, and in turn in the position of peaks, induced by various experimental parameters such as radiation damage or pH of the crystallization solution (Basso et al., 2005; Besnard et al., 2007; Von Dreele, 2007). The multi-pattern Pawley method was applied to the data using the program TOPAS-Academic (Coelho, 2004). 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 a single set of intensities can be fitted to multiple data sets, while benefiting from the effects of the change in the pH of the crystallization solutions. An in-



#### Figure 1

Effect of multi-pattern Pawley intensity extraction for the HEWL samples. Each curve represents an eigenvalue spectrum of a reduced matrix constructed from the purely intensity-related elements of the full Pawley normal matrix. The spectra (normalized by the largest eigenvalue) are given for Pawley intensity extractions from one pattern (at one pH value) and two patterns (at two different pH values) for each sample type: native (Nat), gadolinium derivative (Gd) and holmium derivative (Ho). The quantity of 'good pieces of intensity information' in one or a set of powder patterns is given by the number of eigenvalues that have a significant magnitude (Sivia, 2000; Wright, 2004).

depth description of the method of multi-pattern Pawley refinement for intensity extraction of powder diffraction peaks can be found in the above-mentioned studies of Wright et al. (2008) and Von Dreele (2007).

The information content in powder patterns is given by the eigenvalue spectrum of a reduced matrix constructed from the purely intensity-related elements of the full Pawley normal matrix (Sivia, 2000; Wright, 2004). Hence, the improvement in the information content upon using the multi-pattern Pawley intensity-extraction method can be quantified by comparing the eigenvalue spectra generated from the intensities extracted from a single pattern and from multiple patterns. Fig. 1 shows such a comparison for the three sets of HEWL samples measured in this study. The number of eigenvalues that have a significant magnitude, which corresponds to those appearing before a discernible dip in the spectrum, determines the quantity of 'good pieces of intensity information' in the corresponding set of intensities. Therefore, from Fig. 1 it is possible to conclude that the method of multi-pattern Pawley intensity extraction successfully increased the information content in the intensities extracted for the native and holmium-derivative samples and considerably less so for the gadolinium derivative. The difference in the success rate of the extraction method can be accounted for by the percentage change in the lattice constants upon varying the pH of the crystallization solution (Table 1). In other words, for the gadolinium derivative the anisotropic change in the lattice parameters is not significant enough to profit from the multipattern Pawley intensity-extraction method. This observation demonstrates that the improvement in quality of the extracted intensities is a direct consequence of the anisotropic shift in lattice parameters and is not simply a consequence of the addition of multiple patterns.

#### 3.2. Heavy-atom detection

Prior to the detection of heavy atoms within the structure of a protein molecule in a heavy-atom derivative for an MIR experiment, the extracted intensities from the various powder spectra have to be adjusted to a common scale. In this study, the scaling of the native and two derivative intensity data sets was performed in the CCP4 software package (Collaborative Computational Project, Number 4, 1994). Within this same software suite, programs employing various conventional single-crystal methods were used for crystallographic computations.

Using direct methods as implemented in the SHELXD software (Sheldrick, 2008), two gadolinium sites were found in the first heavy-atom derivative. Using these positions, a single isomorphous replacement analysis was performed. The experimental phase information obtained from the SIR analysis was in turn used to compute an isomorphous difference Fourier map in order to determine the atomic positions of the holmium sites in the second derivative. In this manner, one site was identified. Upon comparison with the known molecular model of HEWL, the Ho atom was discovered to be bound to the side chain of the Asp52 residue situated in the

protein catalytic cleft. This site is in agreement with the second most occupied site reported by Jakoncic et al. (2006). The gadolinium sites were also found to correspond to previously published atomic positions (Girard et al., 2002).

