

# The role of wavelength and source in the search for sulfur-atom positions evaluated in two case studies: lysozyme at room temperature and cryo apocrustacyanin A1

Michele Cianci,<sup>a,b,\*</sup> John R. Helliwell,<sup>a</sup> David Moorcroft,<sup>a</sup> Andrzej Olczak,<sup>a,c,†</sup> James Raftery<sup>a</sup> and Pierre J. Rizkallah<sup>b</sup>

<sup>a</sup>Department of Chemistry, The University of Manchester, Brunswick St., Manchester M13 9PL, UK,

<sup>b</sup>CCLRC Daresbury Laboratory, Daresbury, Warrington, Cheshire WA4 4AD, UK, and <sup>c</sup>Institute of General and Ecological Chemistry, Technical University of Lodz, Zwirki 36, 90-924, Poland.

Correspondence e-mail: m.cianci@dl.ac.uk

Synchrotron radiation beamlines offer automated data collection with faster and larger detectors, a choice of wavelength(s), intense beams and fine collimation. An increasing output of protein crystal structures sustains an interest in streamlining data collection protocols. Thus, more and more investigators are looking into the use of the anomalous signal from sulfur to obtain initial phase information for medium-size proteins. This type of experiment ideally requires the use of synchrotron radiation, softer X-rays and cryocooling of the sample. Here the results are reported of an investigation into locating the weak, *i.e.* sulfur, anomalous scatterers in lysozyme using rotating anode or synchrotron radiation data recorded at room temperature. It was indeed possible to locate the sulfur atoms from a lysozyme crystal at room temperature. Accurate selection of images during scaling was needed where radiation damage effects were detected. Most interestingly, comparisons are provided of high-redundancy data sets recorded with synchrotron radiation at  $\lambda = 2.0$  and  $1.488 \text{ \AA}$ , and with Cu  $K\alpha$  and Mo  $K\alpha$  radiation. Apocrustacyanin A1 was also investigated; from the results of a very high redundancy data collection using softer synchrotron X-rays and a cryo-cooled crystal, it was possible to find the sulfur atoms.

© 2004 International Union of Crystallography  
Printed in Great Britain – all rights reserved

## 1. Introduction

The past decade has certainly seen macromolecular crystallography evolving from a quite sophisticated academic curiosity to a mature industrial tool for drug discovery. At synchrotrons, the availability of tunable beamlines with highly intense and collimated beams allows scientists, from most countries in most continents, to collect more data with more phasing power.

However, it is also clear that, while anomalous data collection benefits from the use of synchrotrons, it remains true that anomalous data can be obtained from laboratory sources. For example, recent work has highlighted the use of Cu  $K\alpha$  anode sources to measure such data, provided that the data are good and highly redundant (Kwiatkowski *et al.*, 2000; Yang & Pflugrath, 2001; Lemke *et al.*, 2002; Debreczeni, Bunkoczi, Girmann & Sheldrick, 2003; Debreczeni, Bunkoczi, Ma *et al.*, 2003; Debreczeni, Bunkoczi, Zeeck *et al.*, 2003).

Moreover, Cr  $K\alpha$  anodes are available, allowing data collection at a softer wavelength with a 'home' laboratory source (Kwiatkowski *et al.*, 2000). An overview of the aspects of softer X-rays is given by Helliwell (2004).

Thus, anomalous dispersion as a tool is still very important for both synchrotron and laboratory X-ray use. The location of anomalous scatterers from anomalous differences has a long history, either using anomalous Patterson maps (Rossmann & Blow, 1962) or direct methods (Mukherjee *et al.*, 1989). The problem of locating anomalous scatterers has been extended to more complicated cases, such as many selenium atoms in the Se-Met MAD method (Hendrickson *et al.*, 1990), by the direct methods approach in programs like *SnB* (Weeks & Miller, 1999) or *SHELXD* (Schneider & Sheldrick, 2002), or by applying Patterson techniques as in *CNS/SOLVE* (Brunger *et al.*, 1998) or *SOLVE* (Terwilliger & Berendzen, 1999).

In the structure solution of apocrustacyanin A1 (Cianci *et al.*, 2001), seeking the sulfur positions was a main goal. The practice and the lessons of that work are reported in detail here. In parallel with these efforts, a xenon derivative was found. Nevertheless, these apocrustacyanin A1 native anomalous differences were harnessed to determine the hand of the

\* Current address: CCLRC Daresbury Laboratory, Daresbury, Warrington, Cheshire WA4 4AD, UK.

† Current address: Institute of General and Ecological Chemistry, Technical University of Lodz, Zeromskiego 116, Poland.

**Table 1**

Theoretical values of  $\langle|\Delta F_{\text{anom}}|\rangle/\langle F \rangle$  (%) for lysozyme and apocrustacyanin A1 at wavelengths of 0.71, 1.54 and 2.0 Å.

$N_A$  = number of anomalous scatterers;  $N_L$  = number of protein (non-H) atoms;  $f_{\text{OL}}$  = average scattering factor of protein atoms (excluding H atoms and sulfurs), at zero scattering angle. Effects of thermal vibrations and atomic scattering-factor variations *versus*  $\sin\theta/\lambda$  are neglected. For lysozyme,  $N_A$  is equal to 10 for resolution better than 3 Å, *i.e.* considering the sulfurs of four disulfide bridges and two methionines resolved separately, while  $N_A$  is equal to 6 for resolution worse than 3 Å, considering the sulfurs of four disulfide bridges not resolved separately. The same calculation approach was used for apocrustacyanin A1.

Wavelength (Å)	Lysozyme						Apocrustacyanin A1					
	0.71		1.54		2.00		0.71		1.54		2.00	
Resolution (Å)	<3	>3	<3	>3	<3	>3	<3	>3	<3	>3	<3	>3
$f_A''$ (e <sup>-</sup> )	0.12	0.24	0.56	1.12	0.90	1.80	0.12	0.24	0.56	1.12	0.90	1.80
$N_A$	10	6	10	6	10	6	12	6	12	6	12	6
$N_L$	1100						3030					
$f_{\text{OL}}$	6.7 e <sup>-</sup>						6.7 e <sup>-</sup>					
$\langle \Delta F_{\text{anom}} \rangle/\langle F \rangle$ (%)	0.2	0.3	1.14	1.77	1.84	2.84	0.1	0.2	0.75	1.06	1.21	1.71

molecule (Cianci *et al.*, 2001). In 1999, Dauter and co-workers reported successful results of finding the sulfur atoms from their anomalous signals and of their use for phasing of hen egg-white lysozyme (HEWL) using synchrotron 1.541 Å wavelength cryodata. The small but significant anomalous dispersion signal from 17 anomalous scatterers, including sulfur and chlorine atoms, was enough to estimate the protein phases of lysozyme, a protein of 129 amino acid residues, by the single-wavelength anomalous dispersion (SAD) approach (Wang, 1985; for more recent reviews see Dauter *et al.*, 2002; Dodson, 2003; Ramagopal *et al.*, 2003).

We wished to enhance the sulfur signal using softer X-rays because of the extra challenge presented by a bigger protein, namely apocrustacyanin A1.

The feasibility of using softer X-rays had already been determined with experimental work using a lysozyme crystal up to 2.6 Å wavelength (Helliwell, 2004), in turn based on the development and utilization of 2, 1.89, 1.74 and 1.488 Å wavelengths with a variety of protein crystals and anomalous dispersion (Helliwell, 1984).

