Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The interdependence of wavelength, redundancy and dose in sulfur SAD experiments

Michele Cianci,[a,b]* John R. Helliwell[b,c] and Atsuo Suzuki[d]

[a]European Molecular Biology Laboratory (EMBL), Hamburg Outstation, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany, [b]CCLRC Daresbury Laboratory, Daresbury, Cheshire WA4 4AD, England, [c]School of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, England, and [d]Department of Biotechnology, Graduate School of Engineering, Nagoya University, Nagoya, Japan

Correspondence e-mail: michele.cianci@embl-hamburg.de

Received 29 May 2008
Accepted 22 September 2008

PDB References: 2w1m; 2w1l; 2w1x; 2w1y.

In the last decade, the popularity of sulfur SAD anomalous dispersion experiments has spread rapidly among synchrotron users as a quick and streamlined way of solving the phase problem in macromolecular crystallography. On beamline 10 at SRS (Daresbury Laboratory, UK), a versatile design has allowed test data sets to be collected at six wavelengths between 0.979 and 2.290 Å in order to evaluate the importance and the interdependence of experimental variables such as the Bijvoet ratio, wavelength, resolution limit, data redundancy and absorbed X-ray dose in the sample per data set. All the samples used in the experiments were high-quality hen egg-white lysozyme crystals. X-radiation damage was found to affect disulfide bridges after the crystals had been given a total dose of 0.20×10^7 Gy. However, with such a total dose, it was still possible in all cases to find a strategy to collect data sets to determine the sulfur substructure and produce good-quality phases by choosing an optimum combination of wavelength, exposure time and redundancy. A $\langle|\Delta_{\rm ano}|/\sigma(\Delta_{\rm ano})\rangle$ greater than 1.5 for all resolution shells was a necessary requirement for successful sulfur SAD substructure location. Provided this is achieved, it seems possible to find an optimum compromise between wavelength, redundancy and dose to provide phasing information. The choice of the wavelength should then follow the sample composition and the diffracting properties of the crystal. For strongly diffracting crystals, wavelengths equal or shorter than 1.540 Å can be selected to capture the available data (provided the Bijvoet ratio is reasonable), while a longer wavelength, to gain as high a Bijvoet ratio as possible, must be used for more weakly diffracting crystals. These results suggest that an approach to a sulfur SAD experiment based on a complete description of the crystal system and the instrument for data collection is useful.

1. Introduction

In the early 1980s, a paper by Hendrickson and Teeter described the structure solution of the 46-residue polypeptide crambin from the anomalous differences for the sulfurs present in its amino-acid sequence, where the sulfurs provided calculated heavy-atom phases to make a choice within the inherent phase ambiguity of SAD (Hendrickson & Teeter, 1981). Four years later, methods to resolve the phase ambiguity of the SAD approach involving solvent density modification were presented (Wang, 1985). In the late 1990s, Dauter and coworkers reported the successful phasing of cryocooled

hen egg-white lysozyme (HEWL; 129 residues) from the anomalous signal of S atoms using synchrotron radiation of wavelength 1.540 Å (Dauter *et al.*, 1999; for recent reviews on the SAD method, see, for example, Dauter *et al.*, 2002; Dodson, 2003; Ramagopal *et al.*, 2003). Soon afterwards, many novel structures were solved by sulfur SAD such as those of obelin (Liu *et al.*, 2000), apocrustacyanin C_1 (Gordon *et al.*, 2001) and other protein structures (Micossi *et al.*, 2002; Lemke *et al.*, 2002; Debreczeni, Bunkóczi, Girmann *et al.*, 2003; Debreczeni, Bunkóczi, Ma *et al.*, 2003; Brown *et al.*, 2002; Doan & Dockland, 2003; Olsen *et al.*, 2004; Weiss *et al.*, 2004; Agarwal *et al.*, 2006; Wagner *et al.*, 2006). In combination with a xenon derivative, a sulfur SAD data set was used to provide the correct enantiomer of the two possible heavy-atom model solutions for apocrustacyanin A_1 (Cianci *et al.*, 2001).

In recent years, with the growing popularity of sulfur SAD phasing, X-radiation damage became a widespread concern, resulting in an increasing number of studies describing the mechanics and the effects of photon interactions with protein crystals. Systematic analysis of synchrotron data sets of unfrozen or frozen samples showed characteristic 'fingerprints' of radiation damage [Helliwell, 1988 (based on an unpublished suggestion of G. A. Petsko); Ravelli & McSweeney, 2000; Burmeister, 2000; Borek *et al.*, 2007]. Radiation damage currently represents a potential issue when collecting multi-wavelength anomalous data sets (Ravelli & Garman, 2006), but also provides an opportunity to collect additional phasing information (see Ravelli *et al.*, 2003; Nanao *et al.*, 2005, and references therein). Despite radiation damage casting a shadow over each data collection, the popularity of sulfur single-wavelength anomalous dispersion (SAD) experiments has spread rapidly amongst synchrotron users as a relatively 'quick and simple' way of solving the phase problem in macromolecular crystallography. One reason for the popularity of sulfur SAD is that it relies on a natural property of proteins, with no need for further chemical modification. Sulfur is present in nearly all proteins and the same methods can also be applied to nucleic acids using phosphorus (Dauter & Adamiak, 2001). Another reason for the popularity of SAD lies in the intrinsic simplicity of the process: once the crystal has been mounted and the wavelength has been selected, users just have to collect sufficiently accurate data in order to be able to locate S-atom positions. Ideally while images are still being collected, users will start integrating and scaling the data to feed the results to software packages that in the most favourable case will output firstly the positions of the sulfur positions and then a traced chain. From a user point of view, the experiment appears to be simple, nearly trivial and possibly completely streamlined. A sulfur single anomalous dispersion data collection may nearly lose its experimental appeal and be seen as a mere routine measurement.

However, before such measurements are performed, many decisions need to be taken. Such a decision-making process will influence the final quality of the data to be collected and will bring back the measurement to its experimental nature. The first step towards a successful SAD experiment should be evaluation of the Bijvoet ratio $\langle|\Delta_{\rm ano}|\rangle/\langle F\rangle$ for the sample as a function of the wavelength by using the formulas reported by Olczak *et al.* (2003) and available online at http://alfa.p.lodz.pl/assc/.

Conditions for cryocrystallography can be optimized beforehand (see Garman & Schneider, 1997; Garman & Owen, 2006 and references within) by testing cryoprotectants (Garman & Mitchell, 1996) and radioprotectants (Kauffmann *et al.*, 2006; Southworth-Davies & Garman, 2007).

Attention should also be paid to crystal-mounting instrumentation, transfer, storage and transport. The original cryomounting technique (Teng, 1990) has recently been improved, with a new method to remove the cryo-buffer and the cryoloop around the protein crystal before data collection in order to reduce X-ray absorption (Kitago *et al.*, 2005) and to produce better results in phasing statistics for the sulfur SAD method with Cr $K\alpha$ radiation.

The correct or 'best' wavelength for the experiment should be considered carefully. For protein structure determination, Chayen *et al.* (2000) and Weiss *et al.* (2001) suggested in two separate initiatives that for the optimal use of sulfur SAD the use of softer X-rays in the wavelength range λ = 1.5–3.0 Å should be harnessed in order to increase the anomalous signal of many elements including sulfur.

Finally, the redundancy of the data required should be decided. The benefits of a highly redundant diffraction data collection for locating anomalous scatterers have been shown by Harrop *et al.* (1996) with Cd and Ca, and by Dauter *et al.* (1999) for locating sulfur and chloride anomalous scatterers. Weiss *et al.* (2001) also championed highly redundant data collection as a pillar for structure determination *via* SAD.

Dauter & Adamiak (2001) showed that the influence of data redundancy in the case of a small DNA oligomer containing phosphates was not only on successful anomalous scatterer location but also at the phasing stage. Higher redundancy data gave more solutions per phase set and overall better figure-of-merit (FOM) values.

In an experiment at room temperature with a crystal mounted in a capillary, a highly redundant data set, the use of softer X-rays and a customized data-scaling approach were the key to locating the sulfur disulfide bridges in HEWL (Cianci *et al.*, 2004).

For longer wavelength data collection, different approaches have been considered such as precise alignment of the crystal (Einspahr *et al.*, 1985; Nieh & Helliwell, 1995) and data processing (Weiss *et al.*, 2004). Strategies for locating anomalous scatterers and phasing protocols for single-wavelength anomalous dispersion have been reviewed by Dauter (2002).

To estimate beforehand how many data are needed for a successful phasing experiment, each experimental parameter must be understood. However, the contribution of many parameters in a sulfur SAD experiment is sometimes seen or considered to be interlinked, such as the Bijovet ratio and the choice of the wavelength, while other parameters are considered separately, such as the resolution limit of the experiment, the exposure time and the redundancy needed; still others, such as the absorbed dose per image, are possibly not considered at all.

Table 1
Data-collection parameters at each wavelength.

The crystal size was ~300 × ~300 × ~300 µm.

Wavelength (Å)	0.979	1.117	1.284	1.540	2.070	2.070	2.290
Exposure time (s)	1.4	1.4	1.4	1.4	1.4	5.4	10.7
2θ (°)	0.0	0.0	0.0	0.0	0.0	30.0	0.0
Rotation interval (°)	1	1	1	1	1	1	1
Detector distance (mm)	145	140	115	85	70	130	55
Slit size (mm)	0.15 × 0.15	0.2 × 0.2	0.1 × 0.1	0.1 × 0.1	0.2 × 0.2	0.2 × 0.2	0.15 × 0.15
Flux (photons s^{-1})†	5.02×10^{10}	3.3×10^{11}	6.7×10^{10}	3.66×10^{10}	3.07×10^{10}	3.07×10^{10}	1.47×10^{10}
Dose rate (Gy s^{-1})	1.11×10^{3}	4.74×10^{3}	4.95×10^{3}	3.6×10^{3}	1.02×10^{3}	1.02×10^{3}	1.0×10^{3}

In this study, data sets have been collected from crystals of lysozyme at wavelengths between 2.290 and 0.979 Å and analysed in order to evaluate the complementary importance of some of the variables above.

All the crystals used in the experiments were of very good quality, as shown by the data-collection statistics, so that the results could be interpreted independently of this. This is a strength but of course also a limitation of the study, together with the fact that only one crystal system has been investigated. This lysozyme crystal form has a solvent content of ~37%. Thus, the contribution of solvent flattening to resolving the inherent phase ambiguity in SAD phasing for such a case is obviously not as effective as for higher solvent contents. This is, therefore, a further limitation of the current study. This work aims to show the relationship between some basic experimental variables and their influence on phasing to make the difference between an uninterpretable electron-density map and a map where the main chain can be autotraced.

