Determination of a novel structure by a combination of long-wavelength sulfur phasing and radiation-damage-induced phasing

**Manfred S. Weiss, Gerd Mander, Reiner Hedderich, Kay Diederichs, Ulrich Ermler
and Eberhard Warkentin**

Determination of a novel structure by a combination of long-wavelength sulfur phasing and radiation-damage-induced phasing

Manfred S. Weiss,[a]***** **Gerd
Mander,**[b] **Reiner Hedderich,**[b]
Kay Diederichs,[c] **Ulrich Ermler**[d]
and Eberhard Warkentin[d]

[a]EMBL Hamburg Outstation, c/o DESY,
Notkestrasse 85, D-22603 Hamburg, Germany,
[b]Max-Planck-Institut für terrestrische
Mikrobiologie, Karl-von-Frisch-Strasse,
D-35043 Marburg, Germany, [c]Fakultät für
Biologie, Universität Konstanz, Box M656,
D-78457 Konstanz, Germany, and [d]Max-
Planck-Institut für Biophysik, Abteilung
Molekulare Membranbiologie, Marie-Curie-
Strasse 15, D-60439 Frankfurt am Main,
Germany

Correspondence e-mail:
msweiss@embl-hamburg.de

The structure of the 115 amino-acid residue protein DsvC was determined based on the anomalous scattering provided by the five S atoms present in the structure. By collecting the diffraction data at a wavelength of 1.9 Å, the anomalous signal provided by the S atoms was enhanced. However, significant radiation damage occurred during the course of the experiment, which led to differences between different parts of the data set. Only by dividing the total data set into five data sets was it possible to obtain phases; these could then be successfully extended to allow structure determination by the automated model-building program *ARP/wARP*. A computational correction for the radiation damage was found to significantly improve the success rate in determining the heavy-atom substructure and to improve phasing and refinement statistics.

1. Introduction

Next to crystallization, the most crucial step of an X-ray structure determination is the determination of the phases of the X-ray reflections. The standard phasing procedures are based either on isomorphous replacement (MIR or SIR) and/or anomalous scattering (MAD or SAD). Owing to the advent of strong and tunable synchrotron sources, fast detectors, cryo-crystallographic methods and the simple substitution of methionine by selenomethionine in proteins using recombinant DNA technology, the MAD or SAD methods have become dominant in the last decade. More recently, further developments with respect to increased brilliance of the X-ray beam, increased detector sensitivity and faster data collection have led to a significant improvement in the data quality obtainable. Together with further developments in the field of crystallographic computing (for instance, the dual-space recycling substructure search and maximum-likelihood phasing) this now allows the utilization of very weak anomalous signals and the calculation of high-quality phases from them (Dauter *et al.*, 1999; Ravelli & McSweeney, 2000; Dauter & Dauter, 2001; Weiss, Sicker & Hilgenfeld, 2001; Dauter, 2002; Evans *et al.*, 2003).

In this study, we describe a successful example of phase determination on the basis of the weak anomalous signal of the S atoms that are intrinsically present in the structure. Hendrickson & Teeter (1981) pioneered this approach more than 20 years ago, but it was only recently rediscovered for broader applications by Dauter *et al.* (1999). In order to increase the otherwise very small anomalous signal provided by the S atoms that are present in the structure, the use of softer X-rays for data collection has been suggested (Weiss, Sicker & Hilgenfeld, 2001). The feasibility of collecting good diffraction data at longer wavelengths (up to λ = 2.6 Å) has

been demonstrated by us and by others (Weiss, Sicker, Djinovic Carugo *et al.*, 2001; Olczak *et al.*, 2003; Mueller-Dieckmann *et al.*, 2004) and structures of both model proteins as well as real examples have successfully been determined based on longer wavelength diffraction data (Liu *et al.*, 2000; Weiss, Sicker & Hilgenfeld, 2001; Brown *et al.*, 2002; Li *et al.*, 2002; Micossi *et al.*, 2002; Ramagopal *et al.*, 2003). In spite of the increase achieved, the signal is still small and high redundancy is typically needed to provide a sufficient signal-to-noise ratio both for substructure determination and phase calculation (Dauter & Adamiak, 2001; Weiss, 2001). An inherent problem associated with collecting data to high redundancy is the advent of radiation damage. However, as has been shown recently, even this can be turned into an advantage when the specific changes caused by radiation are exploited in an isomorphous replacement type of experiment for phase determination (Ravelli *et al.*, 2003). In addition, a computational correction ('zero-dose extrapolation') of the effect of radiation damage on the intensities of individual unique reflections can be performed based on multiple observations of the reflections at different doses (Diederichs *et al.*, 2003).

Our model system is the γ-subunit of the dissimilatory sulfite reductase (DsvC) from *Archaeoglobus fulgidus* (Dahl *et al.*, 1993). Dissimilatory sulfite reductase is the key enzyme in sulfate respiration, which reversibly catalyzes the reduction from sulfite to sulfide. It is organized as either a $(\alpha\beta\gamma)_2$ heterohexamer or, more frequently, as a $(\alpha\beta)_2$ heterotetramer since the γ-subunit is generally only loosely bound to the remaining part of the enzyme (Pierik *et al.*, 1992; Steuber *et al.*, 1995). The α- and β-subunits are both about 50 kDa in molecular weight and each harbours two [Fe$_4$S$_4$] clusters. In addition, the α-subunit contains one sirohaem. The γ-subunit has a molecular weight of about 12 kDa and contains as its most characteristic feature four cysteine residues in two disulfide bridges. These disulfide bridges are believed to exist even inside the cell. One of them is redox-active, which suggests that DsvC may serve as an electron acceptor for sulfite reductase and adenylylsulfate reductase (Mander, 2003). A homologous structure of DsvC was solved by NMR methods (Cort *et al.*, 2001), but structure determination by molecular replacement using the NMR model was unsuccessful.

The present report is focused on the structure determination of DsvC by a combination of long-wavelength sulfur phasing and radiation-damage-induced phasing. Our results suggest that an additional parameter in a diffraction data set, either a time or a dose parameter, can contribute significantly to the success of phase determination and to the quality of the phases obtained.

2. Materials and methods

2.1. Cloning, purification and crystallization

The *dsvC* gene was amplified from the chromosomal gene of *A. fulgidus* by PCR techniques, ligated into the pET24b(+)

Table 1
Data-collection and processing statistics for the 1.1 Å resolution data set *A*.

Values in parentheses are for the outer shell.