The use of softer wavelengths at synchrotrons, to enhance the anomalous signal from sulfur, arose again with the crystal structure solution of obelin (Liu *et al.*, 2000) and shortly after with studies, parallel to that reported here, on the crystal structure of apocrustacyanin C1 (Gordon *et al.*, 2001) and other protein structures (Micossi *et al.*, 2002; Brown *et al.*, 2002; Doan & Dockland, 2003).

In this paper, we report on exploring the feasibility of a high-redundancy approach (Harrop *et al.*, 1996) with HEWL and apocrustacyanin A1 at a variety of wavelengths. At room temperature, HEWL high-redundancy data sets were collected at different wavelengths, namely Cu  $K\alpha$  and Mo  $K\alpha$

using an in-house rotating anode, as well at 2 and 1.488 Å using SRS station 7.2. Dauter *et al.* (1999) had already performed a 1.5 Å wavelength lysozyme cryo experiment. Room-temperature data collection remains of interest for a variety of reasons, including avoiding cryo-induced structural changes (see *e.g.* Deacon *et al.*, 1997) and because some protein crystals will not freeze.

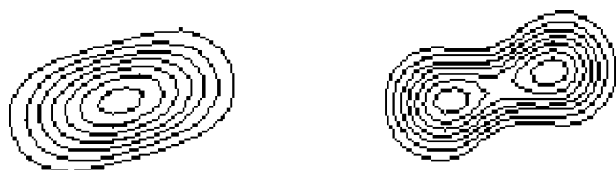
**2. Background theory**

The use of a 2.0 Å wavelength, termed ‘softer X-rays’ (Chayen *et al.*, 2000), produces an increase of 60% in the anomalous signal coming from the sulfur atoms compared with the use of Cu  $K\alpha$  radiation, and 80% relative to Mo  $K\alpha$ . At any wavelength, the signals from the two sulfur atoms in each disulfide bridge can reinforce each other at low and medium resolution (>3 Å). Moreover, anomalous difference Patterson maps calculated at 3 and 2 Å resolution would show a distinct ‘signature’ of splitting of a disulfide Harker peak from one (slightly elongated) into two distinct peaks (Fig. 1). The direct-methods approach to locate anomalous scatterers could, at high resolution, filter *E*-map phase sets, not only with respect to the figure of merit used in direct methods, but also by requiring two peaks separated by 2 Å (as is the case for S—S in a disulfide bridge).

The expected anomalous difference of structure-factor amplitudes *F* can be calculated with the following formula (Olczak *et al.*, 2003):

$$\langle|\Delta F_{\text{anom}}|\rangle/\langle F_{\text{anom}} \rangle = \frac{8}{\pi^{3/2}} \frac{f_A''}{f_{\text{OL}}} \left(\frac{N_A}{N_L}\right)^{1/2},$$

where  $N_L$  is the number of protein atoms (excluding H atoms),  $N_A$  is the number of anomalous scatterers,  $f_{\text{OL}}$  has a value of 6.7 electrons (weighted number of electrons for each protein atom) and  $f_A''$  is the imaginary anomalous scattering-factor component for sulfur at the wavelength used. In Table 1, the calculated values of  $\langle|\Delta F_{\text{anom}}|\rangle/\langle F \rangle$  (%) for lysozyme and apocrustacyanin A1 at wavelengths of 0.71, 1.54 and 2.0 Å are given.



**Figure 1**  
Example of splitting of a disulfide Harker peak from 3 Å (left) to 2 Å resolution (right). This is based on calculated ideal data for sulfur atoms in a disulfide bridge with *B* factor = 0.

**Table 2**

HEWL diffraction data sets (293 K).

The numbers in parentheses are for the appropriate highest resolution shell.

	Mo $K\alpha$ 0.710 Å	Cu $K\alpha$ 1.541 Å		Synchrotron 1.488 Å	Synchrotron 2.000 Å	
Source		Rigaku RU-200			Station 7.2 SRS	
Detector		R-Axis IIC area detector			MAR345 IP	
Crystal	1	1	2	1		1
Distance (mm)	80.0	150.0	80.0	140.0		83.3
Oscillation range (°) per image	1.0; 0.75	1.5	1.0	1.0		1.0
Number of images	58; 416	200	200	500	500	486
Exposure time (min)	20	20	15	1		5
Resolution (Å)	99.0–1.69		35.0–1.79	39.5–2.0		39.5–2.20
Number of unique reflections	13751	11714	8661	6390		6352
Data multiplicity	26		33.4	31.3	35.5	33.3
Highest resolution shell (Å)	1.73–1.69		1.86–1.79	2.00–2.07		2.28–2.20
Overall completeness (%)	99.0 (92.1)		99.1 (91.9)	100.0 (100.0)	98.2 (81.9)	97.5 (80.4)
$R_{\text{merge}}$ (%)	8.1 (41)		11.3 (38)	10.9 (29.7)	8.4 (14.5)	8.3 (17.6)
Overall $I/\sigma(I)$	49.35 (6.6)		37.3 (10.8)	52.8 (17.9)	92.7 (58.4)	63.1 (28.1)
Reflections with $I/\sigma(I) > 3$ (%)	88.2 (65.8)		93.5 (84.9)	94.2 (85.9)	98.8 (97.1)	90.5 (88.3)
Space group		$P4_32_12$			$P4_32_12$	
$a, b, c$ (Å)	78.82, 78.82, 38.18		79.13, 79.13, 37.96	79.31, 79.31, 38.09		78.77, 78.77, 38.14
Crystal size (mm)	$0.7 \times 0.6 \times 0.6$	$0.6 \times 0.7 \times 0.5$	$0.5 \times 0.7 \times 0.5$	$0.6 \times 0.6 \times 0.6$		$0.7 \times 0.7 \times 0.6$

**Table 3**

Correlation coefficient values for lysozyme experimental anomalous difference Patterson maps versus the map based on calculated positions from the successful *Shake 'n' Bake* run and the deposited coordinates of Dauter *et al.* (1999) (PDB code 1LZ8).

Map	Correlation coefficient
Mo $K\alpha$	0.244
Cu $K\alpha$	0.477
Synchrotron $\lambda = 1.488$ Å	0.60
Synchrotron $\lambda = 2.0$ Å	0.661
<i>Shake 'n' Bake</i> solutions	0.807
Calculated positions	1.000

### 3. Methods

#### 3.1. Crystals

The HEWL crystals used for these experiments had been grown by standard methods (see *e.g.* Helliwell, 1992). The images were processed using the *HKL* package (Otwinowski & Minor, 1997). The intensities were converted to amplitudes using the program *TRUNCATE* (Collaborative Computational Project, Number 4, 1994). Protocols and statistics are reported in Table 2.

The apocrustacyanin A1 crystals are those reported by Chayen *et al.* (2000) and Cianci *et al.* (2001).

#### 3.2. In-house HEWL data collection

**3.2.1. Molybdenum anode source.** A large crystal was mounted in a quartz capillary at room temperature. X-ray data were collected using Mo  $K\alpha$  radiation ( $\lambda = 0.710$  Å) from the in-house Rigaku RU-200 X-ray generator and R-AXIS IIC area detector. For the first batch (58 images), an oscillation range of  $1^\circ$  was used. For the second batch (416 images) the oscillation range was  $0.75^\circ$ .