2. Methods

2.1. Sample preparation

Lysozyme powder was purchased from Sigma–Aldrich and was used without further purification. Tetragonal crystals were grown using the hanging-drop vapour-diffusion method by mixing equal amounts of the protein and precipitant solutions at 293 K. The protein-solution droplet (10 µl) consisted of 3% lysozyme and 2% NaCl in 0.1 *M* sodium acetate pH 4.2. The reservoir solution (1 ml) was made up of 4%(*w*/*v*) NaCl in 0.1 *M* sodium acetate pH 4.2. Crystals were quickly immersed into reservoir solution with 10% glycerol added and then into reservoir solution with 20% glycerol added, *i.e.* as a two-step process. Each crystal so treated was mounted in a Hampton cryoloop on an Oxford flat base and flash-frozen without further manipulation. The crystal orientation in the loop was arbitrary, *i.e.* 'random'.

2.2. Data collection

Data were collected on the 2.4 T multipole wiggler beamline 10 at the SRS (Daresbury, UK) using a MAR 225 CCD detector. The design of beamline 10 was optimized for delivering softer X-rays (Chayen *et al.*, 2000; Helliwell, 2004; Djinovic Carugo *et al.*, 2005) with a minimized Be-window thickness and also with a 2θ arm for tilting the detector up to 30° so that at the closest distance a maximum 2θ of 110° is achievable, thus reaching into the back-scattering case (Cianci *et al.*, 2005).

Crystals of an average size of 300 µm in diameter were cooled at 100 K using a CryoJet system from Oxford Instruments. The Si(111) double-crystal monochromator was set at various wavelengths so that about 1000 diffraction images were collected at 0.979 Å, about 450 images at 1.117 Å, 360 images at 1.284 Å, 360 images at 1.540 Å, 360 images at 2.070 Å (plus 360 images with the detector at a tilt of 30°) and finally 360 images at 2.290 Å. A fresh crystal was used for each wavelength data set. Data collection was performed in time mode and with continuous rotation without measuring Bijvoet reflection pairs close in time. No helium-purged path was placed between the sample and the detector. Photon flux was monitored during data collection using the MAR desktop beamline ion-chamber reading. The values were converted into photons s^{-1} using calibration measurements collected using a pin diode over the full energy range of the beamline. Data-collection parameters are reported in Table 1.

2.3. Data processing and experimental phasing protocols

In order to assess the influence of wavelength and redundancy on phasing, a fixed standard set of programs was used. Images were processed using *HKL*-2000 (Otwinowski & Minor, 1997) for integration and scaling. The absorption-correction option during scaling was used for all the jobs. The merged data were input to *SHELXD* (Schneider & Sheldrick, 2002) to locate the anomalous scatterers. $R_{\rm r.i.m.}$ and $R_{\rm p.i.m.}$ were calculated using the *RMERGE* code (Weiss, 2001, and references therein). Intensities were converted to structure factors using *TRUNCATE* (Collaborative Computational Project, Number 4, 1994). Experimental phasing was performed using *MLPHARE* (Otwinowski, 1991) and *DM* (Cowtan, 1994). $\langle|\Delta_{\rm ano}|/\sigma(\Delta_{\rm ano})\rangle$ and $\langle|\Delta_{\rm ano}|\rangle/\langle F\rangle$ were calculated using *SFTOOLS* (Collaborative Computational Project, Number 4, 1994).

The electron-density map correlation between each final experimental map and each refined model was calculated using *OVERLAPMAP* (Collaborative Computational Project, Number 4, 1994). The weighted mean phase error was calculated using the *CLIPPER* utility *PHASEMATCH* as

Table 2
Data processing for 360° of rotation at different wavelengths.

Values in parentheses are for the highest resolution shell.

Data set	0.979-360	1.117-360	1.284-360	1.540-360	2.070-360	2.290-360
Wavelength (Å)	0.979	1.117	1.284	1.54	2.070	2.29
Detector distance (mm)	145	140	115	85	70	55
Resolution (Å)	50.0–1.51 (1.56–1.51)	50.0–1.68 (1.74–1.68)	50.0–1.70 (1.76–1.70)	50.0–1.73 (1.79–17.3)	50.0–2.13 (2.21–2.13)	50.0–2.13 (2.21–2.13)
R_{merge} (%)	4.8 (15.8)	4.9 (13.7)	3.5 (13.2)	3.4 (12.6)	5.1 (11.2)	5.1 (18.1)
$R_{p.i.m.}$ (%)	0.9 (3.1)	0.9 (2.6)	0.7 (2.5)	0.7 (2.5)	1.0 (2.4)	1.0 (4.0)
$R_{r.i.m.}$ (%)	4.7 (16.3)	4.7 (13.8)	3.7 (13.1)	3.5 (12.6)	6.2 (11.6)	5.1 (20.3)
Multiplicity	28.4	28.2	27.4	27.0	25.8	25.3
$I/\sigma(I)$	69.5 (24.6)	82.0 (28.4)	116.5 (27.4)	95.7 (26.7)	66.9 (30.2)	60.9 (21.2)
Completeness (%)	100.0 (100.0)	99.8 (98.1)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	99.4 (94.6)

implemented in *CCP*4 program suite v.6.0.2 (Collaborative Computational Project, Number 4, 1994). Experimental phases were compared with the σ_A-weighted model phases derived from each refined coordinate set. Autotracing of the main chain was performed using *ARP*/*wARP* v.7.0 (Morris *et al.*, 2003) running with *REFMAC*5.4 (Murshudov *et al.*, 1996). Difference Fourier maps such as $wF_{(\text{first 120 images})} - wF_{(\text{last 120 images})}$ were calculated using *SIGMAA* weighting with a *w* coefficient (Read, 1986). Figures were prepared using the *CCP*4 molecular-graphics suite (Potterton *et al.*, 2004).

3. Results

3.1. Wavelength dependencies

3.1.1. Dependence on the estimated Bijvoet ratio. Fig. 1 shows the variation of $\langle|\Delta_{ano}|\rangle/\langle F\rangle$ as a function of the wavelength for a tetragonal lysozyme crystal. When taking into account the presence of ten S atoms and eight Cl^- ions (Dauter *et al.*, 1999), $\langle|\Delta_{ano}|\rangle/\langle F\rangle$ varies from a value of 3.11% at an energy of 5.465 keV to 0.56% at 12.658 keV. The limit for a successful single-wavelength experiment is currently seen to be 0.46% (Wang *et al.*, 2006).

By processing 360° of data for each wavelength, the effect of the wavelength on the Bijvoet ratio can be observed experimentally. The processing statistics for each data set are reported in Table 2.

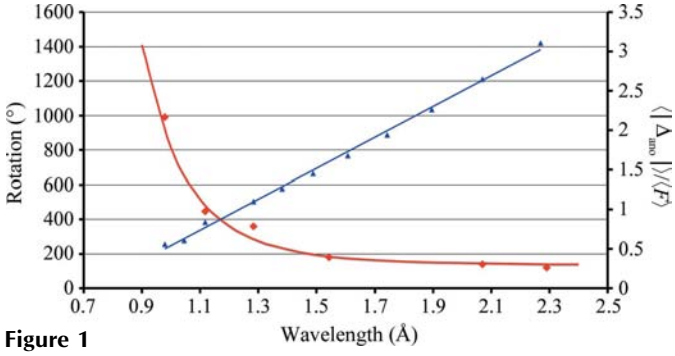

Figure 1
Red curve: plot of the minimum number of degrees of rotation needed *versus* wavelength. Experimental points were fitted with a polynomial function of the type (degrees of rotation) = $A \times (\text{wavelength})^{-6} + B$, with A = 677.48 and B = 135. Blue line: plot of expected value of $\langle|\Delta_{ano}|\rangle/\langle F\rangle$ (%) (calculated using equation 8 from Olczak *et al.*, 2003) *versus* wavelength.

The plot of $\langle|\Delta_{ano}|\rangle/\langle F\rangle$ *versus* resolution (Fig. 2*a*) for 360° of rotation shows good agreement between the predicted avarage values (Fig. 1) and the experimental value. Fig. 2(*a*) shows that the longer the wavelength the higher the Bijvoet ratio value, as expected. The large rise in the value at about 2.2 Å resolution probably includes the effect of errors that inflate the values; a bias correction for such an effect has been described (Dodson *et al.*, 1975).

3.1.2. The anomalous signal *versus* resolution. On beamline 10.1 at the SRS (Daresbury, UK; Cianci *et al.*, 2005) the minimum crystal-to-detector distance is 50 mm, the detector has an active area of 225 × 225 mm and there is an accessible wavelength range for the monochromator of between 0.9 and 2.3 Å. This means that without using a 2θ detector-tilt arm, at a

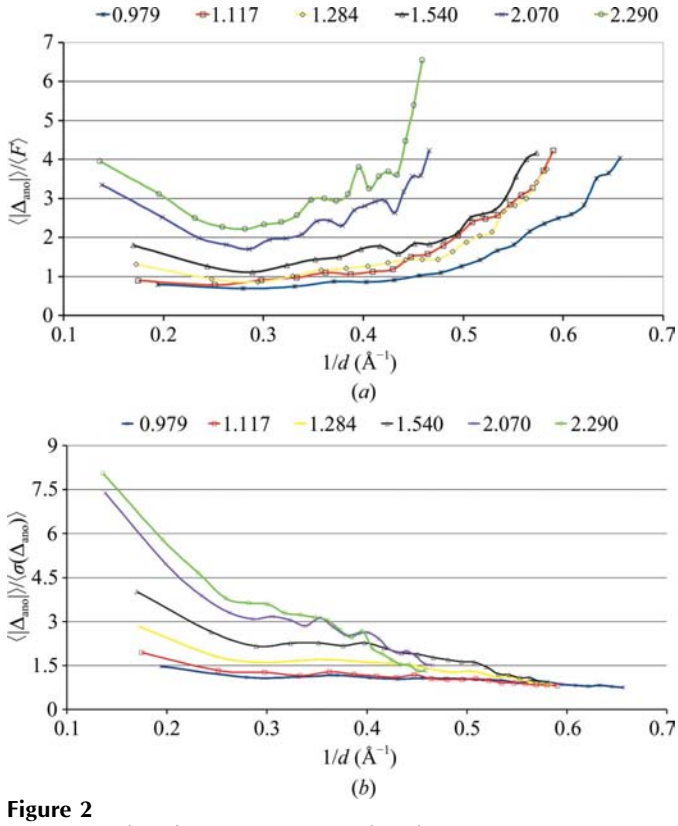

Figure 2
Plot of (*a*) $\langle|\Delta_{ano}|\rangle/\langle F\rangle$ (%) and (*b*) $\langle|\Delta_{ano}|/\sigma(\Delta_{ano})\rangle$ *versus* resolution (expressed as $1/d$) at each different wavelength calculated for data sets with 360° of oscillation.