No. of images	158
$\Delta\varphi$ per image (°)	1.0
Data-collection time (min)	28
Crystal-to-detector distance (mm)	100
Resolution limits (Å)	99.0–1.12 (1.15–1.12)
Unit-cell parameters	
a (Å)	39.93
b (Å)	46.53
c (Å)	57.15
Unit-cell volume (Å^3)	106181
Mosaicity (°)	0.35
Total No. reflections	205663
Unique reflections	40862
Redundancy	5.0
Completeness (%)	93.1 (73.5)
$I/\sigma(I)$	14.8 (4.7)
R_{merge} (%)	9.3 (16.0)
$R_{r.i.m.}/R_{meas}$ (%)	n.d.
$R_{p.i.m.}$ (%)	n.d.
R_{anom} (%)	—
Overall B factor (Å^2)	8.2
Optical resolution (Å)	1.02

vector and expressed in the BL21(DE3) strain (Mander, 2003). The protein was purified in two steps by heat denaturation of the *Escherichia coli* enzymes and Q-Sepharose ion-exchange chromatography. DsvC crystals could be grown in the presence of 2-methyl-2,4-pentanediol (MPD) as a precipitant. The best crystals were obtained by mixing 2 µl protein solution with 2 µl reservoir solution consisting of 58%(v/v) MPD in 0.1 *M* Tris–HCl buffer pH 9.5. The hanging drop was equilibrated against 1 ml reservoir solution. The crystals grow in the orthorhombic space group $P2_12_12_1$, with unit-cell parameters $a = 40.1$, $b = 48.4$, $c = 57.3$ Å and one molecule per asymmetric unit. For data collection they were frozen in a nitrogen stream at 100 K using the drop solution as cryoprotectant.

2.2. Data collection

2.2.1. The 1.1 Å resolution data set *A*. A data set to a maximum resolution of 1.12 Å was collected at a temperature of 100 K and a wavelength of 0.9393 Å at beamline ID 14-4 (ESRF Grenoble, France) using a Q4R ADSC CCD detector. The relevant data-collection parameters are listed in Table 1.

2.2.2. The λ = 1.9 Å data sets *B*-1 to *B*-5. Five passes with a total rotation range of 1742° of diffraction data were collected at 100 K at the diffraction beamline of the ELETTRA synchrotron (Trieste, Italy) using an X-ray wavelength of 1.9 Å. The detector used was a 165 mm MAR CCD detector at a crystal-to-detector distance of either 40 or 70 mm. Three of the passes (*B*-1, *B*-2 and *B*-4) extended to a maximum resolution of 1.85 Å and the other two (*B*-3 and *B*-5) were low-resolution passes that only extended to about 2.3 Å maximum resolution. Data set *B*-2 was offset with respect to data set *B*-1 by 0.5° (half the rotation range per image), whereas for data sets *B*-4 and *B*-5 slightly different orientations of the crystal

Table 2
Data-collection and processing statistics for the $\lambda = 1.9$ Å data sets.

Values in parentheses are for the outer shell.

Data set	B-1	B-2	B-3	B-4	B-5	B-Mini	B-All	B-All$_{XDS}$	B-0-Dose
No. images	360	360	181	360	151	90	1412	1412	1412
$\Delta\varphi$ per image (°)	1.0	1.0	2.0	1.0	2.0	1.0	—	—	—
Data-collection time (h)	2.0	1.9	0.7	1.9	0.5	0.5	7.5	7.5	7.5
Crystal-to-detector distance (mm)	40	40	70	40	70	40	—	—	—
Resolution limits (Å)	99.0–1.85	99.0–1.85	99.0–2.30	99.0–1.85	99.0–2.30	99.0–1.85	99.0–1.85	50.0–1.85	50.0–1.85
Unit-cell parameters									
a (Å)	40.32	40.39	40.42	40.46	40.48	40.28	40.38	40.38	40.38
b (Å)	47.82	47.87	47.88	47.91	47.90	47.79	47.86	47.86	47.86
c (Å)	57.80	57.89	57.93	57.96	57.95	57.74	57.88	57.88	57.88
Unit-cell volume (Å^3)	111444	111929	112112	112352	112365	111148	111858	111858	111858
Mosaicity (°)	0.68	0.67	0.66	0.66	0.67	0.74	0.69	0.25†	0.25†
Total No. of reflections	98545	100547	61761	93878	46112	24488	399272	492887	492867
Unique reflections	9446	9587	5253	9873	5164	9021	9925	9896	9890
Redundancy	10.43	10.49	11.76	9.51	8.93	2.71	40.23	49.81	49.80
Completeness (%)	95.0 (95.0)	95.9 (93.6)	98.5 (97.5)	98.1 (97.3)	97.0 (95.0)	91.6 (91.6)	98.9 (97.2)	98.6 (96.7)	98.5 (96.7)
$I/\sigma(I)$	37.4 (9.5)	41.3 (9.5)	58.4 (28.9)	33.3 (10.4)	35.2 (15.4)	25.6 (13.3)	47.0 (18.1)	48.5 (25.0)	60.2 (29.9)
R_{merge} (%)	5.3 (10.1)	4.4 (12.8)	3.5 (8.9)	5.4 (17.9)	5.7 (12.3)	3.4 (6.1)	8.0 (17.1)	8.7 (17.0)	6.3 (14.2)
$R_{r.i.m.}/R_{meas}$‡ (%)	5.6 (10.7)	4.6 (13.6)	3.6 (9.4)	5.8 (19.1)	6.0 (13.1)	4.2 (7.7)	8.0 (17.4)	8.8 (17.3)	6.4 (14.4)
$R_{p.i.m.}$§ (%)	1.7 (3.5)	1.4 (4.4)	1.1 (2.8)	1.9 (6.6)	2.0 (4.3)	2.3 (4.6)	1.2 (3.4)	1.8 (4.9)¶	1.5 (4.3)¶
R_{anom} (%)	1.9 (3.3)	1.6 (4.2)	1.2 (2.7)	2.1 (5.6)	2.3 (4.2)	2.7 (4.8)	1.3 (3.1)	n.d.	n.d.
Overall B factor (Å^2)	22.3	26.1	35.9	29.0	40.6	22.6	23.5	25.6	26.2
Optical resolution (Å)	1.44	1.49	1.72	1.52	1.76	1.45	1.49	1.47	1.48

† REFLECTING_RANGE_E.S.D. as defined in *XDS*. ‡ For definitions, see Diederichs & Karplus (1997) and Weiss (2001). § For definition, see Weiss (2001). ¶ R_{mrgd-F} as defined in *XDS* and in Diederichs & Karplus (1997).

were achieved using the arcs of the goniometer head. The total data-collection time was 7.5 h including the time required for the changes in the goniometer setting between the different data sets. The total time the crystal was exposed to X-rays amounted to approximately 4.2 h, which very roughly corresponds to a dose of about 2×10^6 Gy based on the equation given in O'Neill *et al.* (2002), a flux of about 10^{12} photons s^{-1} in a 1.2×2 mm beam (see http://www.elettra.trieste.it/experiments/beamlines/xrd1) and a crystal size of about $200 \times 200 \times 200$ μm.

2.2.3. The Cu Kα data set C. A data set to 1.82 Å resolution was collected in-house at 100 K using a Rigaku rotating copper-anode generator ($\lambda = 1.54$ Å) and a Rigaku R-AXIS imaging-plate detector.