#### 3.3. Heavy-atom refinement and phasing

In the same way as for the SIR experiment, the refinement and phasing of the heavy atoms using both derivatives for the MIR method were performed using maximum-likelihood techniques incorporated in the program SHARP (Bricogne et al., 2003; de La Fortelle & Bricogne, 1997). The refinement was carried out using three heavy-atom sites: two for gadolinium (Gd-1 and Gd-2) and one for holmium (Ho-1). The site occupancies refined to approximately 0.60, 0.45 and 0.45 for Gd-1, Gd-2 and Ho-1, respectively. The subsequent map was calculated using the centroids of the probability distribution of each structure factor afforded by SHARP as Fourier coefficients. The correlation coefficients between this map (Fig. 2; MIR1) and that computed from the known molecular structure indicate that phase information has been generated up to about  $5.3 \text{ Å}$ .

### 3.4. Density modification

Taking the analysis further, density-modification methods were used in order to improve the quality of the protein phase angles and in turn that of the electron-density map. Two different density-modification programs, both of which are included in the CCP4 package, were utilized: DM (Cowtan, 1994) and SOLOMON (Abrahams & Leslie, 1996). Independently, each algorithm improved the phase information obtained from SHARP. However, combining the two resulted in a comparatively greater improvement. In light of this, the best results were generated by first launching the algorithm DM in histogram-matching mode, completing two cycles, followed by one cycle of  $SOLOMON$  running in solvent-



#### Figure 2

Correlation coefficients (computed in reciprocal space) between maps computed from the known molecular structure of HEWL (PDB code 6lyt; Young et al., 1993) and maps computed with experimental phases: SIR data (SIR), MIR data (MIR1) and MIR data after density modification (MIR2).

flipping mode. Although alternative routines in which more than one cycle of SOLOMON was used were attempted, the general trend seemed to be that increasing the number of cycles performed somewhat deteriorated the quality of the derived phases. Nevertheless, the effectiveness of the chosen routine can be observed when comparing the columns for MIR1 and MIR2 in Fig. 2. The subsequent electron-density map computed using the improved phase angles is shown in Fig. 3.

### 3.5. Detection of secondary-structure elements

In most cases, it is possible to locate elements of the secondary structure of a protein molecule, such as helices, at a resolution of  $6 \text{ Å}$  or higher. Considering that the phase determination was successful to a resolution of approximately 5.3  $\AA$ , a search for helices was perfomed using the map after density modification. The search was carried out with the program FFFEAR (Cowtan, 1998), also included in the CCP4 package, which uses 'representative' maximum-likelihood search targets to detect molecular fragments in electrondensity maps. More specifically, the search targets provided are helical fragments of nine residues at various resolutions. Using a  $6 \text{ Å}$  resolution helical search target, FFFEAR found all four helices present in HEWL (Blake et al., 1967). The results of the search are shown in Fig. 4, in which the thickness of each helix is representative of its score (an agreement function based on the mean-squared difference between the model and map over a masked region): the thicker the helix, the lower its score and the more accurate the fit. When compared with Fig. 3, it can be seen that the helices found by the search program are well located.

#### 3.6. Handedness of protein molecules

The chirality ('hand') of protein molecules can be a hurdle in X-ray crystallographic structural analyses which employ the method of isomorphous replacement (Harker, 1956) not supplemented by anomalous scattering. A set of heavy-atom coordinates and its inverse set will yield equally good solutions using Patterson or direct-methods techniques, giving rise to a twofold ambiguity. One way to determine which set will ultimately generate the electron-density map representing the true structure of the protein molecule is to closely inspect the handedness of  $\alpha$ -helices. In an isocontour representation of the electron density, the direction of helices only starts to be visible to the naked eye at a resolution of approximately  $4.5 \text{ Å}$ . However, secondary-structure-element search programs profit from the full information contained in an electron-density map and therefore the prospect of the resolution achieved being sufficient for the search program to discriminate between the two solutions was investigated.