Table 3
Phasing statistics for 360° of rotation at different wavelengths with phases produced by the three different phasing scenarios.

Scenario	Wavelength (Å)	*SHELXD* max. resolution (Å)	Phasing resolution (Å)	No. of *SHELXD* solutions/tries	Cullis R for ΔF (highest shell)	Lack of closure, ano	Real-space free R, last cycle (%)	Map_CC	Mean phase error w2$\langle d\varphi \rangle$
I	0.979	2.6	2.2	0/1000	—	—	—	—	—
	1.117	2.6	2.2	0/1000	—	—	—	—	—
	1.284	2.6	2.2	1/1000	0.82	1.99	0.20	0.36	64.3
	1.540	2.6	2.2	13/100	0.77	2.32	0.21	0.45	59.5
	2.070	2.6	2.2	17/100	0.76	2.95	0.25	0.47	59.4
	2.290	2.6	2.2	14/100	0.81	2.86	0.25	0.50	58.2
II	0.979	2.0	2.2	0/1000	—	—	—	—	—
	1.117	2.2	2.2	0/1000	—	—	—	—	—
	1.284	2.2	2.2	4/100	0.73	1.76	0.23	0.52	53.1
	1.540	2.2	2.2	25/100	0.64	1.93	0.19	0.57	48.6
	2.070	2.6	2.2	17/100	0.78	2.95	0.24	0.47	59.4
	2.290	2.7	2.2	14/100	0.80	2.86	0.25	0.49	59.5
III	0.979	2.0	1.5	0/1000	—	—	—	—	—
	1.117	2.2	1.68	0/1000	—	—	—	—	—
	1.284	2.2	1.70	4/100	0.83	1.75	0.23	0.69	45.2
	1.540	2.2	1.73	25/100	0.73	1.83	0.17	0.73	42.5
	2.070	2.6	2.13	17/100	0.76	2.96	0.20	0.50	58.2
	2.290	2.7	2.16	14/100	0.81	2.86	0.27	0.50	58.3

wavelength of 1.5 Å with the detector at the closest position, the highest resolution is 1.38 Å at the edges of the detector and 1.27 Å at its corners. At a wavelength of 2.3 Å the highest resolution is 2.11 Å at the edges and 1.94 Å at the corners. Therefore, for a crystal which diffracts beyond 2.0 Å and for a long wavelength such as 2.3 Å, only a data set to 2.11 Å resolution would be collected unless the detector is tilted.

The plot of $\langle|\Delta_{\rm ano}|/\sigma(\Delta_{\rm ano})\rangle$ *versus* resolution (Fig. 2*b*) for 360° of rotation shows the benefits of using a longer wavelength. Data collected at 0.979 Å and at 1.117 Å have $\langle|\Delta_{\rm ano}|/\sigma(\Delta_{\rm ano})\rangle$ around 0.8, as expected, in almost all resolution ranges, while for the other wavelengths $\langle|\Delta_{\rm ano}|/\sigma(\Delta_{\rm ano})\rangle$ is over 1.5 up to 2.0 Å resolution. Considering that only the 1.284, 1.540, 2.070 and 2.290 Å data sets gave a *SHELXD* solution (Table 3), the difference between successful and unsuccessful data sets is quite clear. For the same amount of data a longer wavelength simply gives better measured anomalous differences. Fig. 2(*b*) also shows evidence of an optimum wavelength for data sets with the same redundancy, in the sense that the 2.290 Å wavelength curve is generally better even than the 2.070 Å wavelength curve. This would suggest that an even longer wavelength, at least using this criterion, would be even better.

3.1.3. Phasing. The effect of the choice of the wavelength on the quality of phasing was analysed by taking data sets of equal redundancy (see Table 3). Data for the first 360° of crystal rotation were integrated and scaled for each wavelength. Applying different resolution cutoffs at the anomalous scattering location stage and at the phasing and density-modification stage produced three different phasing scenarios.

In the first scenario, the resolution of the data for the determination of the anomalous scatterer positions was cut to 2.6 Å and for the phasing and density modification the resolution was truncated to 2.2 Å. With these settings, the data sets collected at wavelengths of 0.979 and 1.117 Å did not produce any plausible *SHELXD* solution. The data set collected at 1.284 Å only produced a solution after 1000 tries and the final phases had a poor weighted mean phase error of 64.3° (map correlation factor of 0.36; Fig. 3*a*). For the following data sets the number of *SHELXD* solutions and the quality of the final calculated phases rapidly increased from shorter to longer wavelengths.

In the second scenario, the anomalous scatterer location resolution cutoff was left to *SHELXC*, which sets the data resolution to about 0.5 Å worse than the diffraction limit of the crystal employed for data collection (Schneider & Sheldrick, 2002). Phasing and density-modification resolution was set to 2.2 Å.

With these settings, the data sets collected at wavelengths of 0.979 and 1.117 Å did not produce a plausible *SHELXD* solution. The data set collected at 1.284 Å produced four clear *SHELXD* solutions over 100 tries at a resolution of 2.2 Å and the final phases had a weighted mean phase error of 53.1° (map correlation factor of 0.52). The data set collected at 1.540 Å wavelength produced 25 clear *SHELXD* solutions over 100 tries at a resolution of 2.2 Å. The calculated phases had a weighted mean phase error of 48.6° (map correlation factor of 0.57). *SHELXD* found 17 solutions over 100 tries with the data set collected at 2.070 Å wavelength at a resolution of 2.6 Å. The calculated phases had a weighted mean phase error of 59.4° (map correlation factor of 0.47). Finally, the data set collected at 2.290 Å wavelength produced 14 clear *SHELXD* solutions over 100 tries at a resolution of 2.7 Å. The final phases had a weighted mean phase error of 59.5° (map correlation factor of 0.49). These values are in agreement with the results reported by Watanabe *et al.* (2005) for the same type of experiment (λ = 2.290 Å, 50.0–2.17 Å resolution, 720° total rotation range) and the same software was used for substructure determination and for phasing.

In the second scenario, the 1.284 and 1.540 Å data sets were cut by *SHELXC* to a resolution of 2.2 Å, while the 2.070 and 2.290 Å wavelength data sets did not extend beyond 2.6 and

2.7 Å resolution. Thus, the quality of the phases for the 1.284 Å (Fig. 3*b*) and 1.540 Å data sets are better than for the first scenario, as one might expect. The improvement of the quality is a consequence of the higher resolution used to locate the anomalous scatterers.

In the third scenario, the anomalous scatterer location resolution cutoff was again left to *SHELXC* and the highest resolution available was used for the phasing and density-modification resolution. The mean phase error and map correlation factor improved slightly for the 2.070 and 2.290 Å wavelength cases. The data sets collected at wavelengths of 0.979 and 1.117 Å still did not produce any plausible *SHELXD* solutions, but the addition of the high-resolution data clearly improved the phasing statistics for the data sets collected at 1.284 and 1.540 Å wavelength, the phases of which now had a weighted mean phase error of 45.2° (map correlation factor of 0.69; Fig. 3*c*) and 42.5° (map correlation factor of 0.73), respectively.

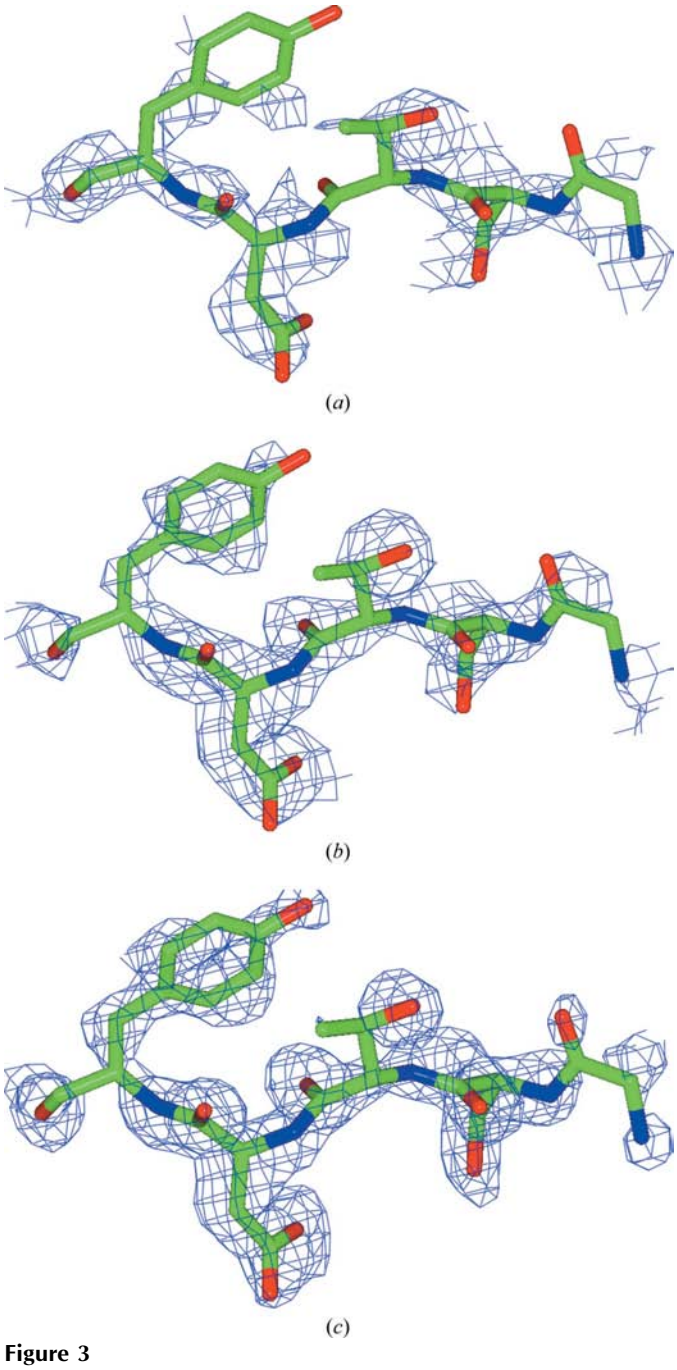

Figure 3
Fourier maps (1.5σ cutoff) covering residues 49–53 and calculated for the 360-image data set collected at 1.284 Å wavelength with phases produced with (*a*) anomalous scatterers located with a data cutoff of 2.6 Å and phases calculated to 2.2 Å, (*b*) anomalous scatterers located with a data cutoff of 2.2 Å and phases calculated to 2.2 Å and (*c*) anomalous scatterers located with a data cutoff of 2.2 Å and phases calculated to 1.7 Å.