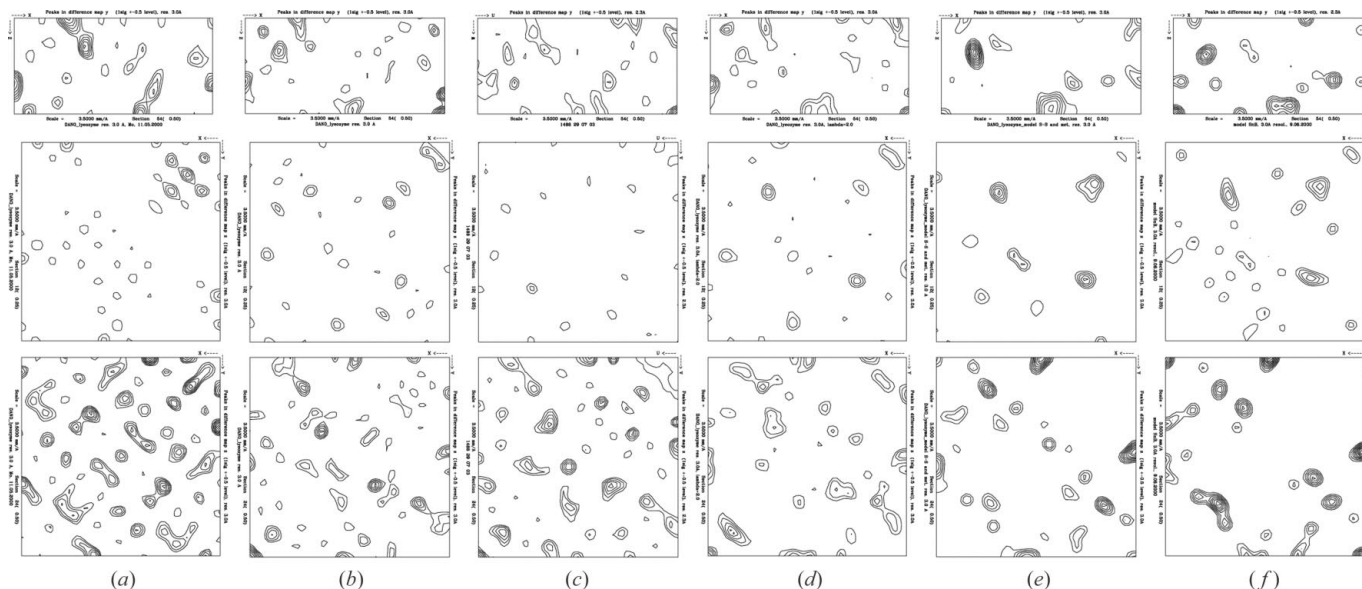
The location of the sulfur anomalous scatterers using *Shake 'n' Bake* (Miller *et al.*, 1994; Weeks & Miller, 1999) was unsuccessful and the anomalous difference Patterson map proved to be of poor quality (Fig. 2a). To convert  $\Delta F_{\text{anom}}$  to  $E_{\text{anom}}$  values, as required by *Shake 'n' Bake*, its subroutine *DREAR* (Blessing & Smith, 1999; Howell *et al.*, 2000) was run.

The correlation coefficient between this anomalous difference Patterson map and the 'ideal' Patterson map calculated from the ten (eight Cys and two Met residues) deposited sulfur-atom positions (PDB code 1LZ8; Dauter *et al.*, 1999) was 0.244 (Table 3); the correlation coefficient was calculated with *OVERLAPMAP* (Collaborative Computational Project, Number 4, 1994).

**3.2.2. Copper anode source.** Two crystals were mounted in glass capillaries at room temperature and X-ray data were collected using Cu  $K\alpha$  radiation ( $\lambda = 1.541$  Å) from a Rigaku RU-200 X-ray generator and an R-AXIS IIC area detector. For the first crystal, 200 images were collected with an oscillation range of  $1.5^\circ$ , 150 mm crystal to detector distance and an exposure time of 20 min. The second crystal data set was 200 images recorded using an oscillation range of  $1.0^\circ$ , 80 mm crystal to detector distance and an exposure time of 15 min.

The location of the anomalous scatterers was unsuccessful with *Shake 'n' Bake* (Miller *et al.*, 1994; Weeks & Miller, 1999), whilst the anomalous difference Patterson appeared to be of good quality (Fig. 2b). The correlation coefficient between this anomalous difference Patterson map and the Patterson map calculated from the ten (eight Cys and two Met residues) deposited sulfur-atom positions (PDB code 1LZ8; Dauter *et al.*, 1999) was 0.477 (Table 3); the correlation coefficient was calculated with *OVERLAPMAP* (Collaborative Computational Project, Number 4, 1994).

**3.2.3. Assessing for radiation damage for in-house Mo  $K\alpha$  and Cu  $K\alpha$  HEWL data collection.** The long exposure times and high number of images collected in both cases naturally warranted the consideration that radiation damage could have limited the quality of data collection. Analysis of the  $B$  factor



**Figure 2** Anomalous difference Patterson Harker sections for lysozyme at room temperature. Columns: (a) Mo  $K\alpha$  wavelength; (b) Cu  $K\alpha$  wavelength; (c) SRS with 1.488 Å wavelength; (d) SRS with 2 Å wavelength, 486 images; (e) calculated Patterson Harker sections for coordinates output by *Shake 'n' Bake* (2 Å wavelength, 486-image run); (f) calculated Patterson Harker sections for the deposited coordinates of the tetragonal lysozyme (PDB code 1LZ8; Dauter *et al.*, 1999). First row: section  $v = 0.5$ ; second row:  $w = 0.25$ ; third row:  $w = 0.5$ .

per diffraction image did not show any significant sign of radiation damage. In particular, in the case of data collection using the Mo  $K\alpha$  wavelength, the maximum value of the  $B$  factor per image was 1, with an average value of 0.5. For the two data collections using the Cu  $K\alpha$  wavelength for both data sets, the maximum value of the  $B$  factor per image was 2, with an average value of around 1. At 2 Å resolution, the impact of that correction at the resolution edge, and therefore of radiation damage, is therefore very small.

### 3.3. HEWL data collection at SRS Daresbury at 1.488 Å

The experiments with lysozyme reported in the previous sections suggested the importance of the use of a synchrotron as the radiation source for this kind of experiment to improve the data statistics. Therefore, the next test experiment was carried out at the SRS in Daresbury (UK) on station 7.2 (Helliwell *et al.*, 1982) with a Ge(111) single-crystal monochromator set to reflect X-rays at  $\lambda = 1.488$  Å. A MAR 345 image-plate detector system was used to collect 500 images of 1° with exposure times of 60 s per frame, in dose mode.