2.3. Data processing

2.3.1. The 1.1 Å resolution data set A. The data were indexed, integrated and scaled using the programs *DENZO* and *SCALEPACK* from the *HKL* package (Otwinowski & Minor, 1997). The relevant scaling and merging statistics are given in Table 1.

2.3.2. The $\lambda = 1.9$ Å data set(s) B-1 to B-5, B-Mini and B-All. The data were indexed and integrated with *DENZO* (Otwinowski & Minor, 1997) and scaled using *SCALA* (Evans, 1993; Collaborative Computational Project, Number 4, 1994). Firstly, the discontinuities between images were corrected for by simple BATCH scaling. Then, the ABSORPTION scaling protocol in *SCALA* was employed to correct for the expected increase in absorption at the rather long wavelength of data collection. The intensities of the scaled and merged data sets were then converted to structure-factor amplitudes using the

Table 3
Data-collection and processing statistics for the Cu Kα data set C.

Values in parentheses are for the outer shell.

No. images	210
Data-collection time (min)	1800
Crystal-to-detector distance (mm)	120
Resolution limits (Å)	99.0–1.82 (1.90–1.82)
Unit-cell parameters	
a (Å)	40.1
b (Å)	48.4
c (Å)	57.3
Unit-cell volume (Å^3)	111183
Mosaicity (°)	0.35
Total No. of reflections	79555
Unique reflections	10835
Redundancy	7.3
Completeness (%)	99.8
$I/\sigma(I)$	25.0 (6.5)
R_{merge} (%)	6.1 (21.4)
$R_{r.i.m.}/R_{meas}$ (%)	n.d.
$R_{p.i.m.}$ (%)	n.d.
R_{anom} (%)	n.d.
Overall B factor (Å^2)	25.0
Optical resolution (Å)	1.47

TRUNCATE procedure (French & Wilson, 1978). The five data sets were processed, scaled and merged separately (B-1 to B-5) as well as together to yield a single data set (B-All). In addition, the data set B-Mini was created just consisting of the first 90 images of data set B-1 in order to assess and possibly reduce the influence of radiation damage. The relevant scaling and merging statistics are listed in Table 2.

2.3.3. The Cu Kα data set C. Similar to data set A, the raw diffraction images were indexed, integrated and scaled using the programs *DENZO* and *SCALEPACK* from the *HKL* package (Otwinowski & Minor, 1997). The relevant scaling and merging statistics are given in Table 3.

Table 4
Survey of the relevant refinement results for the various data sets.

Data set	Resolution limits (Å)	R (%)	R_{free} (%)	Solvent molecules	Average B factor of protein atoms (Å²)	Alternate conformations†	Reduced S—S‡ (%)
C	40.0–1.82	18.3	22.8	105	24.0	3	10
B-Mini	40.0–1.85	18.9	25.4	85	22.8	3	27
B-1	40.0–1.85	18.6	24.3	103	23.0	3	31
B-2	40.0–1.85	17.9	21.8	76	27.6	3	37
B-3	40.0–2.30	18.1	22.9	44	29.8	0	n.d.
B-4	40.0–1.85	18.6	21.5	56	31.2	3	45
B-5	40.0–2.30	18.8	23.1	23	37.2	0	n.d.
B-All	40.0–1.85	18.1	22.2	84	24.1	3	28
B-All$_{XDS}$	40.0–1.85	17.6	21.0	95	21.2	3	30
B-0-Dose	40.0–1.85	18.5	21.4	95	20.3	3	27
A	30.0–1.12	10.8	13.7	149	12.4	17	48

† Except Cys77.   ‡ This applies to the disulfide bridge Cys77–Cys85.

2.3.4. The merged data set A/B-All. Owing to the apparent non-isomorphism between data sets A and B-1 to B-5, an attempt to scale and merge these data sets together by conventional means resulted in an enormous amount of rejected observations. Therefore, this approach was considered to be unsuitable. However, in order to utilize the high-resolution data of data set A, a rather unconventionally merged data set was prepared by simply averaging the structure-factor amplitudes of data sets A and B-All. Firstly, the two data sets were scaled to each other using the program *SCALEIT* (Collaborative Computational Project, Number 4, 1994) and a single scale and isotropic temperature factor. The scaled structure-factor amplitudes were then averaged with equal weights applied to the two data sets. The unit-cell parameters of data set A were kept for the merged data set. The merging R factor based on structure-factor amplitudes between these two data sets was 17.3%. In contrast to data set A alone, this new and averaged data set A/B-All proved very valuable for the solvent-flattening and phase-extension step.

2.3.5. The data sets B-All$_{XDS}$ and B-0-Dose. In order to assess the influence of radiation damage on the structure determination, the method described in Diederichs *et al.* (2003) was used to extrapolate the intensities of data sets B-1 to B-5 to the beginning of the data collection. For this purpose, the five passes were integrated using the program *XDS* (version from December 2002; Kabsch, 1993) and scaled using *XSCALE*. Just scaling and merging the five data sets together yielded the data set B-All$_{XDS}$ (Table 2). Scaled but unmerged intensities were then used for the zero-dose extrapolation. A second scaling step using the program *XSCALE* was then carried out on the corrected intensities. This procedure produced the data set B-0-Dose, the statistics of which are also given in Table 2.

2.4. Substructure determination

The sulfur substructure was determined using the program *SHELXD* (Schneider & Sheldrick, 2002) and the merged data set B-All with all data to 1.85 Å resolution included. Since no significant drop in the peak heights was observed after the first five expected sites, the number of sites to be searched for was increased to ten. In order to compare the success rates of substructure determination, data sets B-1 to B-5 were also tried separately. Furthermore, data sets B-All$_{XDS}$ and B-0-Dose were also tried under the same conditions.

2.5. Phase determination

The occupancies of the top ten sites found using *SHELXD* were further refined using the program *MLPHARE* (Otwinowski, 1991; Collaborative Computational Project, Number 4, 1994). Because of the significant differences between the data sets (Table 4), various combinations of data sets were tried. Since the phase-calculation procedure in *MLPHARE* is an isomorphous replacement type of approach, the role of a data set can be 'native' and/or 'derivative' depending on which parameters are refined and which are kept constant. In single-wavelength anomalous scattering, the same data set is used both as native and as derivative.

2.6. Solvent flattening and phase extension

The phases obtained using the program *MLPHARE* were improved by solvent flattening using the program *DM* (Cowtan, 1994; Collaborative Computational Project, Number 4, 1994). A solvent content of 40% was assumed based on the unit-cell parameters and molecular weight of the protein. If only B data sets were used, the maximum resolution was 1.85 Å; if either data sets A or A/B-All were introduced as the native data sets, the phases were automatically extended to a maximum resolution of 1.20 Å. In order to identify the correct enantiomorph of the anomalous substructure, the phase-determination step and the solvent-flattening and phase-extension step were carried out for both enantiomorphs under otherwise identical conditions.

