In-house sulfur SAD phasing: a case study of the effects of data quality and resolution cutoffs

Ganapathy N. Sarma‡ and
P. Andrew Karplus*

Department of Biochemistry and Biophysics, 2011 ALS, Oregon State University, Corvallis, OR 97331, USA

‡ Present address: Howard Hughes Medical Institute, University of California, San Diego, 432 Leichtag Building, 9500 Gilman Drive, Mail Code 0654, La Jolla, CA 92093, USA.

Correspondence e-mail:
karplusp@science.oregonstate.edu

Single-wavelength anomalous diffraction (SAD) utilizing the weak signal of inherently present S atoms can be successfully used to solve macromolecular structures, although this is mostly performed with data from a synchrotron rather than a laboratory source. Using high redundancy, sufficiently accurate anomalous data may now often be collected in the laboratory using Cu $K\alpha$ X-ray radiation. Systematic analyses of a laboratory-derived data set illuminate the effects of data quality, redundancy and resolution cutoffs on the ability to locate the S atoms and phase the structure of Ptr ToxA, a 13.2 kDa toxin secreted by the fungus *Pyrenophora tritici-repentis*. Three sulfurs contributed to the successful phasing of the structure and were located using the program *SHELXD*. It is observed that data quality improves with increasing redundancy, but after a certain point becomes worse owing to crystal decay, so that there is an optimal amount of data to include for the sulfur substructure solution. Further, the success rate in locating S atoms is dramatically improved at lower resolutions and in a manner similar to data quality, there exists an optimal resolution at which the likelihood of solving the substructure is maximized. Based on these observations, a strategy for SAD data collection and substructure solution is suggested.

1. Introduction

Over the past decade, the sequencing of complete genomes has spurred a growing need for structural information on the encoded gene products that has driven the development of high-throughput crystallography (reviewed in Blundell *et al.*, 2002; Blundell & Patel, 2004; Brunzelle *et al.*, 2003; Heinemann *et al.*, 2001; Kuhn *et al.*, 2002; Loll, 2003; Pusey *et al.*, 2005; Rupp, 2003; Sharff & Jhoti, 2003). At the same time, methodological advances in X-ray crystallography such as the increased availability of synchrotron radiation, cryocrystallography and improved software for heavy-atom location and density modification have increased our ability to solve protein structures using single-wavelength anomalous diffraction (SAD) based on the weak anomalous signal from inherently present S atoms (hereafter called sulfur-SAD or S-SAD).

The anomalous scattering length of S atoms (f'' in electrons) is maximal at $\lambda = 5.02$ Å and, considering the wavelengths normally used for protein crystallography, the strength increases with increasing wavelength, with values of 0.234 at 1.0 Å to 0.557 at 1.54 Å (Cu $K\alpha$) to 0.697 at 1.75 Å to 1.141 at 2.29 Å (Cr $K\alpha$) (Creagh, 2004; Cromer & Liberman, 1970, 1981; Ramagopal *et al.*, 2003). While $\lambda \simeq 1.75$ Å has been commonly used for synchrotron-based S-SAD structure

research papers

Table 1
Data-collection statistics.

Values in parentheses are for the highest resolution shells.

No. of images	1260
$\Delta\varphi$ (°)	1
Wavelength (Å)	1.54
Exposure time per image (min)	20
Resolution limits (Å)	23.6–1.90 (1.97–1.90)
Unique observations	12754
Multiplicity†	112.3 (98.8)
Completeness (%)	99.9 (100)
Average I/σ‡	47.4 (14.2)
Mosaicity (°)	0.42
R_{meas}†§	11.2 (66.8)
$R_{mrgd\text{-}F}$†§	2.5 (11.7)

† Bijvoets were not merged. ‡ An $|F|/\sigma|F| \geq 0$ cutoff was used for data processing. § R_{meas} is the multiplicity-weighted merging R factor and $R_{mrgd\text{-}F}$ is an indicator of the quality of reduced data (Diederichs & Karplus, 1997).