The phase information generated as part of the SIR experiment was obtained using the correct hand of the gadolinium structure (Wright et al., 2008) and hence the subsequent isomorphous difference Fourier map in turn revealed the holmium atomic structure with the correct handedness. The procedures described in  $\S$ 3.3–3.5 were repeated while inverting the atomic coordinates of both derivatives and switching to the enantiomorphic space group  $P_{12,12}$ . Using the same helical search target, only the three longest helices of the four in HEWL were found by FFFEAR. The fact that not all helices were found tends to indicate that the search was less successful when using the mirror image of the heavy-atom substructures in the phasing procedure. This was confirmed by the range of scores corresponding to each search: 69.51–70.91 versus 84.50–86.68 for the searches with the correct and inverted coordinates, respectively. As a result, the information content at  $\sim$ 5.3 Å resolution seems to be sufficient to determine which heavy-atom coordinates lead to the electron density representing the HEWL molecule of correct chirality.

# 4. Discussion

The application of the multiple isomorphous replacement method to protein X-ray powder diffraction data allowed the quality of the extracted information to be evaluated. Considering the extracted intensities as arising from a singlecrystal experiment allowed the use of established singlecrystal software throughout the data analysis. As a result, substantial structural information was obtained from microcrystalline protein samples, thereby reinforcing the complementary nature of the technique with respect to single-crystal X-ray diffraction. It was shown that the addition of a second heavy-atom derivative to the analysis, switching from SIR to MIR, enabled the recovery of phase information to a significantly greater extent. While the SIR data produced an accurate molecular envelope describing the globular shape of the HEWL molecule, an electron-density map in which elements of the secondary structure of the protein molecule could be correctly located was obtained through the MIR analysis. Additionally, the MIR data were of sufficient quality to distinguish which set of heavy-atom coordinates ultimately leads to an electron-density map representing the HEWL



#### Figure 3

Electron-density map for HEWL (green) obtained from MIR data after density modification superimposed onto the known molecular structure (orange) obtained from the PDB (PDB code 6lyt; Young et al., 1993).

molecule of correct chirality. Although this can also be performed by using the Friedel symmetry-breaking effects of anomalous scattering, it is an unsuitable option for powder diffraction data given that Friedel-related reflections exactly overlap in a powder pattern. In a multiple-wavelength experiment, however, the wavelength-dependent (dispersive) intensity modulations could be exploited in the phasing process of the analysis (Helliwell et al., 2005).

Although most of the programs utilized in this study are part of the CCP4 software suite, alternative programs and algorithms were tested in various steps of the analysis, such as scaling and density modification. In some cases, the alternative results were better and in others they were worse, but they were rarely equivalent to those presented here. Although different algorithms may be more or less tolerant of the errors introduced by overlapping peaks, it was not always possible to justify these discrepancies as simply being the consequence of the fact that a different algorithm was applied to the data. However, the underlying fact that single-crystal programs are in the end not entirely adequate for treating powder diffraction data can most probably account for more than just the observed disparities. Single-crystal phasing methods are based on the assumption that the errors are uncorrelated between different reflection intensities or amplitudes. With powder diffraction data, where the intensity data are extracted from patterns which can display substantial peak overlap, this assumption is not justified. Hence, the quality of the phase information could potentially be much better if protein powder-specific programs existed which could make use of the error covariances between reflection intensities. In a similar way, algorithms for extracting reliable structural information from medium-resolution data  $(4.5–8 \text{ Å})$  are, as yet, still relatively scarce (Karmali et al., 2009; Brunger et al., 2009). However, this is expected to change as the gap between electron micoscopy and high-resolution X-ray diffraction is bridged.





The four helices (orange) found by the program FFFEAR superimposed on the electron-density map for HEWL (green) obtained from MIR data after density modification. The thickness of each helix is representative of the accuracy of the fit between the search target and the electron density: the thicker the helix, the lower its score and the more accurate the fit.