Table 4
Data-processing statistics for 2.070 Å data sets using the first 180 images (2.070-180), all 360 images integrated with *HKL*-2000 and scaled using *SCALA* (2.070-360 *SCALA*) and redundancy of 360 images plus a 2θ tilt pass (2.070-2θ).

Values in parentheses are for the highest resolution shell.

Data set	2.070-180	2.070-360 *SCALA*	2.070-2θ
2θ angle (°)	—	—	30
No. of images	180	360	360 + 360
Resolution (Å)	50.0–2.13	50.0–2.13	50.0–1.78
Highest resolution shell (Å)	2.21–2.13	2.24–2.13	1.84–1.78
$R_{\rm merge}$ (%)	3.6 (10.0)	5.4 (11.9)	7.0 (12.9)
$R_{\rm p.i.m.}$ (%)	1.0 (3.2)	1.1 (2.5)	1.1 (5.7)
$R_{\rm r.i.m.}$ (%)	3.7 (10.7)	—	6.8 (14.3)
Multiplicity	12.9	25.6	25.1
$I/\sigma(I)$	85.8 (29.8)	61.0 (22.4)	49.4 (10.5)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	97.9 (88.6)

3.1.4. Scaling protocol. The use of adequate integration and scaling algorithms has been widely advocated and described (Mueller-Dieckmann *et al.*, 2004 and references therein) for data collected using softer X-rays. For this work, a common integrating and scaling protocol was adopted for all the data sets, including those collected at low energies. To test this choice, after being integrated with *DENZO* (*HKL*-2000; Otwinowski & Minor, 1997), the data set collected at 2.070 Å wavelength was also scaled using *SCALA* (Evans, 1993) with the 'SECONDARY BEAM ABSORPTION CORRECTION' option turned on (see Table 4 for processing statistics). *SHELXD* now found 29 solutions over 100 tries with the data set collected at 2.070 Å wavelength using a resolution cutoff at 2.6 Å. After phasing and density modification, the phases had a weighted mean phase error of 58.2° with a map correlation factor of 0.52 (Table 5). The marginal gain obtained at the phasing stage in this case suggests that the scaling protocol implemented in *SCALEPACK* in the *HKL*-2000 suite (Otwinowski & Minor, 1997) with 'ABSORPTION SCALE' turned on was equally appropriate.

3.1.5. 2θ detector-tilt data collection with 2.070 Å wavelength. Data were also collected with the 2θ arm set to a tilt of 30° at 2.070 Å wavelength, which extended the resolution limit of the data set to 1.78 Å resolution (see Table 4 for processing statistics). This in turn resulted in a *SHEXLD* resolution cutoff of 2.3 Å (compared with 2.6 Å for the previous data set without the 2θ tilt data component), yielding 29 clear solutions over 100 tries. Phasing to 2.2 Å resolution produced a weighted mean phase error of 53.3°, equivalent to the values

Table 5
Phasing statistics for different combinations of data collected at 2.070 Å wavelength.

Degrees of rotation	Scenario	*SHELXD* max. resolution (Å)	Phasing resolution (Å)	No. of *SHELXD* solutions/tries	Cullis R for ΔF (highest shell)	Lack of closure, ano	Real-space free R, last cycle (%)	Map_CC	Mean phase error w2⟨dφ⟩
360 *SCALA*†	III	2.6	2.13	29/100	0.78	2.56	0.26	0.52	58.2
180	III	2.6	2.13	10/100	0.79	3.16	0.20	0.47	61.6
140	III	2.6	2.13	2/100	0.86	3.56	0.23	0.41	65.7
2θ	II	2.3	2.2	29/100	0.72	2.82	0.22	0.54	53.3
2.070 Å phases with 0.979 Å data	II	2.6	1.5	17/100	0.78	2.96	0.24	0.49	59.3
2θ	III	2.3	1.78	29/100	0.84	3.00	0.21	0.66	50.0

† Images integrated with *HKL*-2000 and scaled using *SCALA*.

Table 6
Data-processing statistics of minimal data sets for each wavelength.

Values in parentheses are for the highest resolution shell.

Data set	0.979-991	1.117-441	1.284-360	1.540-180	2.070-140	2.290-120
Wavelength (Å)	0.979	1.117	1.284	1.540	2.070	2.290
No. of images	991	441	360	180	140	120
Resolution (Å)	50.0–1.50 (1.50–1.55)	50.0–1.68 (1.74–1.68)	50.0–1.70 (1.76–1.70)	50.0–1.73 (1.79–17.3)	50.0–2.13 (2.21–2.13)	50.0–2.16 (2.24–2.16)
$R_{\rm merge}$ (%)	5.0 (18.6)	4.7 (13.4)	3.5 (13.2)	3.2 (11.3)	3.5 (9.5)	3.9 (15.6)
$R_{\rm p.i.m.}$ (%)	0.5 (1.7)	0.8 (2.3)	0.7 (2.5)	0.9 (3.2)	1.1 (3.5)	1.4 (5.3)
$R_{\rm r.i.m.}$ (%)	4.2 (14.7)	4.7 (13.7)	3.7 (13.1)	3.4 (11.8)	3.6 (10.4)	4.1 (15.7)
Multiplicity	78.0	34.8	27.4	13.5	10.0	8.5
$I/\sigma(I)$	120.2 (31.4)	91.3 (31.6)	116.5 (27.4)	68.0 (19.6)	72.6 (24.8)	68.7 (15.2)
Completeness (%)	98.9 (89.1)	99.8 (98.1)	100.0 (100.0)	99.9 (100.0)	99.9 (99.9)	99.8 (99.9)

Table 7
Phasing statistics of 'minimalized redundancy' test data sets for each wavelength with phases produced by the three different phasing scenarios.

Scenario	Wavelength (Å)	Degrees of rotation	*SHELXD* max. resolution (Å)	Phasing resolution	No. of *SHELXD* solutions/tries	Cullis R for ΔF (highest shell)	Lack of closure, ano	Real-space free R, last cycle (%)	Map_CC	Mean phase error w2⟨dφ⟩
I	0.979	991	2.6	2.2	0/1000	—	—	—	—	—
	1.117	441	2.6	2.2	1/1000	0.96	2.03	0.30	Map not clear	—
	1.284	360	2.6	2.2	1/1000	0.82	1.99	0.20	0.36	64.3
	1.540	180	2.6	2.2	6/100	0.79	2.64	0.22	0.48	54.9
	2.070	140	2.6	2.2	2/100	0.84	3.54	0.21	0.41	65.2
	2.290	120	2.6	2.2	4/100	0.84	3.21	0.21	0.37	68.3
II	0.979	991	2.0	2.2	5/1000	0.93	1.55	0.30	Map not clear	—
	1.117	441	2.2	2.2	1/100	0.92	1.93	0.29	0.35	64.4
	1.284	360	2.2	2.2	4/100	0.73	1.76	0.23	0.52	53.1
	1.540	180	2.2	2.2	6/100	0.79	2.64	0.22	0.49	54.9
	2.070	140	2.6	2.2	2/100	0.84	3.54	0.21	0.41	65.2
	2.290	120	2.7	2.2	4/100	0.84	3.21	0.21	0.37	68.3
III	0.979	991	2.0	1.5	1/100	0.92	1.32	0.17	0.67	46.9
	1.117	441	2.2	1.68	1/100	0.96	1.82	0.27	0.48	60.9
	1.284	360	2.2	1.70	4/100	0.83	1.75	0.23	0.69	45.2
	1.540	1–180	2.2	1.73	6/100	0.85	2.44	0.17	0.65	46.2
	1.540	181–360	2.2	1.73	2/100	0.86	2.59	0.22	0.58	51.6
	2.070	140	2.6	2.13	2/100	0.86	3.56	0.23	0.41	65.7
	2.290	120	2.7	2.16	4/100	0.85	3.10	0.21	0.37	68.4

obtained for the 1.540 and 1.284 Å data sets of 54.9 and 53.1°, respectively (Table 7, scenario II). When applying the full resolution range for phasing and density modification, the weighted mean phase error was 50.0° with a map correlation of 0.66. Fig. 4 shows the map-quality improvement with different resolution and redundancy.

It could be argued that an alternative approach to the data collection with the 2θ arm for the detector could have been to use the 2.070 Å wavelength 360° crystal rotation data set for phasing and then to extend the phases with the high-resolution 0.979 Å data set. The results of this approach, with a 2.6 Å resolution *SHELXD* cutoff and 1.5 Å resolution for phasing,

is a weighted mean phase error of 59.3° (map correlation 0.49; Table 5).

3.2. Analyses *versus* redundancy

3.2.1. On $\langle|\Delta_{ano}|/\sigma(\Delta_{ano})\rangle$ and on $\langle|\Delta_{ano}|\rangle/\langle F\rangle$. Fig. 2(*b*) poses the question whether it would be possible to locate the sulfur positions and phase with them if there was enough redundancy for the 0.979 Å data set. It was found that with 991° (see Table 6 for processing statistics) of sample rotation it was indeed possible to measure anomalous differences with an accuracy high enough to yield the positions of the anomalous scatterers (Table 7).