determination (Dauter *et al.*, 1999; Liu *et al.*, 2000; Gordon *et al.*, 2001; Brown *et al.*, 2002; Li *et al.*, 2002; Micossi *et al.*, 2002; Ramagopal *et al.*, 2003; Lartigue *et al.*, 2004), it has been shown recently that $\lambda \simeq 2.1$ Å is the wavelength at which one can obtain the highest anomalous signal-to-noise ratio for S-SAD data collection using synchrotron radiation (Carugo *et al.*, 2005; Mueller-Dieckmann *et al.*, 2004, 2005). While most S-SAD studies published to date have used synchrotron radiation because of the need for very high accuracy data, home sources are still more readily available and recent efforts have shown that an in-house Cr $K\alpha$ source ($\lambda \simeq 2.29$ Å) can be used (Yang *et al.*, 2003; Kitago *et al.*, 2005; Watanabe *et al.*, 2005) and even a Cu $K\alpha$ X-ray source is a viable venue for S-SAD (Yang & Pflugrath, 2001; Lemke *et al.*, 2002; Debreczeni, Bunkoczi, Girmann *et al.*, 2003; Olsen *et al.*, 2004; Debreczeni, Girmann *et al.*, 2003). Indeed, with $\lambda = 1.54$ Å radiation the Bijvoet ratios for an 'average' 100-residue prokaryotic and eukaryotic protein are ~0.75% (3.3 sulfurs) and 0.86% (4.4 sulfurs), respectively (Ramagopal *et al.*, 2003; Nagem *et al.*, 2005), both of which are above the value of 0.6% historically considered as the minimum Bijvoet ratio sufficient to solve a protein structure using S-SAD (Wang, 1985). While it is known that the Wang limit is based on error-free data and is therefore not the best indicator to predict successful structure solution (Zwart, 2005), the theoretical values for prokaryotic and eukaryotic proteins still suggest that in theory the in-house X-ray source holds great promise for use in high-throughput crystallography.

For the structures solved *de novo* using S-SAD based on Cu $K\alpha$ radiation (Lemke *et al.*, 2002; Debreczeni, Bunkóczi, Girmann *et al.*, 2003; Olsen *et al.*, 2004; Debreczeni, Girmann *et al.*, 2003), highly redundant data were collected from strongly diffracting crystals to maximize the anomalous signal. While one structure had a low Bijvoet ratio of 0.86 (Lemke *et al.*, 2002), the others had Bijvoet ratios that were much higher, between 1.6 and 3.1, and thus these structures were not stringent tests of the method. The common themes of strongly diffracting crystals and high resolution are often considered to be essential for successful phasing by S-SAD (Dauter & Nagem, 2002; Debreczeni, Bunkóczi, Ma *et al.*, 2003; Nagem *et*

al., 2005), despite some evidence to the contrary from both theory (Wang, 1985) and experiment (Liu *et al.*, 2000; Dauter & Nagem, 2002; Ramagopal *et al.*, 2003).

We have recently solved the structure of a 13.2 kDa protein, Ptr ToxA from the fungal pathogen *Pyrenophora tritici-repentis*, using highly redundant laboratory-based (Cu $K\alpha$) sulfur SAD data to locate and phase based on three strong sulfur sites (one Met, one Cys and one sulfate) corresponding to a Bijvoet ratio of 0.68% (Sarma *et al.*, 2005). Accounting for a fourth weak sulfate ion discovered during this retrospective analysis, the true Bijvoet ratio for the structure is closer to 0.77%. In either case, the ratio is low enough to be considered a challenging case. Here, we present a systematic analysis of the effects of data quality and resolution on S-SAD substructure solution and phasing. Our results counter the commonly held views that increased redundancy and higher resolution are always better.

2. Materials and methods

2.1. Purification and crystallization

Mature ToxA was purified from culture filtrates as previously described (Sarma *et al.*, 2005; Tuori *et al.*, 1995). Crystals with cubic morphology and dimensions of $0.15 \times 0.15 \times 0.15$ mm were grown from a 1:1 drop using a reservoir solution of 0.5 M (NH$_4$)$_2$SO$_4$, 15% dioxane and 0.1 M MES pH 6.5. The protein crystallizes in the cubic space group $P2_13$, with unit-cell parameters $a = b = c = 78.2$ Å, a solvent content of 60% and one molecule in the asymmetric unit. The reservoir solution with 10% glycerol served as a cryoprotectant. Prior to data collection, the crystals were flash-frozen in loops by dipping in liquid nitrogen after a 2 min incubation in the cryoprotectant.