2.7. Automated model building

Both improved and extended phase sets were submitted to the automated model-building and refinement software *ARP*/*wARP* (Perrakis *et al.*, 1999). Since the phases were essentially single-wavelength anomalous scattering-phases and since these phases are notoriously difficult to work with in *ARP*/*wARP* (Victor Lamzin, personal communication), the slow protocol was chosen. The density was firstly further improved by 20 cycles of free-atom density modification and then automated model building was iterated with ten cycles of model refinement and free atoms updating.

2.8. Manual model building

After visual inspection of the electron-density map after the first *ARP*/*wARP* building cycle, it was realised that the

corresponding 1.2 Å electron density was essentially perfect. Nearly all of the 115 amino acids of DsvC were visible. Therefore, the gaps between the four automatically built chains were filled in by hand. At this stage, the model consisted of 77 amino-acid residues plus 744 free atoms inserted into the electron density. The easiest way to build the gaps between the amino acids was to identify the correctly placed free atoms and rename and renumber them according to the amino acid the position of which they occupied.

2.9. Refinement

The structure of DsvC was refined against several of the data sets discussed. A survey of the relevant refinement statistics is given in Table 4.

2.9.1. Refinement against data set _A_. The model was refined against the high-resolution data set _A_ using the programs _CNS_ (Brünger _et al._, 1998) and _O_ (Jones _et al._, 1991) to R and R_{free} values of 18.0 and 19.8%, respectively, using standard scripts. 135 H_2O molecules were picked automatically in two rounds from the difference map; peaks were discarded again when their thermal displacement parameter B refined to greater than 50 Å^2. Incorporating alternate conformations for 15 residues finally reduced R to 17.1% and R_{free} to 18.9% in the resolution range 30.0–1.12 Å. Because of the high resolution of the data, refinement was continued using the program _SHELXL_ (Sheldrick & Schneider, 1996). Refinement of anisotropic thermal displacement parameters, definition of alternative conformations for a total of 18 residues (for ten of them all atoms exhibit dual conformations; for the remaining eight only the side-chain atoms do), the incorporation of H atoms and additional water molecules, the latter being partly assigned half-occupied, lowered R to 10.6% and R_{free} to 13.5% for reflections with $F_o > 4\sigma(F_o)$ and R to 10.8% and R_{free} to 13.7% ($wR2 = 27.5\%$) for all 39 635 reflections. The final model comprises 114 amino-acid residues and 149 H_2O molecules, 55 of which were assigned to half-occupancy.

The quality of the model was checked using the programs _PROCHECK_ (Laskowski _et al._, 1993) and _WHATCHECK_ (Hooft _et al._, 1996). All non-Gly and non-Pro residues are within the allowed (86.6% in the most favoured) regions of the Ramachandran plot. The root-mean-square deviation of the bond lengths from ideal values is 0.016 Å and that of the bond angles is 2.1°. The model occupies a total of 70% of the unit-cell volume.

2.9.2. Refinement against data sets _B_-All, _B_-All$_{XDS}$, _B_-Mini, _B_-1 to _B_-5, _B_-0-Dose. Full models of DsvC were refined against the _B_ data sets using the program _CNS_ (Brünger _et al._, 1998) starting from the high-resolution protein model without the alternate conformations. The temperature factors were set to their corresponding isotropic values and refined isotropically. Water molecules were automatically picked as above. For data sets _B_-1, _B_-2 and _B_-4 alternate conformations were fitted and refined for all atoms of residues 41–43 and for Cys77 S^{γ}. To allow for the 'partial presence' of the disulfide bridge between Cys77 and Cys85, the constraints file used in

Table 5
R factors (%) between the scaled B data sets.

All R factors are based on structure-factor amplitudes with the maximum resolution limited to 2.3 Å.

Data set	_B_-1	_B_-2	_B_-3	_B_-4	_B_-5	_B_-All	_B_-All$_{XDS}$	_B_-0-Dose
_B_-Mini	2.9	8.1	11.0	13.0	14.8	8.2	9.9	5.7
_B_-1	—	5.7	8.7	10.7	12.5	5.7	7.9	4.9
_B_-2		—	3.6	5.3	7.3	1.4	4.0	6.8
_B_-3			—	3.7	4.4	3.0	4.2	9.5
_B_-4				—	3.7	4.8	5.5	11.4
_B_-5					—	6.4	6.9	12.5
_B_-All						—	4.2	7.8
_B_-All$_{XDS}$							—	7.0

Table 6
R factors (%) between the scaled B data sets and the data sets A and A/B-All.

All R factors are based on structure-factor amplitudes with the maximum resolution limited to 2.3 Å.

Data set	_A_	_B_-Mini	_B_-1	_B_-2	_B_-3	_B_-4	_B_-5	_B_-All	_B_-All$_{XDS}$	_B_-0-Dose
_A_	—	29.1	29.0	29.4	30.4	30.4	31.2	29.4	29.4	29.4
_A_/_B_-All	14.5	15.3	14.3	13.3	13.6	14.7	15.1	12.6	13.2	14.5

CNS had to be modified appropriately. A second position for Cys85 S^{γ} had to be kept for technical reasons but with the occupancy set to 0.01. The relative occupancy of the two positions of Cys77 S^{γ} was refined. Because of the strong correlation between the temperature factor and the occupancy, both parameters were refined alternatively and the occupancy checked for stability.

2.9.3. Refinement against data sets _C_. The refinement against data set _C_ was carried out as for the _B_ data sets. Alternate conformations were fitted and refined for all side-chain atoms of Glu31, Asp59 and Ile90 and in addition tentatively for the single atom Cys77 S^{γ}.

3. Results and discussion

3.1. Data collection and processing

At first glance, the data collection and processing proceeded with no apparent problems (Tables 1, 2 and 3). For the long-wavelength data sets _B_-1 to _B_-5, the strategy had been to collect a highly redundant data set, since it was not clear at the outset what redundancy was needed to solve the structure. However, even though the estimated dose was significantly below the Henderson limit of 2×10^7 Gy, the radiation damage occurring during the course of the experiment led to significant differences between the data sets _B_-1 to _B_-5 (Table 5). Since data set _B_-3 was the first low-resolution data set collected in which no low-resolution reflections were overloaded, data sets _B_-1 and _B_-2 turned out to be essentially useless by themselves because of the reduced completeness at low resolution. In retrospect, it would have been better to start with the collection of the low-resolution data set _B_-3 and then proceed with the high-resolution data sets.

The data set _B_-All, which resulted from the merging of data sets _B_-1 to _B_-5, exhibits significantly higher merging statistics

Table 7
Success rate for substructure determination using the program *SHELXD* (Schneider & Sheldrick, 2002).

The number of trials was 100 in all cases.