The benefits of redundancy on data quality can be seen in the plot of $\langle|\Delta_{ano}|\rangle/\langle F\rangle$ *versus* resolution (Fig. 5*a*). The curves calculated for 360°, 540°, 720° and 991° of sample rotation range approach the expected value of $\langle|\Delta_{ano}|\rangle/\langle F\rangle$ with the increased redundancy. It has previously been proposed (Dauter & Adamiak, 2001) that this plot could be used to judge the accuracy of anomalous measurements. The closer the experimental level of $\langle|\Delta_{ano}|\rangle/\langle F\rangle$ is to the expected value, the higher the accuracy of the collected data. The effect of redundancy is clear in the plot of $\langle|\Delta_{ano}|/\sigma(\Delta_{ano})\rangle$ *versus* resolution (Fig. 5*b*), where the values of $\langle|\Delta_{ano}|/\sigma(\Delta_{ano})\rangle$ *versus* resolution are only above 1.5 up to 2.0 Å resolution when 991° of data are integrated. Considering that only in the last case could a *SHELXD* solution be found, it is evident that a $\langle|\Delta_{ano}|/\sigma(\Delta_{ano})\rangle$ greater than 1.5 up to the highest resolution bins is an essential prerequisite for phasing. While the criterion of experimental $\langle|\Delta_{ano}|\rangle/\langle F\rangle$ close to its expected value would be an important qualitative criterion, $\langle|\Delta_{ano}|/\sigma(\Delta_{ano})\rangle > 1.5$ should be a more definitive quantitative

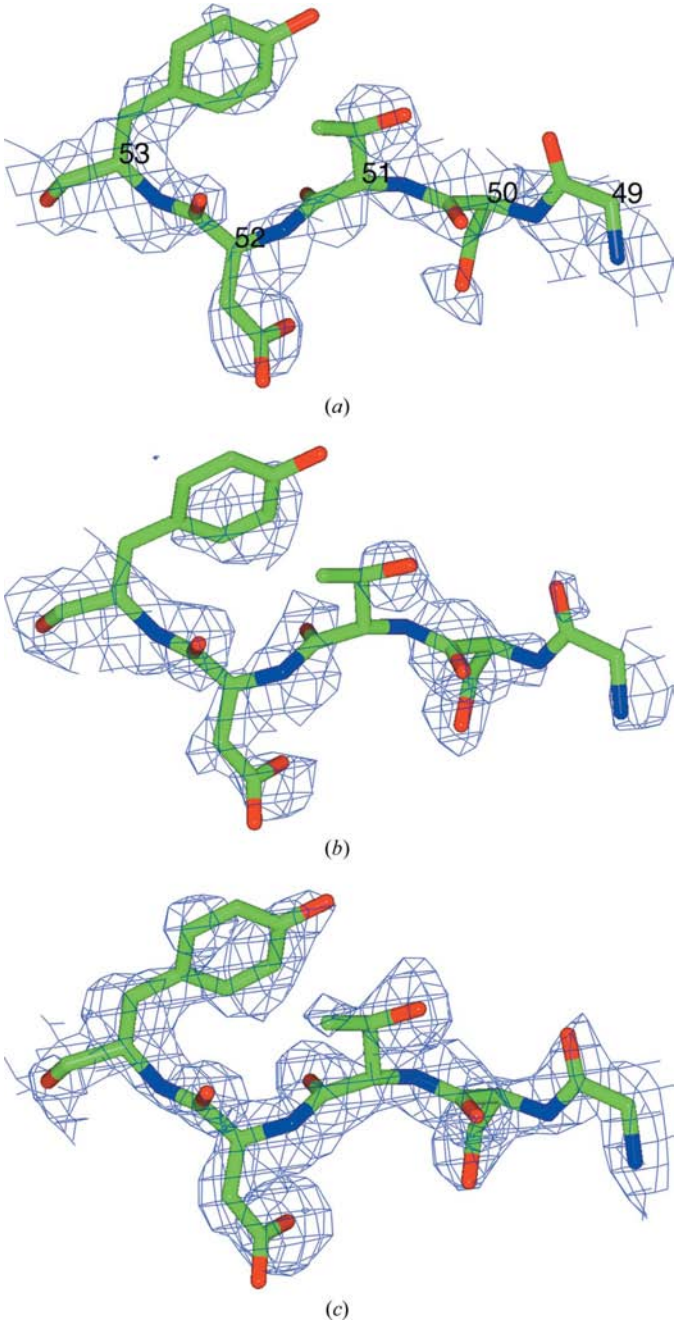

Figure 4
Fourier maps (1.5σ cutoff) covering residues 49–53 and calculated for the data set collected at 2.070 Å wavelength: (*a*) redundancy of 180 images, anomalous scatterers located with a data cutoff of 2.6 Å and phases calculated to 2.16 Å, (*b*) redundancy of 360 images, anomalous scatterers located with a data cutoff of 2.6 Å and phases calculated to 2.16 Å and (*c*) redundancy of 360 images plus 2θ tilt pass, anomalous scatterers located with a *SHELXD* data cutoff of 2.2 Å and phases calculated to 1.73 Å.

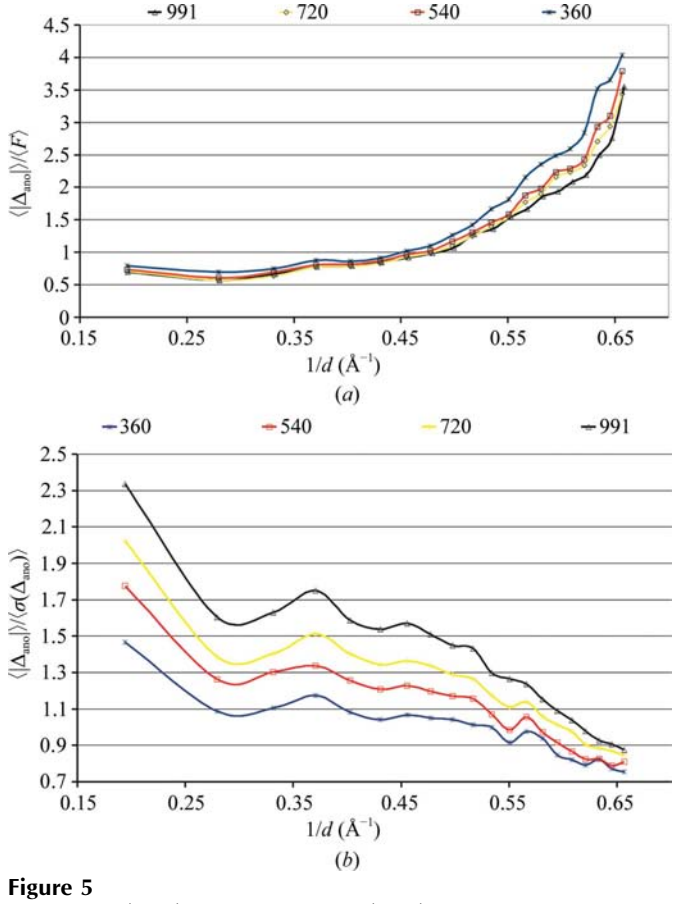

Figure 5
Plot of (*a*) $\langle|\Delta_{ano}|\rangle/\langle F\rangle$ (%) and (*b*) $\langle|\Delta_{ano}|/\sigma(\Delta_{ano})\rangle$ *versus* resolution (expressed as 1/*d*) for the data set collected at 0.979 Å wavelength *versus* number of degrees of oscillation.

criterion to estimate the required redundancy. This latter statistic would prove to be particularly useful if quoted routinely in publications reporting a newly phased structure.

3.2.2. Anomalous scatterer location at different wavelengths. The effect of the redundancy on anomalous scatterer location can be described by testing how many data are needed to find a clear *SHELXD* solution at each different wavelength, while the second question is to ask what happens to the phasing statistics when the redundancy is reduced but it is still possible to obtain a *SHELXD* solution.

In order to verify the effect of data redundancy on phasing, all the data sets were checked for the mimimum quantity of data that was neccessary to obtain a *SHELXD* solution and provide meaningful phasing information. For each data set, the full resolution range available was used. The minimum number of degrees or the minimum redundancy necessary was found to increase with the energy of the photon beam, as expected. At the extremes, with $\lambda = 0.979$ Å 991° were needed before *SHELXD* could locate the anomalous scatterers (one solution over 100 tries), while with $\lambda = 2.290$ Å only 120° were necessary (four solutions over 100 tries; Table 7).

By plotting $\langle|\Delta_{ano}|/\sigma(\Delta_{ano})\rangle$ *versus* resolution for each minimal data set (Fig. 6), it becomes clear that the wavelength and data redundancy both work in increasing the $\langle|\Delta_{ano}|/\sigma(\Delta_{ano})\rangle$ values across all the resolution ranges. This means that the higher the photon energy, the higher the redundancy needed to keep the ratio $\langle|\Delta_{ano}|/\sigma(\Delta_{ano})\rangle$ above 1.5 across all resolution ranges. Provided this is achieved, it seems possible to find an optimum compromise between wavelength and redundancy to allow phasing.

For the 1.117 Å wavelength data set with 441° of rotation it was possible to locate the anomalous scatterers, but the quality of the phases was somewhat lower than for the other data sets of similar resolution. This could be explained by the fact that $\langle|\Delta_{ano}|/\sigma(\Delta_{ano})\rangle$ was just below the threshold of 1.5 (Fig. 6).

The quality of the electron-density maps obtained using longer wavelength data was somewhat poorer and this could be attributed to the lower resolution of the data sets. For instance, this effect has already been observed in the 1.540 Å data sets, in which on changing the resolution at the anomalous scatterer location stage the final weighted mean phase error increased from 48.6° to 59.5° (Table 3), as might be expected.

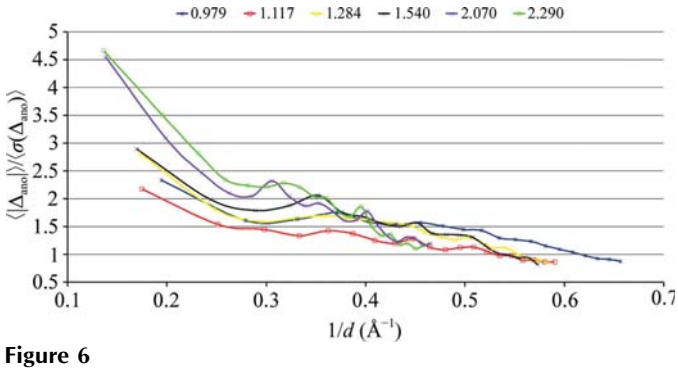

Figure 6
Plot of $\langle|\Delta_{ano}|/\sigma(\Delta_{ano})\rangle$ *versus* resolution (expressed as $1/d$) for each minimal data set.

3.2.3. Phasing. The effect of the redundancy on the phasing statistics was assessed by progressively reducing the redundancy while ensuring that it was still possible to obtain a *SHELXD* solution. The correctness of the positioning of the anomalous scatterers was estimated by looking at the values of the correlation coefficient (CC_{all}/CC_{weak}) statistics reported by *SHELXD* for the best solutions (Schneider & Sheldrick, 2002). The data set collected at 2.070 Å wavelength had an overall resolution of 2.13 Å (Table 4). With a 360° crystal rotation data set *SHELXD* found 17 solutions (Table 3; the best solution had CC_{all}/CC_{weak} statistics of 46.47/32.55) over 100 tries and subsequent phasing gave a weighted mean phase error of 58.2° (map correlation of 0.50). On halving the redundancy, *SHELXD* found ten solutions (the best solution had CC_{all}/CC_{weak} statistics of 45.42/26.34) over 100 tries and subsequent phasing gave a weighted mean phase error of 61.6° (map correlation 0.47). With 140° of rotation, *SHELXD* found two solutions (the best solution had CC_{all}/CC_{weak} statistics of 38.01/16.06) over 100 tries and subsequent phasing gave a weighted mean phase error of 65.7° (map correlation of 0.41; Table 7). Finally, with 120° of rotation *SHELXD* found one solution (the best solution had CC_{all}/CC_{weak} statistics of 30.46/12.26) over 100 tries but no meaningful map was produced. The CC_{all}/CC_{weak} statistics decrease sensibly with the reduced redundancy. Notably, collecting more than 180° of rotation only increases the number of solutions, not their correctness.