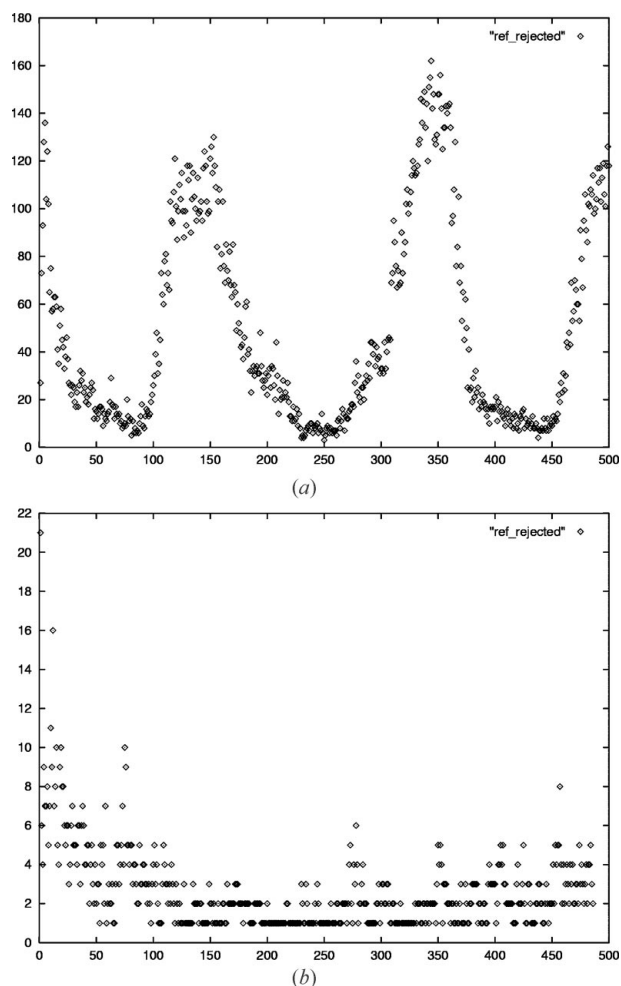
The location of the anomalous scatterers was unsuccessful with *Shake 'n' Bake* (Miller *et al.*, 1994; Weeks & Miller, 1999). The correlation coefficient between this anomalous difference Patterson map (Fig. 2c) and the Patterson map calculated from the ten (eight Cys and two Met residues) deposited sulfur-atom positions (PDB code 1LZ8; Dauter *et al.*, 1999) was improved at 0.60 (Table 3); the correlation coefficient was calculated with *OVERLAPMAP* (Collaborative Computational Project, Number 4, 1994).

### 3.4. HEWL data collection at SRS Daresbury at 2.0 Å

The previous experiment, where the Patterson map correlation coefficient was better (1.488 Å) with synchrotron radiation than with Cu  $K\alpha$ , confirmed the need for increasing the anomalous signal coming from the sulfurs, using a softer wavelength of  $\lambda = 2.0$  Å, as highlighted by preliminary experiments with apocrustacyanin A1 (Chayen *et al.*, 2000).

Therefore, another test was carried out at station 7.2 SRS in Daresbury (UK) using  $\lambda = 2.0$  Å. A MAR 345 image-plate detector system was used to collect 943 images of 1° with exposure times of 5 min per frame, in dose mode. A large crystal, 6 years old, mounted in a quartz capillary at room temperature, had been grown in the ESA's Advanced Protein Crystallization Facility on the NASA Shuttle during the LMS mission (Boggon, 1998). In spite of its age, the crystal's faces were of high optical quality. The first 500 images were scaled all together as a preliminary check for searching for the right parameters to be used with *SCALEPACK* and *Shake 'n' Bake*. The statistics of the data set are reported in Table 2.

The plot of rejected reflections *versus* image number (Fig. 3) showed a periodic behaviour with maxima for the number of rejected reflections every 180°. This phenomenon behaved like an absorption and/or radiation damage effect. Moreover, the search for the anomalous scatterers using *Shake 'n' Bake* was inconclusive. Since the number of rejected reflections for each image was very well correlated with the values of  $\chi^2$  and  $R_{\text{merge}}$ , only images with  $\chi^2$  and  $R_{\text{merge}}$  below a certain threshold were kept, while 'topping up' from the additional 443 unused images. The process was iterated several times. Each time, the chosen images were rescaled and the anomalous scatterer positions sought with *Shake 'n' Bake*. Finally, when 486 'good images' out of 943 images were selected,



**Figure 3**

Rejected reflections for the lysozyme data set at  $\lambda = 2.0 \text{ \AA}$ . (a) The first 500 images. There is a distinct signature of periodic behaviour with maxima for the number of rejected reflections every  $180^\circ$ . (b) The selected 486 images. In (a) the average number of rejected reflections is around 70; for (b) it is around 5. Ordinate: number of rejected reflections; abscissa: image number.

*Shake 'n' Bake* output was a bimodal distribution confirming the successful location of sulfur-atom positions from anomalous differences.<sup>3</sup> The disulfide bridges in lysozyme involve residues Cys6–Cys127, Cys30–Cys115, Cys64–Cys80 and Cys76–Cys94. *Shake 'n' Bake* correctly located the four disulfide bridges as four unique positions and also located the sulfur atoms of Met12 and Met105. The experimental anomalous difference Patterson map (Fig. 2d) was in good agreement with the calculated Patterson map for *SnB* solutions (Fig. 2e) and for the deposited coordinates (Fig. 2f) for the tetragonal lysozyme (PDB code 1LZ8; Dauter *et al.*, 1999). The positions for the peaks were correctly conserved throughout the Harker sections; the coordinates were matched *via* manual

<sup>3</sup> A supplementary table of these *SnB*-derived coordinates and the coordinates of the same set of atoms deposited by Dauter *et al.* (1999) (PDB code 1LZ8) is available from the IUCr electronic archives (Reference: EA5014). Services for accessing these data are described at the back of the journal.

inspection using *PDBSET* and *GENSYM* programs (Collaborative Computational Project, Number 4, 1994). The correlation coefficient between these anomalous difference Patterson maps and the Patterson maps calculated from the ten (eight Cys and two Met residues) deposited sulfur-atom positions (PDB code 1LZ8; Dauter *et al.*, 1999) was 0.661 (Table 3). The correlation coefficient was calculated with *OVERLAPMAP* (Collaborative Computational Project, Number 4, 1994).

The statistics of this 486-image data set are reported in Table 2.

**3.4.1. Assessing for radiation damage for the synchrotron data collection.** Analysis of the *B* factors per image for the data collected at  $1.488 \text{ \AA}$  did show a signature of radiation damage. The data set shows an increase of the *B* factors for the diffraction images around  $180$  and  $360^\circ$ . *B* values up to 12 were manifest. This is a serious effect and strongly suggests that the beam traversing through the crystal at the start of the run steadily caused damage. As a result of this analysis, 151 images were removed and *Shake 'n' Bake* re-run. This still did not yield the sulfur sites. For the synchrotron  $2 \text{ \AA}$  wavelength data, we have already documented above that the number of rejected reflections for the merged diffraction images indicated problems every  $180^\circ$  and these were removed. That approach finally gave the sites, as described above.