Resolution	2.30 Å	1.85 Å
B-1	0	0
B-2	2	1
B-3	0	—
B-4	0	0
B-5	0	—
B-All	46	78
B-All$_{XDS}$	45	67
B-0-Dose	67	88

than the single data sets *B*-1 to *B*-5. The redundancy-independent merging R factor $R_{r.i.m.}$ (or R_{meas}) of data set *B*-All is about 50% higher than that for data sets *B*-1 to *B*-5, although theoretically they should be identical if the major source of error is counting statistics. This is clearly a result of radiation damage. Nevertheless, the $R_{p.i.m.}$ of data set *B*-All is lower than the $R_{p.i.m.}$ of data sets *B*-1 to *B*-5, indicating that the merging did decrease the noise in the averaged data at least to some extent. The occurrence of radiation damage is also evident from the R factors between data sets presented in Table 5. The further the data sets are apart along the course of the experiment, the higher the R factor between them.

Significant non-isomorphism exists between the *A* and *B* data sets as can be inferred from the R factor of almost 30% between data set *A* and all *B* data sets (Table 6). Also, the unit-cell parameters are different even though both sets were determined from native DsvC crystals. It should be mentioned here that the morphology of the crystal used for collecting data set *A* appeared to be clearly more compact and not as long as the crystal used for collecting the *B* data sets. Whether this is a consequence of the oxidation state of the enzyme is not clear. It may just as well be the case that the shock-cooling to 100 K led to this apparent non-isomorphism.

The radiation damage correction by zero-dose extrapolation leads to a significantly improved internal agreement of reflections as demonstrated by the overall $R_{r.i.m.}$ (or R_{meas}) of 6.4% for data set *B*-0-Dose compared with the 8.8% for the uncorrected data set *B*-All$_{XDS}$ (Table 2). The correction also translates into a much improved signal-to-noise ratio (Table 2). In terms of degrees of freedom, the effect of fitting two parameters (zero-dose intensity and decay factor) to n observations of a unique reflection should correspond to that of fitting a single parameter (intensity) to only $(n − 1)$ observations. As the redundancy is very high (around 40), we conclude that the observed reduction in $R_{r.i.m.}$ (or R_{meas}) is highly significant. This demonstrates clearly that the systematic error brought about by the occurrence of radiation damage can be dealt with satisfactorily by zero-dose extrapolation. In the case presented here, the high redundancy of the data certainly prevents overfitting of the data, with the additional parameter introduced by the linear function used to fit the intensity *versus* dose relationship of each unique reflection. The significance of the zero-dose extrapolation can

also be inferred from the fact that the extrapolated data set *B*-0-Dose exhibits the lowest R factors to data sets *B*-Mini and *B*-1, which were the first data sets collected from that crystal (Table 5).

The original aim of collecting data set *C* had been to detect the anomalous signals of the S atoms at the wavelength of the Cu $K\alpha$ radiation. Unfortunately, it turned out that the acquired redundancy was insufficient to achieve this goal. The role of data set *C* now is that of a reference measurement exhibiting very low radiation damage and essentially no disulfide-bond reduction.

3.2. Substructure determination

Despite the significant radiation damage present in the merged data set *B*-All, the substructure determination using this data set and the computer program *SHELXD* (Schneider & Sheldrick, 2002) proceeded smoothly and the success rate for the substructure determination was surprisingly high: 78 out of 100 trials (Table 7) yielded significantly higher correlation coefficients than the remaining 22. Nevertheless, with values of 0.25–0.27 and 0.13–0.16 for all data and for the weak data, respectively, the correlation coefficients obtained are lower than expected for a correct solution. This effect is probably because of radiation damage, which led to significant differences in the data sets *B*-1 to *B*-5 which are the constituents of *B*-All (Tables 5 and 6).

If data sets *B*-1 to *B*-5 were tried separately (Table 7), the success rate essentially drops to zero. Only in the case of *B*-2 was a single solution found if the maximum resolution was used; two solutions were found when the resolution was truncated to 2.3 Å. This effect has been noted previously (Weiss, 2001; Dauter & Adamiak, 2001) and is likely to be a consequence of the decreased accuracy of the data as a result of the lower redundancy of the data sets. However, since we have not ensured that the same subset of data was used in the different data sets, we cannot exclude the possibility that the presence or absence of a few strong normalized structure factors also has some influence on the observed behaviour. The exclusion of overloaded reflections resulted in a lower completeness at low resolution of data sets *B*-1, *B*-2 and *B*-4. However, this is not the case for data sets *B*-3 and *B*-5, which are essentially complete at low resolution. Here, the reduction in the success rate is almost certainly a result of the low data redundancy.

In order to evaluate the influence of the zero-dose extrapolation (Diederichs *et al.*, 2003), *SHELXD* was run again on data sets *B*-All$_{XDS}$ and *B*-0-Dose. The *XDS/XSCALE*-processed data set *B*-All$_{XDS}$ yields a slightly lower success rate than the *DENZO/SCALA*-processed data set *B*-All (Table 7). The success rate with data set *B*-0-Dose, however, is significantly higher than that of any of the uncorrected data sets. This clearly demonstrates the potential of the method in cases when radiation damage degrades the signal and it also calls for the inclusion of the zero-dose extrapolation method in all the available scaling programs.

The peak heights of the sites found using *SHELXD* are given in Table 8. Since no significant drop was observed after

Table 8
Substructure determination using the program *SHELXD* (Schneider & Sheldrick, 2002) and substructure refinement using *MLPHARE* (Collaborative Computational Project, Number 4, 1994).

For the substructure determination the data set *B*-All was used; for the refinement data sets *B*-1 to *B*-5 were kept separate.

					Anomalous occupancy (*MLPHARE*)					
Site	*x*	*y*	*z*	Occupancy (*SHELXD*)	*B*-1	*B*-2	*B*-3	*B*-4	*B*-5	Atom
1	0.296	0.584	0.052	1.00	0.62	0.57	0.58	0.48	0.56	Cys114 S$^\gamma$
2	0.341	0.586	0.990	0.97	0.68	0.66	0.63	0.62	0.60	Cys103 S$^\gamma$
3	0.368	0.364	0.171	0.72	0.50	0.47	0.48	0.44	0.50	Cys85 S$^\gamma$
4	0.079	0.437	0.068	0.60	0.43	0.41	0.42	0.40	0.36	Met72 S$^\delta$
5	0.372	0.343	0.131	0.44	0.34	0.36	0.36	0.27	0.25	Cys77 S$^\gamma$
6	0.345	0.369	0.097	0.28	0.22	0.20	0.21	0.24	0.21	Next to Cys77
7	0.271	0.606	0.127	0.27	0.07	0.08	0.03	0.07	0.08	Near Val70
8	0.406	0.416	0.991	0.23	0.07	0.05	0.01	0.03	0.07	Near Ile53
9	0.366	0.579	0.033	0.23	0.08	0.04	0.05	0.00	0.01	Next to Cys103
10	0.278	0.577	0.009	0.17	0.05	0.05	0.04	0.05	0.04	Next to Cys114

Table 9
Phase determination in *MLPHARE* and *DM*.