2.2. Data collection

As summarized in Table 1, extensive SAD data were collected using no special data-collection strategies such as collecting Bijvoets close in time.

Data extending to 1.9 Å resolution were collected from a single crystal that was not fully fresh, but had been exposed for ~13 min to synchrotron radiation [beamline 8.2.1 of the Advanced Light Source (ALS); Berkeley National Laboratory; 2×2 CCD array (ADSC Q210) detector; $\lambda = 1.75$ Å]. For this in-house data collection, the crystal was rotated around the φ-axis to collect a total of 1260 images (~17 d total exposure) using Cu $K\alpha$ radiation (Rigaku RU-H3R rotating-anode generator operating at 50 kV, 100 mA with an R-AXIS IV image-plate detector; Table 1). The exposure time (20 min) per image was a compromise designed to collect highly redundant medium-resolution data even though longer exposure times could have yielded higher quality single-pass data. After the first 360 images, the φ angle was offset by 0.5° for the next 360 images in order to collect as 'fullies' those reflections that were collected as partials in the earlier scan and the offset was repeated every 360 images. With the data represented as seven sequential 180° data sets (designated A

Table 2
Unique 180° data sets.

All data were processed to 1.90 Å resolution, with the values in parentheses indicating the values for the highest resolution shell (1.97–1.90 Å). For each 180° of data collection, the redundancy is ∼21-fold. For the later data sets, the lower than expected redundancies are a consequence of a higher rate of rejection.

	A	B	C	D	E	F	G
Oscillation range (°)	0–180	180–360	0.5–180.5	180.5–360.5	0–180	180–360	0.5–180.5
Unique obs.	12578	12605	12542	12587	12479	12479	12348
Multiplicity†	21.0 (18.9)	21.0 (19.1)	21.0 (18.5)	21.0 (18.8)	20.7 (17.0)	20.8 (17.3)	20.0 (12.5)
Average I/σ‡	34.7 (8.0)	33.5 (7.7)	30.4 (6.0)	31.1 (6.1)	28.0 (4.5)	28.3 (4.3)	25.1 (2.8)
R_{meas}†§ (%)	7.3 (46.5)	7.0 (45.4)	7.8 (60.5)	7.8 (59.8)	8.8 (76.2)	8.6 (80.1)	9.8 (102.0)
R_{mrgd-F}†§ (%)	4.5 (20.6)	4.2 (19.8)	5.3 (28.5)	5.0 (26.0)	6.4 (36.1)	6.3 (39.0)	8.1 (55.0)
R_{mrgd-I}†§ (%)	3.0 (21.0)	3.0 (19.9)	3.2 (28.7)	3.2 (27.0)	3.7 (36.2)	3.7 (39.9)	4.2 (56.9)
$R_{p.i.m.}$¶ (%)	1.6 (10.5)	1.5 (10.3)	1.7 (13.8)	1.7 (13.5)	1.9 (18.0)	1.9 (18.9)	2.2 (27.8)
R_{anom}†† (%)	3.1 (16.7)	3.1 (18.7)	3.3 (23.5)	3.3 (23.9)	3.6 (29.6)	3.6 (32.6)	4.0 (44.4)
Unit cell‡‡ (Å)	78.203	78.225	78.243	78.256	78.267	78.277	78.278

† Bijvoets were not merged. ‡ An $|F|/\sigma|F| \geq 0$ cutoff was used for data processing. § R_{meas} is the multiplicity-weighted merging R factor. R_{mrgd-F} and R_{mrgd-I} are indicators of reduced data quality (Diederichs & Karplus, 1997). ¶ $R_{p.i.m.}$ is the precision-indicating merging R factor, similar to R_{mrgd-I} (Weiss, 2001). †† $R_{anom} = \sum_{hkl}[(|I^+ - I^-|)/(|I^+ + I^-|/2)]$. ‡‡ Post-refined unit-cell parameter as output from *SCALEPACK* (Otwinowski & Minor, 1997).

Table 3
Data sets with increasing redundancies.

The data were processed to 1.90 Å, with the values in parentheses indicating values for the highest resolution shell (1.97–1.90 Å). For each 180° of data collection, the redundancy is ∼21-fold. For the highest redundancy data sets, the lower than expected redundancies arise from a higher rate of rejection.