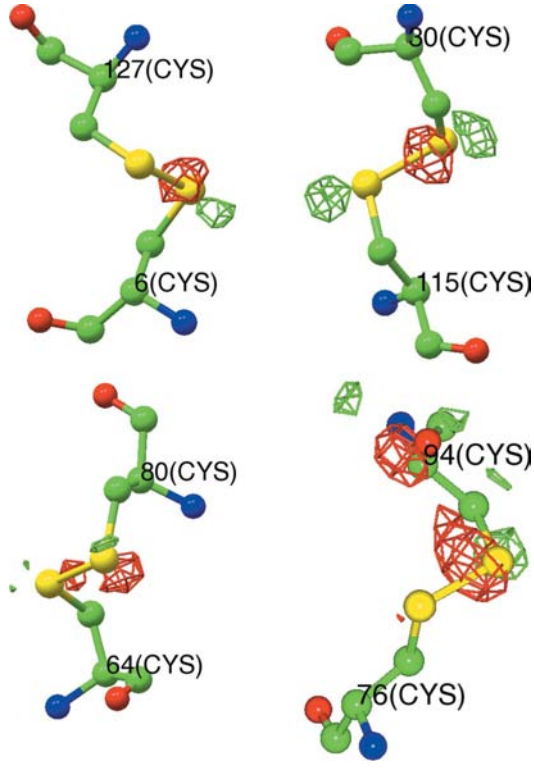

Figure 7
Difference Fourier maps *wF*(images 001–120) − *wF*(images 871–991) (where *w* indicates a *SIGMAA* weighting; Read, 1986) for the data set collected at 0.979 Å wavelength at 1.50 Å resolution. Maps contoured at 4σ (positive density green, negative density red). Signs of early X-radiation damage effects on the four disulfide bonds in HEWL are clearly visible.

3.3. Autotracing with *ARP*/*wARP*

The quality of the experimental phasing of each data set was assessed using the main-chain autotracing algorithm of *ARP*/*wARP* (Morris *et al.*, 2003). Independently of the wavelength used for data collection, only data sets with a final weighted mean phase error less than or equal to 50° could produce a fully autotraced structure. Such data sets were the λ = 0.979 Å with 991° of rotation (structure-factor PDB file: r2w1lsf, Table 7, scenario III), λ = 1.284 Å with 360° (structure-factor PDB file: r2w1xsf, Table 7, scenario III), λ = 1.540 Å with either 360 or 180° of rotation (structure-factor PDB file: r2w1ysf), λ = 2.070 Å with 360° of rotation at 2θ = 0° combined with 2θ tilt = 30° tilt data (mean phase error 50.0°; structure-factor PDB file: r2w1msf, Table 5, scenario III). The combination of the high-resolution 0.979 Å wavelength data set with the 2.070 Å wavelength 360° crystal rotation phases extended from 2.2 to 1.5 Å resolution did not produce a map that was good enough to allow the autotracing exercise to work (mean phase error of 59.3°; Table 5).

3.4. Dose *versus* redundancy analyses

The dose rate was monitored for each experiment by logging the ion-chamber reading and slit positions (Table 1). The ion-chamber readings were then translated into photon flux (photons s^{-1}) values *via* calibration of the flux measurement made with a pin-diode probe (S. G. Buffey, personal communication). The dose rate and total absorbed dose were calculated using *RADDOSE* (Murray *et al.*, 2005). The dose values (Table 8) are calculated for a stationary crystal without taking into account the sample rotation and the consequent change of the volume irradiated. The calculated total dose rate shows that provided a map of quality good enough for autotracing is obtained (see §3.3), a total dose in the range 0.1–0.42 × 10^7 Gy is needed for each phasing experiment; this is independent of the wavelength used for data collection and the redundancy.

Among these data sets, the highest dose was given to the λ = 1.284 Å with 360° rotation data set and the λ = 2.070 Å with 360° rotation at 2θ = 0° and 2θ = 30° data set (0.26–0.29 × 10^7 Gy). It is worth noting that the λ = 2.290 Å with 360° rotation data set was given three times the total dose (0.42 × 10^7 Gy) of the λ = 0.979 Å with 991° rotation data set (0.14 × 10^7 Gy), with completely different results. This might be explained by partial photon absorption by air and less optimal detector efficiency.

3.4.1. Radiation damage in HEWL at a tenth of the Henderson limit. A difference Fourier map $wF_{(\text{images } 001\text{–}120)} - wF_{(\text{images } 871\text{–}991)}$ (where *w* indicates a *SIGMAA* weighting; Read, 1986; Fig. 7) clearly shows signs of early X-radiation damage effects on the four disulfide bonds in HEWL.

All four disulfide bridges show negative density along the S—S bonds and positive density outside (Fig. 7; the disulfide bond between residues 30 and 115 is particularly clear). This can be interpreted as an indication of bond splitting, with the two S atoms moving further apart. With λ = 0.979 Å, radiation damage on HEWL disulfide bridges took place with a total dose of 0.14 × 10^7 Gy (Table 8). Similar doses have been previously reported to cause radiation damage to disulfide bridges (Wang *et al.*, 2006; Shimizu *et al.*, 2007). This is roughly equivalent to a tenth of the predicted absorbed dose of 2 × 10^7 Gy (Henderson, 1990), which is sufficient to reduce the total diffracted intensity by half.

Table 8
Dose calculation for the different data sets by *RADDOSE*.

Wavelength	Total rotation (°)	Resolution (Å)	Dose (Gy)	Henderson dose percentage (%)
0.979	991	50.0–1.50	0.14 × 10^7	7.0
1.117	441	50.0–1.68	0.31 × 10^7	15.5
1.284	360	50.0–1.70	0.26 × 10^7	13.0
1.540	180	50.0–1.73	0.10 × 10^7	5.0
2.070 (2θ = 0°)	360	50.0–2.13	0.6 × 10^6	3.0
2.070 (2θ = 0°)	180	50.0–2.13	0.3 × 10^6	1.5
2.070 (2θ = 30°)	360	50.0–1.78	0.23 × 10^7	11.5
2.070 (overall)	360 + 360	50.0–1.78	0.29 × 10^7	14.5
2.070 + 0.979	360 + 360	50.0–1.78	0.10 × 10^7	5.0
2.290	120	50.0–2.16	0.14 × 10^7	7.0
2.290	360	50.0–2.16	0.42 × 10^7	21.0

However, it was still possible to perform a successful sulfur SAD experiment without radiation damage affecting the quality of the anomalous signal. Data quality overall was not compromised and S-atom positions could still be usefully determined. However, it is not so easy to estimate how many more degrees of rotation could be added without spoiling data collection, since the first signs of radiation damage were already appearing after 991° of rotation. The other data collections (λ = 1.284 Å with 360° rotation, λ = 1.540 Å with 360° rotation and λ = 2.070 Å with 360° rotation plus 2θ data) received total doses of 0.26 × 10^7, 0.10 × 10^7 and 0.23 × 10^7 Gy, respectively, and still produced successful phasing despite likely radiation damage to disulfide bridges. On the other hand, the λ = 1.117 Å with 441° of rotation data set received a total dose of about 0.31 ×10^7 Gy and still gave S-atom positions, but the phasing statistics were of somewhat lower quality when compared with other data sets. A dose of 0.31 × 10^7 Gy could thus already be producing too much damage to the protein structure at the atomic level and affecting the phasing process.

3.4.2. Consequences for phasing. This was assessed by searching for anomalous scatterers in two separate processed wedges (one for the first 180° of rotation and one for the second 180° of rotation) of the data set collected at 1.540 Å wavelength. Despite the fact that the second wedge only accumulated a dose of 0.10 × 10^7 Gy greater than the first, the number of the anomalous scatterers found by *SHELXD* and their *MLPHARE* refined positions differ (see supplementary material[1]). All of the sulfurs (ten atoms) and three chloride ions were located (six solutions over 100 tries) from the anomalous signal of the first wedge , which in turn produced a final weighted mean phase error of 46.6° (map correlation of 0.65). On the other hand, only the SG atoms of Cys64, Cys76,

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: EA5090). Services for accessing this material are described at the back of the journal.

Cys80, Cys94, Cys115 and Cys127, and the SD atoms of Met12 and Met105 as well as two Cl atoms were located clearly (two solutions over 100 tries) from the anomalous signal of the second wedge. Cys30 SG was not located at all, while Cys6 SG could perhaps be generously attributed to a peak 3.84 Å away from the expected position. For the SG atoms of Cys64, Cys76, Cys80, Cys94, Cys115 and Cys127, and the SD atoms of Met12 and Met105 there were also marginal movements of around 0.1 Å. In turn, phasing of the second wedge gave poorer statistics, with a final weighted mean phase error of 51.1° (map correlation of 0.59; Fig. 8, yellow curve). At this point, the second wedge had accumulated a dose of 0.20×10^7 Gy. Clearly, the two wedges behave as two separate data sets. In this case the fraction of the sulfurs in the wrong positions is not so many of the total, but still the X-rays can cause changes.

The putative disappearance or small movements of S atoms upon irradiation affects the overall quality of phasing. In fact, when using the anomalous scatterer positions derived from the 181–360° wedge to phase the 1–180° wedge, the weighted mean phase error is slightly worse at all resolutions (Fig. 8, red curve) than that produced when using the anomalous scatterer positions derived from the same wedge (Fig. 8, blue curve). In this case, one or more atoms were not located. Small atom movements upon X-irradiation can be simulated by applying small (0.5 Å) coordinate changes to the anomalous scatterer positions. In this case, the phases calculated for the the 1–180° wedge are affected more at high resolution (Fig. 8, green curve).

4. Discussion

4.1. Role of wavelength

The longer wavelengths gave a higher success rate in locating the anomalous scatterers and producing useful phases at low redundancy. In fact, with one 360° crystal rotation redundancy there were no *SHELXD* solutions after 1000 trials with the λ = 0.979 and 1.117 Å data sets, while with data sets at wavelengths between 1.540 and 2.290 Å there were *SHELXD* solutions after tens of trials, all of which produced useful phases (Table 3).