	MLPHARE							*DM*						
Exp.	F_{native}†	$F_{derivative}$†	ISO‡	ANO‡	d_{min}§	FOM¶	CC††	F_{native} †	d_{min}§	FOM¶	CC††	d_{min}§	FOM¶	CC††
1	*B*-All	*B*-All	No	Yes	1.85	0.08	0.31	*B*-All	1.85	0.64	0.44	—	—	—
2	*B*-1	*B*-1, *B*-2, *B*-3, *B*-4, *B*-5	No	Yes	1.85	0.26	0.35	*B*-All	1.85	0.63	0.52	—	—	—
3	*B*-1	*B*-1, *B*-2, *B*-3, *B*-4, *B*-5	Yes	Yes	1.85	0.58	0.14	*B*-All	1.85	0.71	0.09	—	—	—
4	*B*-All	*B*-1, *B*-2, *B*-3, *B*-4, *B*-5	No	Yes	1.85	0.25	0.38	*B*-All	1.85	0.64	0.57	—	—	—
5	*B*-All	*B*-1, *B*-2, *B*-3, *B*-4, *B*-5	Yes	Yes	1.85	0.53	0.15	*B*-All	1.85	0.68	0.14	—	—	—
6	*B*-All	*B*-All	No	Yes	1.85	0.08	0.31	*A*	1.85	0.64	0.27	1.20	0.49	0.25
7	*B*-All	*B*-1, *B*-2, *B*-3, *B*-4, *B*-5	No	Yes	1.85	0.25	0.38	*A*	1.85	0.64	0.42	1.20	0.43	0.39
8	*B*-All	*B*-All	No	Yes	1.85	0.08	0.31	*A*/*B*-All	1.85	0.70	0.37	1.20	0.45	0.36
9	*B*-All	*B*-1, *B*-2, *B*-3, *B*-4, *B*-5	No	Yes	1.85	0.25	0.38	*A*/*B*-All	1.85	0.74	0.48	1.20	0.48	0.46
10	*A*	*B*-All	No	Yes	1.85	0.05	0.23	*A*	1.85	0.63	0.16	1.20	0.49	0.14
11	*A*	*B*-All	Yes	Yes	1.85	0.16	0.10	*A*	1.85	0.69	0.08	1.20	0.53	0.07
12	*A*	*B*-1, *B*-2, *B*-3, *B*-4, *B*-5	No	Yes	1.85	0.16	0.32	*A*	1.85	0.66	0.33	1.20	0.50	0.30
13	*A*	*B*-1, *B*-2, *B*-3, *B*-4, *B*-5	Yes	Yes	1.85	0.64	0.05	*A*	1.85	0.73	0.05	1.20	0.51	0.05
14	*A*/*B*-All	*B*-All	No	Yes	1.85	0.05	0.23	*A*/*B*-All	1.85	0.68	0.31	1.20	0.38	0.29
15	*A*/*B*-All	*B*-All	Yes	Yes	1.85	0.10	0.16	*A*/*B*-All	1.85	0.73	0.12	1.20	0.55	0.10
16	*A*/*B*-All	*B*-1, *B*-2, *B*-3, *B*-4, *B*-5	No	Yes	1.85	0.15	0.29	*A*/*B*-All	1.85	0.74	0.42	1.20	0.50	0.40
17	*A*/*B*-All	*B*-1, *B*-2, *B*-3, *B*-4, *B*-5	Yes	Yes	1.85	0.55	0.06	*A*/*B*-All	1.85	0.73	0.06	1.20	0.41	0.05
18	*B*-All$_{XDS}$	*B*-All$_{XDS}$	No	Yes	1.85	0.08	0.32	*B*-All$_{XDS}$	1.85	0.59	0.34	—	—	—
19	*B*-All$_{XDS}$	*B*-1, *B*-2, *B*-3, *B*-4, *B*-5	No	Yes	1.85	0.24	0.37	*B*-All$_{XDS}$	1.85	0.66	0.56	—	—	—
20	*B*-0-Dose	*B*-0-Dose	No	Yes	1.85	0.10	0.33	*B*-0-Dose	1.85	0.54	0.35	—	—	—
21	*B*-0-Dose	*B*-1, *B*-2, *B*-3, *B*-4, *B*-5	No	Yes	1.85	0.26	0.35	*B*-0-Dose	1.85	0.63	0.52	—	—	—
22	*B*-0-Dose	*B*-1, *B*-2, *B*-3, *B*-4, *B*-5	Yes	Yes	1.85	0.57	0.07	*B*-0-Dose	1.85	0.61	0.07	—	—	—

† F_{native} and $F_{derivative}$ are distinguished according to their role and definition in *MLPHARE* and/or *DM*, respectively.   ‡ ISO and ANO describe the use of dispersive and/or anomalous differences for phase determination in *MLPHARE*.   § d_{min} denotes the maximum resolution used for phase determination and solvent flattening.   ¶ FOM is the figure of merit for phase determination in *MLPHARE* and the combined figure of merit after solvent flattening in *DM*, respectively.   †† CC are map correlation coefficients between experimentally phased (or solvent-flattened) maps and the final $(2F_o - F_c)$ maps. For the cases in which $d_{min} = 1.85$ Å the model phases used were taken from a refinement of DsvC against data set *B*-All; for the cases in which $d_{min} = 1.2$ Å the model phases were those from a refinement against data set *A*.

site number five, all the top ten sites were used for further refinement and phase determination. In retrospect, this observation is also a result of radiation damage. The disulfide bridge Cys77–Cys85 became reduced during the course of the experiment and a second position appeared for one of the S atoms (Cys77 S$^\gamma$). In the same table, the refined anomalous occupancy values for data sets *B*-1 to *B*-5 are also given. A clear trend towards a decrease of the signal can be observed for sites 1, 2, 4 and 5, although one has to be somewhat careful

with the interpretation since data sets *B*-3 and *B*-5 extend to a maximum resolution of only 2.3 Å, whereas *B*-1, *B*-2 and *B*-4 extend to 1.85 Å.

3.3. Phase determination

After an unsuccessful first attempt to calculate phases from the data set *B*-All alone in a single-wavelength anomalous scattering approach, various combinations of native and

derivative data sets were tried (Table 9). It was found to be critical that the five data sets *B*-1 to *B*-5 were treated as separate data sets, thus allowing some specific radiation-damage-induced changes to be accommodated. It was also realised that only the anomalous differences within each of the five data sets could be used for refining the occupancies. If the isomorphous differences between the data sets that arose from radiation damage were also used, the figures of merit were overestimated and at the same time the quality of the phases deteriorated. It is not clear to us whether this is a problem of the phasing program *MLPHARE* and whether it may be that with a more sophisticated phase-determination program such as *SHARP* (de La Fortelle & Bricogne, 1997) or *BP*3 (Pannu *et al.*, 2003) better phases would have been obtained. However, as the structure was solved successfully in the end using *MLPHARE*, the need to use another program became obsolete.