## 4. Sulfur-location approach applied to lobster apocrustacyanin A1

### 4.1. Background

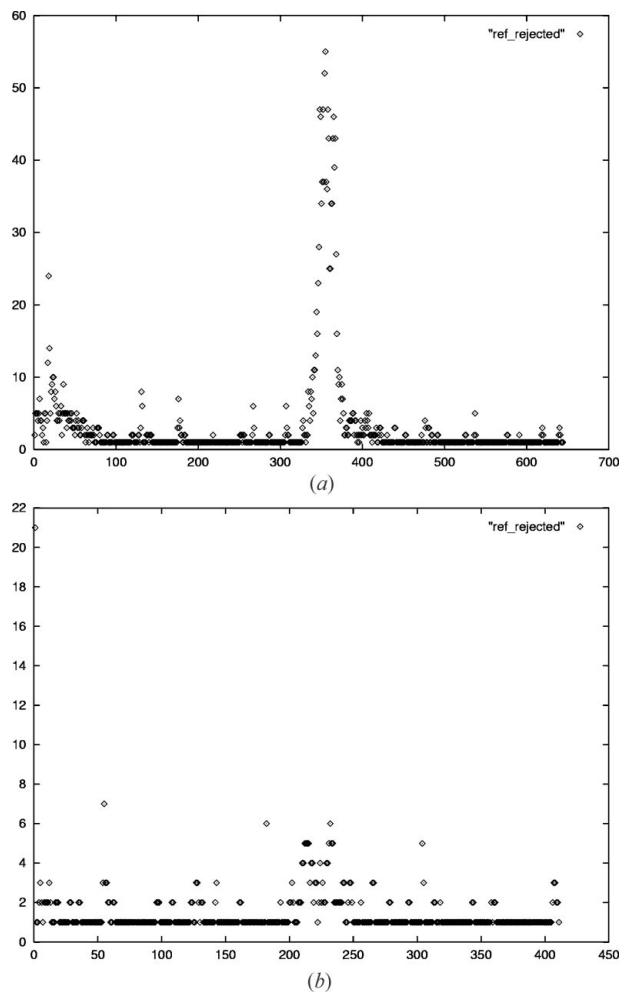
The blue colour of the lobster *Homarus gammarus* carapace is provided by a 16 protein subunit complex, with 16 bound astaxanthin molecules (for a recent review, see Chayen *et al.*, 2003). Crystal structures were needed to provide the structural basis of the spectral shifts of the carotenoid in the complex state. Crystals of  $\alpha$ -crustacyanin were first reported some 30 years ago, though without any recorded diffraction. Later, single subunits of this complex were targeted, but their crystal structure determination proved problematic through lack of a suitable molecular replacement model even from the same family of lipocalin proteins, lack of good heavy-atom derivatives, and lack of seleno-methionine variants.

The lack of success with ordinary techniques of protein crystal structure determination for apocrustacyanin A1 warranted consideration of new approaches (Chayen *et al.*, 2000). One possibility was the use of anomalous differences from the three putative disulfides per monomer (six per asymmetric unit, from 12 cysteines) to locate anomalous scatterers *via* anomalous difference Patterson maps (Brunger *et al.*, 1998; Terwilliger & Berendzen, 1999), by a direct methods approach (Mukherjee *et al.*, 1989; Weeks & Miller, 1999; Schneider & Sheldrick, 2002), or, in a more esoteric way, with molecular replacement (Antson *et al.*, 1995). For apocrustacyanin A1, early results (Chayen *et al.*, 2000) suggested other basic requirements for locating the weak anomalous scatterers such as sulfur atoms, namely the need for accurate

**Table 4**  
Apocrustacyanin A1 data collection statistics.

The numbers in parentheses are for the appropriate highest resolution shell.

	Room-temperature low-redundancy 1.488 Å	Room-temperature low-redundancy 2.0 Å	Cryo high-redundancy 2.0 Å
Published	Chayen <i>et al.</i> (2000)	Chayen <i>et al.</i> (2000)	Cianci <i>et al.</i> (2001)
Source	Station 7.2 SRS	Station 7.2 SRS	Station 9.5 SRS
Temperature (K)	293	293	100
Detector	MAR 345	MAR 345	MAR165 CCD
Distance (mm)	150	83.3	83.3
Oscillation range (°) per image	1.0	1.0	1.0
Number of images	90	90	644 411
Exposure time (s)	60	300	50
Resolution (Å)	25.0–2.0	25.0–2.3	64.5–2.7
Number of unique reflections	27426	16464	10420 10392
Data multiplicity	4	3.6	23.5 14.7
Highest resolution shell (Å)	2.11–2.0	2.42–2.30	2.80–2.70
Overall completeness (%)	99.8 (99.5)	99.7 (99.7)	99.5 (100.0) 99.1 (99.7)
$R_{\text{merge}}$ (%)	9.7 (33.5)	6.4 (12.2)	8.3 (20.2) 5.8 (14.4)
Overall $I/\sigma(I)$	4.8 (1.7)	7.9 (5.1)	55.6 (26.7) 48.5 (22.3)
Reflections with $I/\sigma(I) > 3$ (%)	–	–	96.0 (88.2) 95.5 (86.5)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
$a, b, c$ (Å)	41.8, 81.0, 110.6	41.9, 80.7, 110.8	41.14, 79.76, 109.73



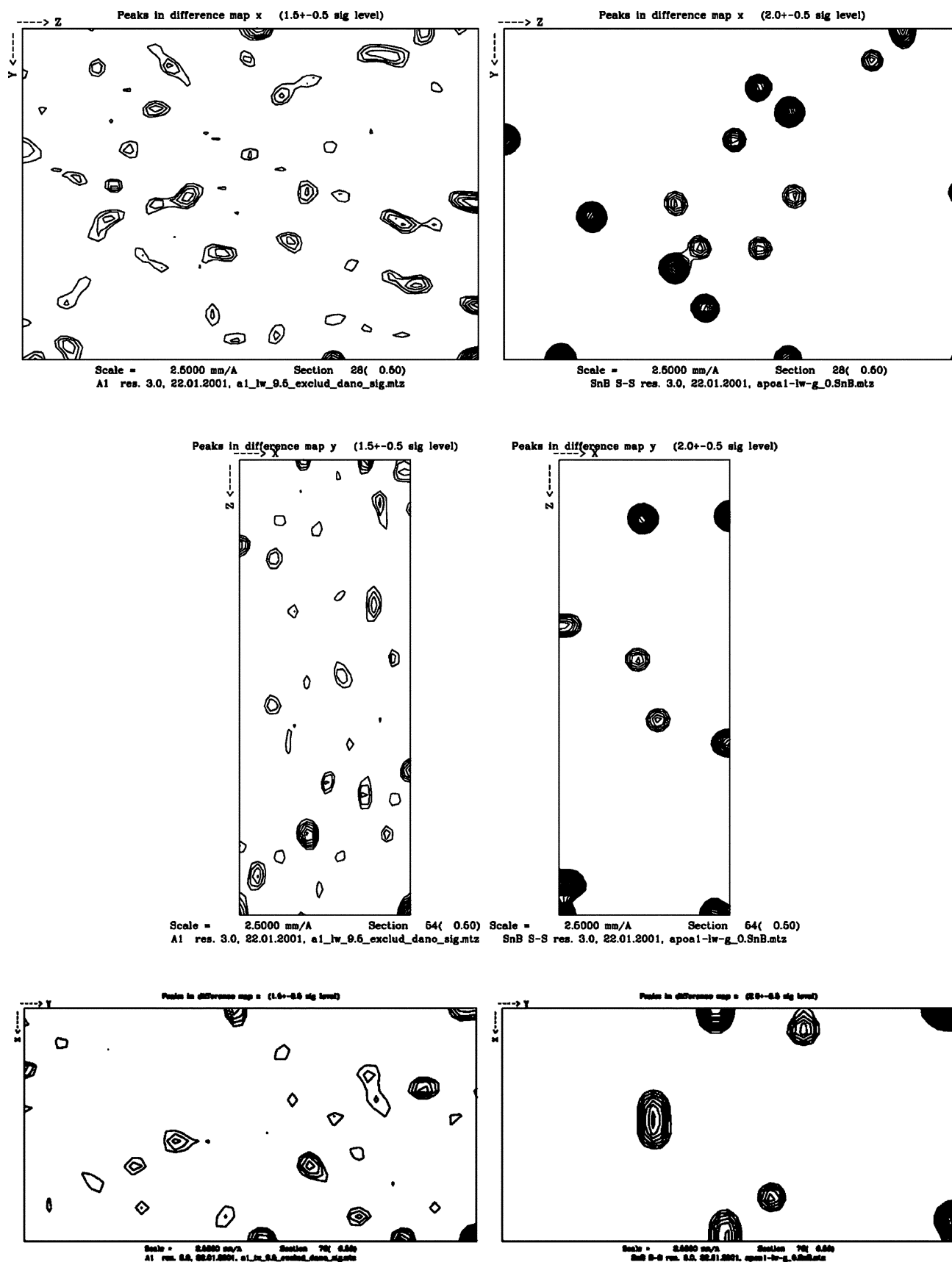
**Figure 4**  
Rejected reflections for (a) the 644-image highly redundant data set of apocrustacyanin A1 at  $\lambda = 2.0$  Å. There is an evident distinct signature of periodic behaviour with a maximum for the number of rejected reflections every  $360^\circ$ . (b) The selected 411 images of the highly redundant data set of apocrustacyanin A1 at  $\lambda = 2.0$  Å. Ordinate: number of rejected reflections; abscissa: image number.