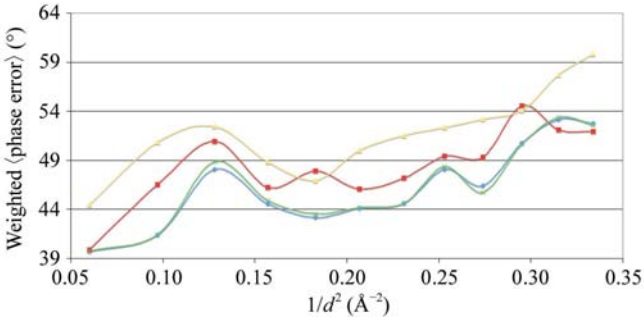

Figure 8
Weighted mean phase error *versus* resolution (expressed as $1/d^2$) for the 1.540 Å wavelength data set phased using the first 1–180° wedge with the *SHELXD* solution from the same data set (blue curve), phased using the first 1–180° wedge with the *SHELXD* solution from the same data set with small (0.5 Å) coordinate changes (green curve), phased using the second 181–360° wedge with the *SHELXD* solution from the same data set (yellow curve) and phased using the first 1–180° wedge with the *SHELXD* solution from the second 181–360° wedge (red curve).

When a crystal diffracts to beyond 2.0 Å, as in this case with lysozyme, a shorter wavelength and a highly redundant data set can yield phases of quality comparable to those produced by a data set collected with a longer wavelength with a mean phase error value of 46.9° (Table 7, scenario III). As shown, it is feasible to collect a very highly redundant data set at shorter wavelength (λ < 1.540 Å) to perform a sulfur SAD experiment.

The previous observations can be explained by looking at the possible correlation between resolution and the $\langle|\Delta_{ano}|\rangle/\langle F\rangle$ value. From Fig. 1 and Table 7, it is noticeable that in this experiment with a $\langle|\Delta_{ano}|\rangle/\langle F\rangle$ value of 0.56 (at λ = 0.979 Å) the minimum resolution necessary to find the substructure with five successes in 1000 trials was 2.0 Å and with a $\langle|\Delta_{ano}|\rangle/\langle F\rangle$ value of 1.1 (at λ = 1.284 Å) the minimum resolution necessary for four successes in 100 trials was 2.2 Å, while with a $\langle|\Delta_{ano}|\rangle/\langle F\rangle$ value of 1.68 (at λ = 1.540 Å) with six successes in 100 trials the minimum resolution necessary was 2.6 Å. The wavelength for the experiment should be chosen to make efficient use of the diffracting capabilities of the crystal and to keep the anomalous diffraction ratio $\langle|\Delta_{ano}|\rangle/\langle F\rangle$ above a reasonable value.

When resolution cutoffs were applied at the anomalous scatterer location or phasing stages, the differences between data sets collected with high-energy or low-energy X-rays were emphasized. With a *SHELXD* resolution of 2.6 Å and a phasing resolution of 2.2 Å, scenario I reproduced the difficulties of phasing a moderately low-resolution structure with the use of the anomalous effect of sulfur. The mean phase error values were not only dependent on the final phasing resolution, but also on the resolution used by *SHELXD* for the location of the anomalous scatterers, as proved by the second scenario (Tables 3 and 7). In fact, the quality of the phases as compared between the data sets showed, for example, that the mean phase error of 59.4° with one 360° crystal rotation with λ = 1.540 Å (Table 3; scenario I) was high compared with the value of 50.0° for λ = 2.070 Å with 2θ tilt data (Table 5, scenario III). Thus, when longer wavelengths are used it is advisable to use a 2θ detector tilt to collect a high-resolution data pass and include these data for phasing. If a 2θ tilt is not available at the beamline or if the crystal has a low-symmetry space group, then the use of a shorter wavelength should be promptly considered in order to preserve completeness even if this is at the cost of the number of images. The speed of data collection and data storage are much less of an issue than in the past, so this should be feasible.

4.2. Effect of redundancy and dose

All five data sets (λ = 0.979 Å with 991° of rotation, λ = 1.284 Å with 360° of rotation, λ = 1.540 Å with either 360° or 180° of rotation and λ = 2.070 Å with 360° of rotation at 2θ = 0° or combined with 2θ = 30° data) had a final weighted mean phase error that was smaller than 50°. That means that provided similar resolutions could be achieved, there was no

evident difference in phase quality between a data set collected with shorter or longer wavelengths.

It is therefore interesting to plot a graph of the number of degrees of rotation needed to locate the anomalous scatterers *versus* the data-collection wavelength (Fig. 1). This curve is basically a plot of the number of degrees of rotation per data set needed to achieve a $\langle|\Delta_{ano}|/\sigma(\Delta_{ano})\rangle$ greater than 1.5 for a lysozyme crystal diffracting to 1.5 Å. As such, the curve is a function of properties of the sample such as the Bijvoet ratio, the space group and all the experimental variables that are a property of the instrument used for data collection, such as the flux or the detector used. Despite the complexity of the function, the curve appears to follow a predictable trend while having to follow two experimental constraints.

The first constraint is that of locating anomalous scatterers successfully, for which the ratio $\langle|\Delta_{ano}|/\sigma(\Delta_{ano})\rangle$ should be at least 1.5 across all resolution bins. This condition is shared by other protein structure data sets for which a successful sulfur-substructure determination has been reported [see Fig. 1(*f*) of Wagner *et al.* (2006) and Fig. 2(*d*) of Kitago *et al.* (2005)]. The second constraint is that the total dose should then probably stay within $0.20–0.25 \times 10^7$ Gy or one tenth of the Henderson limit, as shown by these experiments.

For a lysozyme crystal diffracting to better than 1.75 Å resolution, it was possible to collect sufficient data to produce good-quality phases to autotrace the structure with such a dose, using different combinations of wavelength and redundancy. The dose of one tenth of the Henderson limit is characteristic of HEWL, but it may be different for other crystal samples. However, early signs of X-radiation damage are observed for the four disulfide bonds in HEWL.

It should also be borne in mind that for crystals with dimensions that exceed the beam size, the dose calculated using a stationary crystal model is not a good approximation. In fact, when the sample-volume change during rotation is considered for dose calculations, the average dose values are obviously lower. The rotation must be taken into account for very detailed estimation of radiation damage (Schulze-Briese *et al.*, 2005).

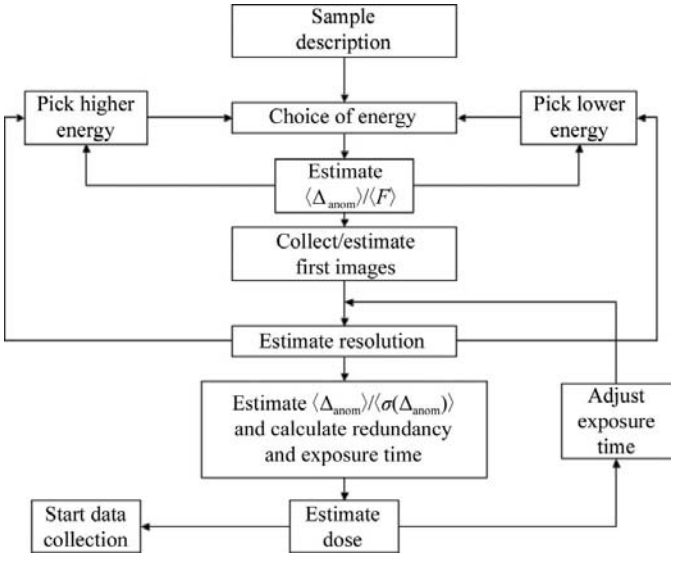

Figure 9
Proposed workflow of the decision-making model for a sulfur SAD experiment.

4.3. Suggestions for planning a sulfur SAD experiment

Within the limitations arising from the fact that only one crystal system has been investigated in this study, a possible strategy for planning a sulfur SAD experiment may be suggested. It now appears possible to change or (better) optimize wavelength, exposure time and consequently the redundancy needed of a data set in an anomalous data-collection experiment while keeping the total final dose constant.

The sample description should give the first indications about the feasibility of a sulfur SAD experiment such as the wavelength to be used to obtain a good Bijvoet ratio. The investigator should then decide on which beamline to perform the experiments. When at the beamline, the first diffraction images will give an indication of the resolution limit of the crystal and the overall value of $I/\sigma(I)$. These values should in turn be used to estimate the exposure time, the orientation of the crystal and the number of images needed to obtain the ratio $\langle|\Delta_{ano}|/\sigma(\Delta_{ano})\rangle > 1.5$ across all the resolution bins. Incidently, a similar approach is currently used by *BEST* (Popov & Bourenkov, 2003; Bourenkov & Popov, 2006) to estimate the final $\langle I/\sigma(I)\rangle$ in the highest resolution shells, optimize data-collection parameters and to obtain complete data sets.

In addition to the sample description (space group, unit-cell parameters *etc.*), *BEST* takes into account instrumental parameters such as the type of detector at the beamline *etc.*

From a practical point of view, the exposure time should be the shortest practically possible. In other words if there does not appear to be much gain in terms of resolution from increasing the exposure of a crystal from 1 to 10 or 15 s, then the shorter exposure should be considered.

The value of the photon flux should be used to evaluate whether or not the suggested total exposure time will give an X-ray dose exceeding the warning limits for radiation damage, in the same way as performed by *RADDOSE* (Murray *et al.*, 2004). If the cycle described (Fig. 9) does not converge on a plausible combination of Bijvoet ratio, wavelength, resolution, redundancy and dose, it should then be reiterated by optimizing one of the parameters until convergence. Prior to setting off to the synchrotron, some preliminary calculations would help in planning and clarifying options.

4.4. Instrumentation for the end station

Other implications arise for the end-station design of beamlines dedicated to the use of long wavelengths and for SAD experiments in general. The end station should be equipped with a detector 2θ arm and its use should be encouraged by beamline scientists, given the fact that the images collected with such a geometry do not pose any problem for integration.

The use of a large area detector with a long wavelength could cause problems with spot profiles and a different absorption cross-section through a phosphor as described by Helliwell (1992). Such problems would be less severe with a detector 2θ arm. The use of a detector tilt and a close distance setting allows back-scattering reflection measurements to be obtained.

Together with the exposure time, the value of the photon flux should be recorded at the time of the data collection. Consequently, an efficient way of measuring photon flux in real time should be in place. Provided good crystals are available, a sulfur SAD experiment could be used to evaluate the quality of a given beamline. Flot *et al.* (2006) showed how the beamline instrumentation can affect the quality of data collection. The need for a finely tuned beamline has been consistently advocated over the years, most recently by Wang *et al.* (2006), as necessary to collect the very weak anomalous signal from a sample with a small Bijvoet ratio. In the future, a 'full back-scattering' option diffractometer, as is common in chemical crystallography, would allow the most flexible choice for instrument usage, *i.e.* longer wavelengths, while still retaining high resolution for both poor and well scattering protein crystals.