The unconventionally merged data set *A/B*-All was introduced as the native data set here despite the significant non-isomorphism (experiments 14–17 in Table 9). This way, it was ensured that the phases calculated would be directly applied to the same structure-factor amplitudes used in the solvent-flattening procedure (see below), which proved to be advantageous over introducing the high-resolution data set *A* only at the solvent-flattening/phase-extension stage below (experiments 6 and 7 in Table 9). Experiments 8 and 9 (Table 9) turned out to be the combination which in the end led to the successful structure solution. The figures of merit after *DM* were 0.792 and 0.778 for the two enantiomorphs, slightly favouring the correct one. In retrospect, however, it would have been sufficient or even better to introduce *A/B*-All at the solvent-flattening stage only (experiment No. 9 in Table 9) as can be inferred from the map correlation coefficients given in Table 9.

It is also interesting to note that the data set *B*-0-Dose did not provide any significant advantage in the phase-determination step over the uncorrected data sets. This is to some extent surprising given that the success rate in substructure determination was higher than for the uncorrected data sets *B*-All and *B*-All$_{XDS}$ and that both applications depend on the accuracy of the anomalous difference. However, in both cases the overall low quality of the phases may mask a positive effect of the zero-dose extrapolation.

3.4. Automated model building

For the correct enantiomorph, *ARP/wARP* built 77 amino-acid residues in four chains in the first building cycle, whereas for the incorrect enantiomorph not a single residue was built. Unfortunately, the structure could not be completed automatically using *ARP/wARP*. The number of residues built could not be increased after the first cycle. In contrast, as the number of cycles increased, the model was successively degraded. According to one of the authors of *ARP/wARP* (V. Lamzin, personal communication), the problem here was that the resolution of the electron-density map was too high and that the program did not have good enough structural

templates for such high resolution. After the production of suitable templates, *ARP/wARP* managed to build essentially the complete structure within three cycles.

3.5. Refinement

When comparing the values for the *R* and the free *R* factor of the refinement runs against the various data sets (Table 4), it appears as if they are independent of the incident X-ray dose. This means that all data sets are of good quality and that each of them describes the corresponding structure equally well. Nevertheless, the various structures do change as a function of the X-ray dose. This can be inferred from the average temperature-factor values of the protein atoms, which increase substantially from about 30 Å^2 in the data sets *B*-Mini and *B*-0-Dose to about 40 Å^2 in data set *B*-4, as well as the number of visible water molecules, which decreases by about 50% during the course of the experiment (see Table 4). Similar observations were also made in previous studies (Ravelli & McSweeney, 2000). This means that concomitant with the absorbed X-ray dose, the disorder in the crystal increases as a consequence of radiation damage.

A three-dimensional comparison of the structures derived from data sets *A*, *B*-1 to *B*-5, *B*-All, *B*-All$_{XDS}$, *B*-0-Dose and *C* indicates no significant differences with respect to the main-chain atoms. The root-mean-square deviations between the C$^\alpha$-atom positions are in the range 0.1–0.3 Å, which is well inside the expected coordinate error, especially when considering that the data were collected from different crystals. When only the structures obtained from refinements against the various *B* data sets are considered, the respective differences are below 0.15 Å. Significant conformational differences are only found for some solvent-exposed side chains: Lys10, Glu45, Trp51 and Arg55 are highly flexible, in particular in data sets measured after longer X-ray exposures. The most striking dose-dependent effect in DsvC is the breakage of the disulfide bond between Cys77 and Cys85, which can be qualitatively followed in the electron-density maps derived from the mentioned data sets (Fig. 1) and in Table 4. Photoreduction of a redox-active disulfide bridge by synchrotron radiation has also been observed recently for tryparedoxins (Alphey *et al.*, 2003). In that case, the concomitant conformational change was even implicated in the function of the protein.

A detailed look at the ratio of the oxidized to the reduced form of the Cys77–Cys85 disulfide bridge in the *B* data sets (Table 4) suggests that either most of the radiation damage occurs at the very beginning of data collection or that the occupancy of the reduced form is about 20% even prior to data collection. Interestingly, a split data set *A* did not reveal a significant difference in the oxidized:reduced ratio dependent upon the dose, suggesting that the oxidation state of the disulfide does actually differ between crystals. Owing to its high resolution, the *A* structure most clearly shows the structural alterations induced by the radiation damage. The S atoms of Cys77 and Cys85 both move, although the extent of the movement and the decrease of occupancy are much

smaller for the latter. The occupancy values for the second conformations of Cys77 S^γ and Cys85 S^γ are 0.48 and 0.20, respectively. The sulfur of Cys77 is turned away from Cys85 by a rotation of 98° around the $C^\alpha - C^\beta$ bond. This structural change induces a shift of the main-chain atoms of the segment between residues 75 and 78 and a rearrangement of the side chain of Leu73, thereby changing the side-chain conformation of Leu57. The S atom of Cys85 moves only slightly. The rotation around the $C^\alpha - C^\beta$ bond amounts to 24° and no further conformational changes are induced in its surroundings.

In contrast to the structures corresponding to the *B* data sets, the structure of the Cu $K\alpha$ data set *C* is compatible with a completely oxidized disulfide bridge prior to X-ray exposure. This supports our hypothesis that the oxidation state may be a function of the crystal as much as it is a function of the X-ray dose.
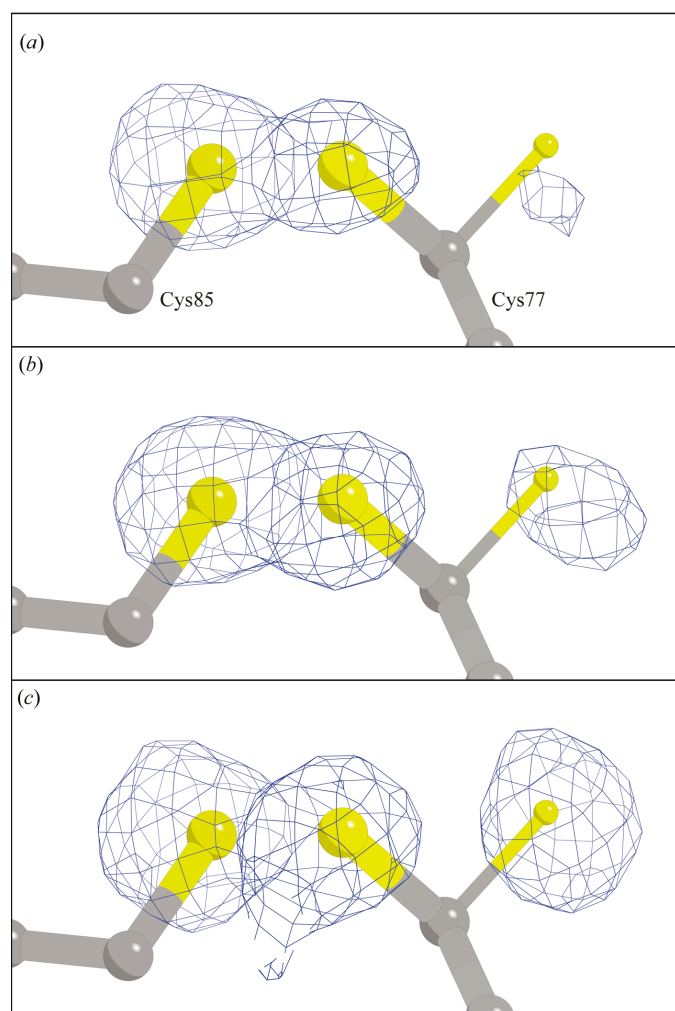