data measurement and use of a softer wavelength. However, crystals of apocrustacyanin protein, available in limited amount, proved to be notoriously fragile and not adequate for any soaking (Boggon, 1999; Cianci, 2002), thus making the preparation of suitable cryo conditions difficult.

#### 4.2. Crystals and methods

The apocrustacyanin A1 crystals were those reported by Chayen *et al.* (2000) and Cianci *et al.* (2001). Protocols and statistics of the room-temperature data, referred to as ‘room-temperature low-redundancy 1.488 Å’ (Chayen *et al.*, 2000) and ‘room-temperature low-redundancy 2.0 Å’ (Chayen *et al.*, 2000), and the cryo data, ‘cryo high-redundancy 2.0 Å’ (Cianci *et al.*, 2001), analysed here are reported in Table 4. To attempt the location of the sulfur-atom positions from the anomalous signal harnessed with the 644-image ‘cryo high-redundancy 2.0 Å’ data set, *Shake ‘n’ Bake* (Miller *et al.*, 1994; Weeks & Miller, 1999) was used. The subroutine *DREAR* (Blessing & Smith, 1999; Howell *et al.*, 2000) was run, prior to *Shake ‘n’ Bake*, to convert the  $\Delta F_{\text{anom}}$  to  $E_{\text{anom}}$  values. However, after several runs of *Shake ‘n’ Bake* it was clear that it was not possible to locate the sulfur-atom positions from the weak sulfur-atom anomalous signal of this data set.

The plot of rejected reflections *versus* number of images (Fig. 4) showed, similar to the lysozyme test (Fig. 3), a maximum for the number of rejected reflections at  $0^\circ$  and at  $360^\circ$ . For *Shake ‘n’ Bake*, the same protocol was then used as for the lysozyme case. The data processing was repeated progressively, excluding those images whose  $R_{\text{merge}}$  and  $\chi^2$  values were highest. Finally, 411 images were selected for which *Shake ‘n’ Bake* yielded the positions of the anomalous scatterers, as evidenced by a bimodal distribution. Comparison of these coordinates with those based on the SIROAS Xenon phased maps (Cianci *et al.*, 2001) reflected their correctness. The positions of the peaks are conserved throughout the various sections (Fig. 5).



**Figure 5**  
 Patterson Harker sections for apocrustacyanin A1 high-redundancy data (411-image subset). First column: anomalous difference maps; second column: calculated from *Shake 'n' Bake* solution positions. First row: section  $u = 0.5$ ; second row:  $v = 0.5$ ; third row:  $w = 0.5$ . The correlation coefficient between the two maps is 0.54 (*OVERLAPMAP*; Collaborative Computational Project, Number 4, 1994).

## 5. Discussion

### 5.1. HEWL results

Our investigations into the problems of location of weak anomalous scatterers using a variety of sources and wavelengths, reported here, show the advantages of using intense synchrotron radiation with 2 Å wavelength for finding sulfur-atom positions. In fact, differently from the previous case of Dauter *et al.* (1999), the sulfurs in HEWL were successfully located from a single lysozyme crystal mounted on a quartz capillary at room temperature rather than at cryo temperature. Looking at the HEWL case, data quality did not particularly suffer from the use of a longer wavelength. These data sets are each of very high redundancy, which affects the  $R_{\text{merge}}$ . It should be borne in mind that the HEWL data were room-temperature and not cryo data. The percentage of data above  $3\sigma$  is a different way to show the data quality; all data sets here have about 95%  $> 3\sigma(I)$  overall and  $>85\%$  in the outer shell, except the Mo  $K\alpha$  set, with percentages of 88% and 66%, respectively. These data sets are significantly strong at the edge of the pattern.

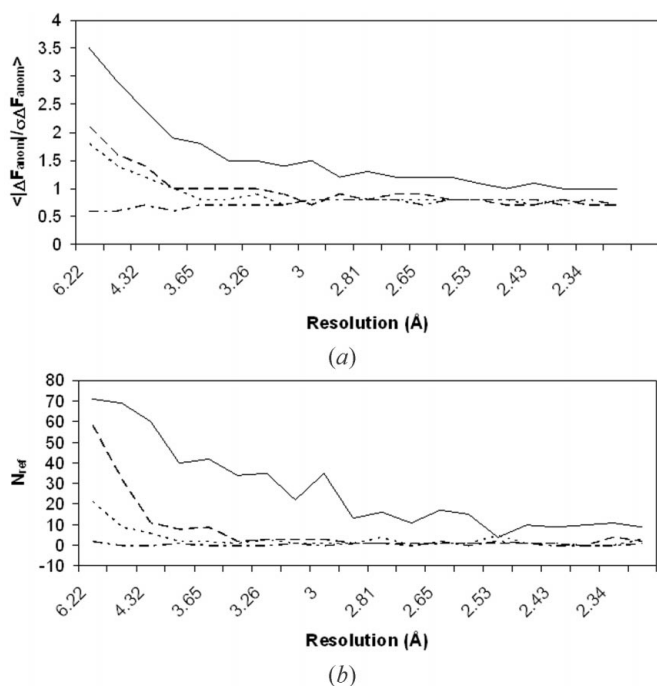
The processing of images and monitoring of rejected reflections per image was carefully done in order to use the weak anomalous signal of the sulfur atoms in *Shake 'n' Bake*. Different protocols (*i.e.* using other processing packages) may give different (better or worse) results, as explained by Weiss *et al.* (2001) in a study parallel to this one. In our case, discarding of images was achieved *via*  $\chi^2$  and  $R_{\text{merge}}$  monitoring. In particular, given the quality of the anomalous

difference Patterson Harker sections of Fig. 2(b) and the correlation coefficient of 0.477 (Table 3), it is quite possible that our room-temperature HEWL Cu  $K\alpha$  data set may be close to yielding a *Shake 'n' Bake* solution.