5. Conclusions

Data sets have been collected from a single crystal of HEWL at wavelengths between 0.979 and 2.290 Å. Such data sets allowed determination of the substructure and the production of good-quality phases. Radiation damage was found to affect disulfide bridges after the crystals had been given a total dose of around 0.2×10^7 Gy; that is, roughly a tenth of the predicted absorbed dose sufficient to destroy crystalline diffraction from protein crystals (Henderson, 1990). These findings agree with other reports in the literature.

However, with such a given total dose it was still possible to develop a strategy to collect data with $\langle|\Delta_{\rm ano}|/\sigma(\Delta_{\rm ano})\rangle > 1.5$ across most of the resolution bins by choosing the optimum combination of exposure time, wavelength and redundancy for this crystal system. The satisfaction of the rule for $\langle|\Delta_{\rm ano}|/\sigma(\Delta_{\rm ano})\rangle > 1.5$ appears to be a necessary requirement for sulfur SAD substructure location.

The choice of the wavelength should then follow the sample composition and the diffracting properties of the crystal. With the typical current instrumentation, for strongly diffracting crystals wavelengths equal or shorter than 1.540 Å should be chosen (provided the Bijvoet ratio is reasonable) so as to not lose diffraction data owing to detector aperture effects, while a longer wavelength can be used for weakly diffracting crystals (and also for strongly diffracting crystals provided a tilting detector or a back-scattering solution is available) to take advantage of the increased sulfur anomalous signal. These results, with the limitations imposed by the study of one single-crystal system, but with the advantage of a versatile instrument design such as SRS beamline 10, suggest a new approach to the preparation of a sulfur SAD experimental strategy. For a given sample, once an acceptable level of total X-radiation dose has been estimated, descriptors of the crystal system and of the instrument should allow assessment of the feasibility of sulfur SAD data collection by computing these as the complete parameters for data collection.

This work is dedicated to the memory of Dr A. P. Fordham-Skelton (STFC, College of Biology and Medicine, Daresbury Laboratory). Beamline 10 at SRS (UK) was funded by BBSRC grants (719/B15474 and 719/REI20571) and an NWDA project award (N0002170). We would like to thank Professor E. Dodson (University of York, England) for her kind suggestions on SAD phasing using the *CCP*4 program suite. We would like to thank Dr T. R. Schneider and Dr G. P. Bourenkov (EMBL Hamburg Outstation, Germany) for helpful discussions. Data were collected under MC's beam time award for beamline 10 commissioning from CCLRC (now STFC), to whom we are also grateful. S. G. Buffey and R. Birchall (STFC, Daresbury Laboratory) are thanked for their dedicated engineering and technical support and useful discussions.

References

Agarwal, R., Bonanno, J. B., Burley, S. K. & Swaminathan, S. (2006). *Acta Cryst.* D**62**, 383–391.
Borek, D., Ginell, S. L., Cymborowski, M., Minor, W. & Otwinowski, Z. (2007). *J. Synchrotron Rad.* **14**, 24–33.
Bourenkov, G. P. & Popov, A. N. (2006). *Acta Cryst.* D**62**, 58–64.
Brown, J., Esnouf, R. M., Jones, M. A., Linnell, J., Harlos, K., Hassan, A. B. & Jones, E. Y. (2002). *EMBO J.* **21**, 1054–1062.
Burmeister, W. P. (2000). *Acta Cryst.* D**56**, 328–341.
Chayen, N. E., Cianci, M., Olczak, A., Raftery, J., Rizkallah, P. J., Zagalsky, P. F. & Helliwell, J. R. (2000). *Acta Cryst.* D**56**, 1064–1066.
Cianci, M. *et al.* (2005). *J. Synchrotron Rad.* **12**, 455–466.
Cianci, M., Helliwell, J. R., Moorcroft, D., Olczak, A., Raftery, J. & Rizkallah, P. J. (2004). *J. Appl. Cryst.* **37**, 555–564.
Cianci, M., Rizkallah, P. J., Olczak, A., Raftery, J., Chayen, N. E., Zagalsky, P. F. & Helliwell, J. R. (2001). *Acta Cryst.* D**57**, 1219–1229.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Dauter, Z. (2002). *Curr. Opin. Struct. Biol.* **12**, 674–678.
Dauter, Z. & Adamiak, D. A. (2001). *Acta Cryst.* D**57**, 990–995.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **28**, 83–92.
Dauter, Z., Dauter, M. & Dodson, E. J. (2002). *Acta Cryst.* D**58**, 494–506.
Debreczeni, J. É., Bunkóczi, G., Girmann, B. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 393–395.
Debreczeni, J. É., Bunkóczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 688–696.
Djinović Carugo, K., Helliwell, J. R., Stuhrmann, H. & Weiss, M. S. (2005). *J. Synchrotron Rad.* **12**, 410–419.
Doan, D. N. & Dockland, T. (2003). *Structure*, **11**, 1445–1451.
Dodson, E. (2003). *Acta Cryst.* D**59**, 1958–1965.
Dodson, E. J., Evans, P. R. & French, S. (1975). *Anomalous Scattering*, edited by S. Ramaseshan & S. C. Abrahams, pp. 423–436. Copenhagen: Munksgaard.
Einspahr, H., Suguna, K., Suddath, F. L., Ellis, G., Helliwell, J. R. & Papiz, M. Z. (1985). *Acta Cryst.* B**41**, 336–341.
Evans, P. R. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 114–122. Warrington: Daresbury Laboratory.

Flot, D., Gordon, E. J., Hall, D. R., Leonard, G. A., McCarthy, A., McCarthy, J., McSweeney, S., Mitchell, E., Nurizzo, D., Ravelli, R. G. B. & Shepard, W. (2006). *Acta Cryst.* D**62**, 65–71.
Garman, E. F. & Mitchell, E. P. (1996). *J. Appl. Cryst.* **29**, 584–587.
Garman, E. F. & Owen, R. L. (2006). *Acta Cryst.* D**62**, 32–47.
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.* D**57**, 1230–1237.
Harrop, S. J., Helliwell, J. R., Wan, T. C. M., Kalb (Gilboa), A. J., Tong, L. & Jariv, J. (1996). *Acta Cryst.* D**52**, 143–155.
Helliwell, J. R. (1988). *J. Cryst. Growth*, **90**, 259–272.
Helliwell, J. R. (1992). *Macromolecular Crystallography with Synchrotron Radiation*, p. 206. Cambridge University Press.
Helliwell, J. R. (2004). *J. Synchrotron Rad.* **11**, 1–3.
Henderson, R. (1990). *Proc. R. Soc. London Ser. B*, **241**, 6–8.
Hendrickson, W. A. & Teeter, M. (1981). *Nature (London)*, **290**, 107–113.
Kauffmann, B., Weiss, M. S., Lamzin, V. S. & Schmidt, A. (2006). *Structure*, **14**, 1099–1105.
Kitago, Y., Watanabe, N. & Tanaka, I. (2005). *Acta Cryst.* D**61**, 1013–1021.
Lemke, C. T., Smith, G. D. & Howell, P. L. (2002). *Acta Cryst.* D**58**, 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.* D**58**, 21–28.
Morris, R. J., Perrakis, A. & Lamzin, V. S. (2003). *Methods Enzymol.* **374**, 229–244.
Mueller-Dieckmann, C., Polentarutti, M., Djinovic Carugo, K., Panjikar, S., Tucker, P. A. & Weiss, M. S. (2004). *Acta Cryst.* D**60**, 28–38.
Murray, J. W., Garman, E. F. & Ravelli, R. B. G. (2004). *J. Appl. Cryst.* **37**, 513–522.
Murray, J. W., Rudiño-Piñera, E., Owen, R. L., Grininger, M., Ravelli, R. B. G. & Garman, E. F. (2005). *J. Synchrotron Rad.* **12**, 268–275.
Murshudov, G., Vagin, A. & Dodson, E. (1996). *Proceedings of the CCP4 Study Weekend. Macromolecular Refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 93–104. Warrington: Daresbury Laboratory.
Nanao, M. H., Sheldrick, G. M. & Ravelli, R. B. G. (2005). *Acta Cryst.* D**61**, 1227–1237.
Nieh, Y.-P. & Helliwell, J. R. (1995). *J. Synchrotron Rad.* **2**, 79–82.
Olczak, A., Cianci, M., Hao, Q., Rizkallah, P. J., Raftery, J. & Helliwell, J. R. (2003). *Acta Cryst.* A**59**, 327–334.
Olsen, J. G., Flensburg, C., Olsen, O., Bricogne, G. & Henriksen, A. (2004). *Acta Cryst.* D**60**, 250–255.
Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Popov, A. N. & Bourenkov, G. P. (2003). *Acta Cryst.* D**59**, 1145–1153.
Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G. N., Cohen, S., Perrakis, A. & Noble, M. (2004). *Acta Cryst.* D**60**, 2288–2294.
Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 1020–1027.
Ravelli, R. B. G. & Garman, E. F. (2006). *Curr. Opin. Struct. Biol.* **16**, 624–629.
Ravelli, R. B. G., Leiros, H.-K., Pan, B., Caffrey, M. & McSweeney, S. (2003). *Structure*, **11**, 217–224.
Ravelli, R. B. G. & McSweeney, S. (2000). *Structure*, **8**, 315–328.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Schulze-Briese, C., Wagner, A., Tomizaki, T. & Oetiker, M. (2005). *J. Synchrotron Rad.* **12**, 261–267.
Shimizu, N., Hirata, K., Hasegawa, K., Ueno, G. & Yamamoto, M. (2007). *J. Synchrotron Rad.* **14**, 4–10.
Southworth-Davies, R. J. & Garman, E. F. (2007). *J. Synchrotron Rad.* **14**, 73–83.
Teng, T.-Y. (1990). *J. Appl. Cryst.* **23**, 387–391.
Wagner, A., Pieren, M., Schulze-Briese, C., Ballmer-Hofer, K. & Prota, A. E. (2006). *Acta Cryst.* D**62**, 1430–1434.
Watanabe, N., Kitago, Y., Tanaka, I., Wang, J., Gu, Y., Zheng, C. & Fan, H. (2005). *Acta Cryst.* D**61**, 1533–1540.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Wang, J., Dauter, M. & Dauter, Z. (2006). *Acta Cryst.* D**62**, 1475–1483.
Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.
Weiss, M. S., Mander, G., Hedderich, R., Diederichs, K., Ermler, U. & Warkentin, E. (2004). *Acta Cryst.* D**60**, 686–695.
Weiss, M. S., Sicker, T., Djinovic-Carugo, K. & Hilgenfeld, R. (2001). *Acta Cryst.* D**57**, 689–695.