Upon examination of the map correlation coefficients displayed in Fig. 2, it is apparent that switching from SIR to MIR not only improved the values where SIR had been successful, but the resolution limit at which significant phase information was derived increased from  $\sim$ 6 Å to  $\sim$ 5.3 Å. Provided that a third heavy-atom derivative of HEWL could be crystallized, including it in the phasing process would in theory improve the quality of the recovered phase-angle information. However, it is expected that the resolution limit reached in this study is primarily a consequence of the powder-specific problem of overlapping peaks rather than the lack of any additional heavy-atom derivatives.

We acknowledge financial support from the Swiss National Science Foundation through grant Nos. 200021-107637/1 and 200021-113339/1. We thank the European Synchrotron Radiation Facility and the Swiss–Norwegian Beamline Consortium for provision and allocation of beam time in the framework of the Long-Term Proposal No. CH-1985 (Development of Synchrotron Powder Diffraction Methods for Biomolecules).

#### References

- Abrahams, J. P. & Leslie, A. G. W. (1996). Acta Cryst. D52, 30–42.
- Basso, S., Fitch, A. N., Fox, G. C., Margiolaki, I. & Wright, J. P. (2005). Acta Cryst. D61, 1612–1625.
- Besnard, C., Camus, F., Fleurant, M., Dahlström, A., Wright, J. P., Margiolaki, I., Pattison, P. & Schiltz, M. (2007). Z. Kristallogr. Suppl. 26, 39–44.
- Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967). Proc. R. Soc. Lond. B Biol. Sci. 167, 365–377.
- Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). Acta Cryst. D59, 2023–2030.
- Brunger, A. T., DeLaBarre, B., Davies, J. M. & Weis, W. I. (2009). Acta Cryst. D65, 128–133.
- Coelho, A. (2004). TOPAS-Academic. http://www.topas-academic.net/.
- Collaborative Computational Project, Number 4 (1994). Acta Cryst. D50, 760-763.
- Cowtan, K. (1994). Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr. 31, 34–38.
- Cowtan, K. (1998). Acta Cryst. D54, 750–756.
- Doebbler, J. A. & Von Dreele, R. B. (2009). Acta Cryst. D65, 348–355.
- 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.
- Girard, É., Chantalat, L., Vicat, J. & Kahn, R. (2002). Acta Cryst. D58, 1–9.
- Harker, D. (1956). Acta Cryst. 9, 1–9.
- Helliwell, J. R., Helliwell, M. & Jones, R. H. (2005). Acta Cryst. A61, 568–574.
- Jakoncic, J., Di Michiel, M., Zhong, Z., Honkimaki, V., Jouanneau, Y. & Stojanoff, V. (2006). J. Appl. Cryst. 39, 831–841.
- Karmali, A. M., Blundell, T. L. & Furnham, N. (2009). Acta Cryst. D65, 121–127.
- La Fortelle, E. de & Bricogne, G. (1997). Methods Enzymol. 276, 472–494.
- Margiolaki, I. & Wright, J. P. (2008). Acta Cryst. A64, 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., Wilmanns, M., Fitch, A. N. & Pinotsis, N. (2007). J. Am. Chem. Soc. 129, 11865–11871.
- Pawley, G. S. (1981). J. Appl. Cryst. 14, 357–361.
- Sheldrick, G. M. (2008). Acta Cryst. A64, 112–122.
- Sivia, D. S. (2000). J. Appl. Cryst. 33, 1295–1301.
- Von Dreele, R. B. (1999). J. Appl. Cryst. 32, 1084–1089.
- Von Dreele, R. B. (2003). Methods Enzymol. 368, 254–267.
- Von Dreele, R. B. (2007). J. Appl. Cryst. 40, 133–143.
- 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. (2008). J. Appl. Cryst. 41, 329–339.
- Young, A. C. M., Dewan, J. C., Nave, C. & Tilton, R. F. (1993). *J. Appl.* Cryst. 26, 309–319.