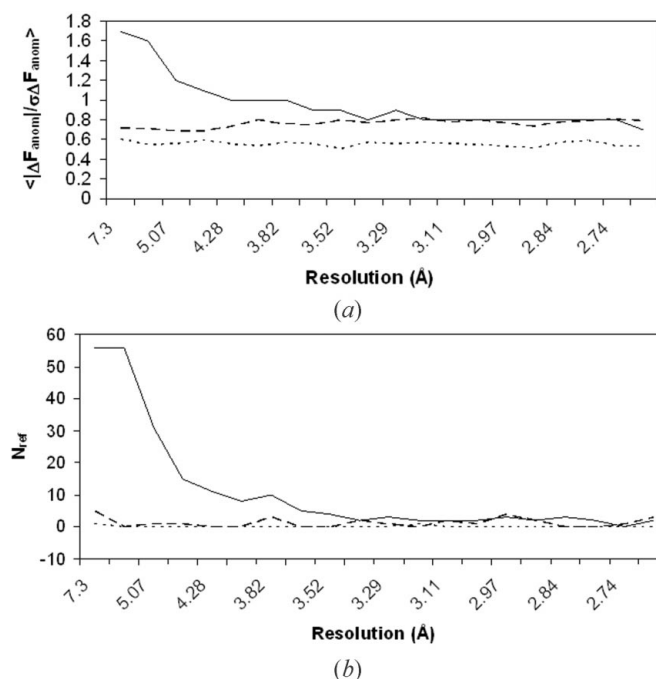
*Shake 'n' Bake* (Miller *et al.*, 1994; Weeks & Miller, 1999) could locate the positions of the anomalous scatterers (Smith *et al.*, 1998) only when a sufficient number of  $N_{\text{ref}}$  [number of reflections per shell with  $\Delta F_{\text{anom}}$  greater than three times  $\sigma(\Delta F_{\text{anom}})$ ] were present (see Figs. 6 and 7). A low number of  $N_{\text{ref}}$  means that in *Shake 'n' Bake*, and probably in other programs exploiting direct methods, the ratios atoms:reflections, reflections:invariants and  $|E|/\sigma(E)$  have to be relaxed from their default values, thereby reducing the likelihood of locating anomalous scatterers (Wang & Ealick, 2003).

For HEWL, because of the large crystal size, perhaps the data suffered from absorption variations. The crystal was chosen so as to be big enough to survive the likely radiation damage that a high-redundancy data collection at room temperature would produce. The crystal was nearly isodimensional ( $0.7 \times 0.7 \times 0.6$  mm) with  $T_{\text{max}}/T_{\text{min}} = 0.91$  at 1.488 Å wavelength. Absorption errors can therefore probably be ruled out to explain why sulfur could not be found at 1.488 Å wavelength for HEWL.

For the HEWL case, the experimental values of  $(\langle \Delta F_{\text{anom}} / |F| \rangle)$  of each data set (all have high redundancy) increase while moving to a longer wavelength, *i.e.* Mo  $K\alpha$ , synchrotron 1.488 Å, Cu  $K\alpha$ , to synchrotron 2 Å wavelength. It should be expected that the source has no effect on this parameter, but only for synchrotron 2.0 Å the calculated values of  $(\langle \Delta F_{\text{anom}} /$



**Figure 6** Statistics for HEWL [Mo  $K\alpha$  (dot-dashed line), Cu  $K\alpha$  (dotted line), 1.488 Å synchrotron (dashed line) and 2.0 Å synchrotron (full line), 486 images]:  $\langle |\Delta F_{\text{anom}}| / \sigma(\Delta F_{\text{anom}}) \rangle$  versus resolution (a);  $N_{\text{ref}}$  versus resolution (b).



**Figure 7** Statistics for apocrustacyanin A1 [1.488 Å synchrotron (dotted line), 2.0 Å synchrotron low redundancy (dashed line) and 2.0 Å high redundancy (full line), 411 images]:  $\langle |\Delta F_{\text{anom}}| / \sigma(\Delta F_{\text{anom}}) \rangle$  versus resolution (a);  $N_{\text{ref}}$  versus resolution (b).

$|F|$ ) versus resolution are around 2–3.5% in the resolution range between 3.0 and 5.0 Å as expected from calculated values (Table 1).

Fig. 6 compares, for each lysozyme data set,  $\langle \Delta F_{\text{anom}} / \sigma(\Delta F_{\text{anom}}) \rangle$  and  $N_{\text{ref}}$  versus resolution, where  $N_{\text{ref}}$  is the number of reflections per shell with  $\Delta F_{\text{anom}}$  greater than 3 times  $\sigma(\Delta F_{\text{anom}})$ .  $\langle \Delta F_{\text{anom}} / \sigma(\Delta F_{\text{anom}}) \rangle$  and  $N_{\text{ref}}$  appear to be dependent on the type of source and wavelength used, with a definite increment in the number of reflections with significant anomalous signal in favour of synchrotron radiation. Mo  $K\alpha$  has no resolution range with  $\langle \Delta F_{\text{anom}} / \sigma(\Delta F_{\text{anom}}) \rangle > 1$ ; Cu  $K\alpha$  had  $\Delta \geq 1\sigma$  up to 3.93 Å resolution; synchrotron 1.488 Å had  $\Delta \geq 1\sigma$  nearly up to 3.0 Å resolution, and the synchrotron 2 Å wavelength data set had  $\Delta \geq 1\sigma$  in all the resolution ranges. Table 3 also shows an improvement of the correlation coefficient of the observed with calculated Patterson maps, while shifting to a longer wavelength, *i.e.* Mo  $K\alpha$  versus Cu  $K\alpha$ , synchrotron 1.488 Å versus synchrotron 2 Å wavelength. The improvement of  $\langle I \rangle / \langle \sigma(I) \rangle$  in the statistics of the data sets when collected at SRS station 7.2 at 2 Å is also evident.

## 5.2. Apocrustacyanin A1 results

Since no solution could be found applying *Shake 'n' Bake* to the room-temperature data, 'room-temperature low-redundancy 1.488 Å' and 'room-temperature low-redundancy 2.0 Å' (Chayen *et al.*, 2000), and since the image-rejection approach could not be employed either, a high-redundancy 2 Å wavelength data set had to be measured, with cryo conditions harnessed to allow the crystal to survive possible radiation damage (Garman & Schneider, 1997). A high-quality subset of images was selected, for which *Shake 'n' Bake* produced a bimodal diagram. As Table 4 shows, the  $R_{\text{merge}}$  improved, as one would expect.

The values of  $\langle \Delta F_{\text{anom}} / |F| \rangle$  resolution for the high-redundancy 411-image data set are around 1.9–2.1% in the resolution range between 3.2 and 5.0 Å, as expected from calculated values for data collected at  $\lambda = 2.0$  Å (Table 1). Fig. 7 shows the benefits of a very high redundancy cryo approach (411 images used), *i.e.* the 2 Å wavelength high-redundancy curve versus the other curves.

## 5.3. Common results between the two case studies

If we now compare the two case studies, firstly, for the apocrustacyanin A1 data sets collected at 1.488 Å and at 2.0 Å wavelength at room temperature and with low redundancy,  $\langle \Delta F_{\text{anom}} / \sigma(\Delta F_{\text{anom}}) \rangle$  and  $N_{\text{ref}}$  values across all resolution shells do not improve as much as observed in the HEWL case (*i.e.* compare Figs. 6 and 7). Only for the apocrustacyanin A1 cryo data set collected at 2.0 Å wavelength with a high multiplicity does the ratio  $\langle \Delta F_{\text{anom}} / \sigma(\Delta F_{\text{anom}}) \rangle$  reach beyond a value of 1.0 in the lower resolution shells. The behaviour of the parameter  $N_{\text{ref}}$  for the apocrustacyanin A1 case, across the same range of resolution, is less good, even though *Shake 'n' Bake* obviously worked.