Figure 1
Anomalous difference electron-density maps at 1.85 Å resolution superimposed on the disulfide bridge Cys77—Cys85 in the structures refined against data sets *B*-1 (*a*), *B*-2 (*b*) and *B*-4 (*c*). The progress of radiation damage is clearly visible as an increase in the electron density at the alternative position of Cys77 S^γ. The contour level for all three electron-density maps is 0.022 e $Å^{-3}$. The figure was prepared using the programs *MOLSCRIPT* (Kraulis, 1991) and *BOBSCRIPT* (Esnouf, 1997).

4. Conclusions

The structure of the γ-subunit of the dissimilatory sulfite reductase was determined based on the anomalous signal provided by the five S atoms that are present in the native structure. Diffraction data collection at a wavelength of 1.9 Å helped to enhance the weak signal. The advent of radiation damage during the data-collection stage on the one hand degraded the signal but on the other hand provided extra phasing information. In addition, a high-resolution data set was necessary to obtain phases of sufficient quality for automated model building.

With the advent of radiation damage, the method of zero-dose extrapolation results in a significant increase in the success rate for substructure determination and also yields better refinement statistics. Further improvements of the method, *e.g.* improved models for the influence of radiation damage on single reflections and better integration into existing scaling procedures, may thus be able to offer even greater benefits for substructure determination, phasing and refinement.

For the phase-determination step it proved to be critical to take the radiation damage into account. We would therefore like to suggest that an additional parameter in a diffraction data set, either a time or a dose parameter, should be kept. This may contribute significantly to the success of phase determination and to the quality of the phases obtained.

Since the overall sulfur content of DsvC is approximately that of a protein of average composition, our results suggest that a combination of long-wavelength sulfur phasing and proper treatment of radiation damage has the potential to become a general method in macromolecular crystallography. Needless to say, further studies on larger proteins with even lower sulfur content have to be carried out in order to uncover the limits of the method. It also remains to be studied how the mix of Cys and Met residues will influence the outcome of a long-wavelength sulfur phasing experiment with respect to the occurrence of radiation damage.

References

Alphey, M. S., Gabrielsen, M., Micossi, E., Leonard, G. A., McSweeney, S. M., Ravelli, R. B., Tetaud, E., Fairlamb, A. H., Bond, C. S. & Hunter, W. N. (2003). *J. Biol. Chem.* **278**, 25919–25925.

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.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cort, J. R., Mariappan, S., Kim, C.-Y., Park, M. S., Peat, T. S., Waldo, G. S., Terwilliger, T. C. & Kennedy, M. A. (2001). *Eur. J. Biochem.* **268**, 5842–5850.

Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.

Dahl, C., Kredlich, N. M., Deutzmann, R. & Trüper, H. G. (1993). *J. Gen. Microbiol.* **139**, 1817–1828.

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. (2001). *Structure*, **9**, R21–R26.

Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.

Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.

Diederichs, K., McSweeney, S. & Ravelli, R. B. G. (2003). *Acta Cryst.* D**59**, 903–909.

Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.

Evans, G., Polentarutti, M., Dijnovic Carugo, K. & Bricogne, G. (2003). *Acta Cryst.* D**59**, 1429–1434.

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.

French, G. S. & Wilson, K. S. (1978). *Acta Cryst.* A**34**, 517–525.

Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.

Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. (1996). *Nature (London)*, **381**, 272–272.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Li, S., Finley, J., Liu, Z. J., Qiu, S. H., Chen, H., Luan, C. H., Carson, M., Tsao, J., Johnson, D., Lin, G., Zhao, J., Thomas, W., Nagy, L. A., Sha, B., DeLucas, L. J., Wang, B.-C. & Luo, M. (2002). *J. Biol. Chem.* **277**, 48596–48601.

Liu, Z.-J., Vysotski, E. S., Chen, C.-J., Rose, J. P., Lee, J. & Wang, B.-C. (2000). *Protein Sci.* **9**, 2085–2093.

Mander, G. (2003). PhD thesis. Philipps-Universität Marburg/Lahn, Marburg/Lahn, Germany.

Micossi, E., Hunter, W. N. & Leonard, G. A. (2002). *Acta Cryst.* D**58**, 21–28.

Mueller-Dieckmann, C., Polentarutti, M., Djinovic Carugo, K., Panjikar, S., Tucker, P. A. & Weiss, M. S. (2004). *Acta Cryst.* D**60**, 28–38.

Olczak, A., Cianci, M., Hao, Q., Rizkallah, P. J., Raftery, J. & Helliwell, J. R. (2003). *Acta Cryst.* A**59**, 327–334.

O'Neill, P., Stevens, D. L. & Garman, E. (2002). *J. Synchrotron Rad.* **9**, 329–332.

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.

Pannu, N. S., McCoy, A. J. & Read, R. J. (2003). *Acta Cryst.* D**59**, 1801–1808.

Perrakis, A., Morris, R. J. H. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Pierik, A. J., Duyvis, M. G., van Helvoort, J. M. L. M., Wolbert, R. B. G. & Hagen, W. R. (1992). *Eur. J. Biochem.* **205**, 111–115.

Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 1020–1027.

Ravelli, R. B., Leiros, H. K., Pan, B., Caffrey, M. & McSweeney, S. (2003). *Structure*, **11**, 217–224.

Ravelli, R. B. & McSweeney, S. (2000). *Structure*, **8**, 315–328.

Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.

Sheldrick, G. M. & Schneider, T. R. (1996). *Methods Enzymol.* **277**, 319–343.

Steuber, J., Arendsen, A. F., Hagen, W. R. & Kroneck, P. M. H. (1995). *Eur. J. Biochem.* **3**, 873–879.

Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.

Weiss, M. S., Sicker, T., Djinovic Carugo, K. & Hilgenfeld, R. (2001). *Acta Cryst.* D**57**, 689–695.

Weiss, M. S., Sicker, T. & Hilgenfeld, R. (2001). *Structure*, **9**, 771–777.