Secondly, in both the HEWL and the apocrustacyanin A1 case, the highly redundant data sets were pruned, *i.e.* images

systematically removed in order to achieve a bimodal *Shake 'n' Bake* distribution. The basis for the removal of any image was the same in each case, namely monitoring  $\chi^2$ ,  $R_{\text{merge}}$ , rejected reflections and the  $B$  factor per diffraction image. The removal of a significant number of images, in several iterations, is neither an optimal use of data collection beam time or the investigator's analysis time. However, rapid data collection has allowed this approach. A longer exposure for each image, but with fewer images for the available beam time, would have perhaps been the more automatic route, but with room-temperature data radiation damage is then the cause of anxiety.

## 6. Conclusion

Working at room temperature with lysozyme as a test revealed that SRS  $\lambda = 2$  Å data were needed over a full revolution for *Shake 'n' Bake* to succeed, but still with careful trimming of images out of the data set. Patterson map correlation coefficients with the ideal map improved from short to long wavelength, and from Cu  $K\alpha$  to 1.488 Å synchrotron radiation. Guided by these evaluations with HEWL, the more challenging case of apocrustacyanin A1 warranted use of synchrotron radiation, yet higher redundancy, use of a softer X-ray wavelength ( $\lambda = 2$  Å), and cryo conditions. Trimming out of images was also needed. *Shake 'n' Bake* was thereby successful. Our approach to the location of sulfur atoms and the experience detailed here add to a growing enthusiasm for the use of softer synchrotron X-rays in structure determination in protein crystallography.

We thank The Leverhulme Trust and the EU Crystallogenesis Project for salary support (AO) and the EPSRC for studentship support (MC). MC's salary support has continued, as a PDRA, under a BBSRC Award at Daresbury [PI's: J. R. Helliwell and S. S. Hasnain with nine other members of the North West Structural Genomics Consortium (NWSGC)]. Access to the SRS facilities was awarded by the Joint Biology Program of the Research Councils of the UK. This paper is a third article contribution from the NWSGC, Manchester node with Daresbury Laboratory.

## References

- Antson, A. A., Otrudge, J., Brzozowski, A. M., Dodson, E. J., Dodson, G. G., Wilson, K. S., Smith, T. M., Yang, M., Kurecki, T. & Gollnik, P. (1995). *Nature (London)*, **374**, 693–700.
- Blessing, R. H. & Smith, G. D. (1999). *J. Appl. Cryst.* **32**, 664–670.
- Boggon, T. J. (1998). PhD thesis, Department of Chemistry, University of Manchester.
- Brown, J., Esnouf, R. M., Jones, M. A., Linnell, J., Harlos, K., Hassan, A. B. & Jones, E. Y. (2002). *EMBO J.* **21**, 1054–1062.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst. D* **54**, 905–921.
- Chayen, N. E., Cianci, M., Grossmann, J. G., Habash, J., Helliwell, J. R., Nneji, G. A., Raftery, J., Rizkallah, P. J. & Zagalsky, P. F. (2003). *Acta Cryst. D* **59**, 2072–82.

- Chayen, N. E., Cianci, M., Olczak, A., Raftery, J., Rizkallah, P. J., Zagalsky, P. F. & Helliwell, J. R. (2000). *Acta Cryst.* **D56**, 1064–1066.
- Cianci, M. (2002). PhD thesis, Department of Chemistry, University of Manchester.
- Cianci, M., Rizkallah, P. J., Olczak, A., Raftery, J., Chayen, N. E., Zagalsky, P. F. & Helliwell, J. R. (2001). *Acta Cryst.* **D57**, 1219–1229.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Dauter, Z., Dauter, M. & Dodson, E. (2002). *Acta Cryst.* **D58**, 494–506.
- Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **28**, 83–92.
- Deacon, A., Gleichmann, T., Kalb (Gilboa), A. J., Price, H., Raftery, J., Bradbrook, G., Yariv, J. & Helliwell, J. R. (1997). *Faraday Trans.* **93**, 4305–4312.
- Debreczeni, J. E., Bunkoczi, G., Girmann, B. & Sheldrick, G. M. (2003). *Acta Cryst.* **D59**, 393–395.
- Debreczeni, J. E., Bunkoczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). *Acta Cryst.* **D59**, 688–696.
- Debreczeni, J. E., Girmann, B., Zeeck, A., Kratzner, R. & Sheldrick, G. M. (2003). *Acta Cryst.* **D59**, 2125–2132.
- Doan, D. N. & Dockland, T. (2003). *Structure*, **11**, 1445–1451.
- Dodson, E. (2003). *Acta Cryst.* **D59**, 1958–1965.
- Garman, E. F. & Schneider, T. R. (1997). *J. Appl. Cryst.* **30**, 211–237.
- Gordon, E. J., Leonard, G. A., McSweeney, S. & Zagalsky, P. F. (2001). *Acta Cryst.* **D57**, 1230–1237.
- Harrop, S. J., Helliwell, J. R., Wan T. C. M., Kalb (Gilboa), A. J., Tong, L. & Jariv, J. (1996). *Acta Cryst.* **D52**, 143–155.
- Helliwell, J. R. (1984). *Rep. Prog. Phys.* **47**, 1403–1497.
- Helliwell, J. R. (1992). *Macromolecular Crystallography with Synchrotron Radiation*. Cambridge University Press.
- Helliwell, J. R. (2004). *J. Synchrotron Rad.* **11**, 1–3.
- Helliwell, J. R., Greenhough, T. J., Carr, P. D., Rule, S. H., Moore, P. R., Thompson, A. W. & Worgan, J. S. (1982). *J. Phys. E*, **15**, 1563–1372.
- Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
- Howell, P. L., Blessing, R. H., Smith, G. D. & Weeks, C. M. (2000). *Acta Cryst.* **D56**, 604–617.
- Kwiatkowski, W., Noel, J. P. & Choe, S. (2000). *J. Appl. Cryst.* **33**, 876–881.
- Lemke, C. T., Smith, G. D. & Howell, P. L. (2002). *Acta Cryst.* **D58**, 2096–2101.
- Liu, Z. J., Vysotski, E. S., Chen, C. J., Rose, J. P., Lee, J. & Wang, B. C. (2000). *Protein Sci.* **9**, 2086–2093.
- Micossi, E., Hunter, W. N. & Leonard, G. A. (2002). *Acta Cryst.* **D58**, 21–28.
- Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
- Mukherjee, A., Helliwell, J. R. & Main, P. (1989). *Acta Cryst.* **A45**, 715–718.
- Olczak, A., Cianci, M., Hao, Q., Rizkallah, P. J., Raftery, J. & Helliwell, J. R. (2003). *Acta Cryst.* **A59**, 327–334.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* **D59**, 1020–1027.
- Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24.
- Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* **D58**, 1772–1779.
- Smith, G. D., Nagar, B., Rini, J. M., Hauptman, H. A. & Blessing, R. H. (1998). *Acta Cryst.* **D54**, 799–804.
- Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* **D55**, 501–505.
- Yang, C. & Pflugrath, J. W. (2001). *Acta Cryst.* **D57**, 1480–1490.
- Wang, B. C. (1985). *Methods Enzymol.* **115**, 90–112.
- Wang, J. & Ealick, S. E. (2003). *J. Appl. Cryst.* **36**, 1397–1401.
- Weeks, C. M. & Miller, R. (1999). *J. Appl. Cryst.* **32**, 120–124.
- Weiss, M. S., Sicker, T., Djinic Carugo, K. & Hilgenfeld, R. (2001). *Acta Cryst.* **D57**, 689–695.