	'30'	'60'	'90'	'120'	'150'	A	A–B	A–C	A–D	A–E	A–F	A–G
No. of images	30	60	90	120	150	180	360	540	720	900	1080	1260
Unique obs.	12246	12473	12504	12531	12547	12578	12689	12719	12743	12749	12753	12754
Multiplicity†	3.5 (3.1)	7.0 (6.3)	10.5 (9.4)	14.0 (12.6)	17.5 (15.7)	21.0 (18.9)	42.0 (37.5)	62.9 (56.0)	83.9 (74.9)	102.8 (90.9)	110.4 (99.4)	112.3 (98.8)
Average I/σ‡	17.3 (3.9)	23.8 (5.4)	25.5 (6.2)	28.6 (6.7)	31.6 (7.3)	34.7 (8.0)	42.4 (11.2)	50.7 (12.8)	52.2 (14.0)	52.9 (14.6)	50.6 (14.9)	47.4 (14.2)
R_{meas}†§ (%)	5.6 (33.5)	5.8 (36.2)	6.3 (40.0)	6.8 (44.4)	7.1 (45.9)	7.3 (46.5)	7.6 (46.0)	8.2 (50.7)	8.7 (53.0)	9.5 (57.3)	10.3 (62.0)	11.2 (66.8)
R_{mrgd-F}†§ (%)	8.5 (35.4)	6.6 (28.8)	5.7 (24.6)	5.1 (22.6)	4.8 (22.1)	4.5 (20.6)	3.1 (14.2)	2.8 (12.5)	2.5 (11.4)	2.4 (11.2)	2.4 (10.9)	2.5 (11.7)
R_{mrgd-I}†§ (%)	6.2 (37.5)	4.5 (29.2)	3.8 (24.0)	3.4 (22.9)	3.2 (22.1)	3.0 (21.0)	2.2 (14.6)	2.0 (13.0)	1.8 (11.8)	1.8 (11.2)	1.8 (11.6)	1.9 (12.2)
$R_{p.i.m}$¶ (%)	2.9 (17.4)	2.2 (14.3)	1.9 (12.8)	1.8 (12.3)	1.7 (11.4)	1.6 (10.5)	1.2 (7.4)	1.0 (6.6)	0.9 (5.2)	0.8 (5.0)	0.8 (5.0)	0.8 (5.5)
R_{anom}†† (%)	6.1 (32.6)	4.8 (27.0)	3.9 (23.1)	3.5 (21.2)	3.3 (19.7)	3.1 (17.9)	2.3 (12.6)	2.1 (11.2)	2.0 (10.4)	2.0 (10.1)	2.2 (10.7)	2.7 (11.9)
Unit cell‡‡ (Å)	78.189	78.189	78.191	78.196	78.201	78.203	78.214	78.222	78.229	78.235	78.241	78.246

† Bijvoets were not merged. ‡ An $|F|/\sigma|F| \geq 0$ cutoff was used for data processing. § R_{meas} is the multiplicity-weighted merging R factor. R_{mrgd-F} and R_{mrgd-I} are indicators of reduced data quality (Diederichs & Karplus, 1997). ¶ $R_{p.i.m}$ is the precision-indicating merging R factor, similar to R_{mrgd-I} (Weiss, 2001). †† $R_{anom} = \sum_{hkl}[(|I^+ - I^-|)/(|I^+ + I^-|/2)]$. ‡‡ Post-refined unit-cell parameter as output from *SCALEPACK* (Otwinowski & Minor, 1997).

to *G*; Table 2), the offset strategy was such that the oscillation ranges of images in data sets *A*, *B* and *C* exactly matched those in data sets *E*, *F* and *G*, respectively.

As reported in §3, the data were partitioned in various ways to create a series of test data sets (Tables 2 and 3). Each of these data sets was indexed and integrated using *DENZO* (Otwinowski & Minor, 1997). Scaling and merging were carried out using *SCALEPACK* (Otwinowski & Minor, 1997). No special scaling protocols (Weiss *et al.*, 2005) or radiation-damage corrections (Diederichs *et al.*, 2003) were used. The raw diffraction data are available on request.

2.3. The reference model

As described by Sarma *et al.* (2005), the 13.2 kDa ToxA structure was determined at 3.0 Å resolution using the sulfur anomalous signal in the first 540 images of the in-house data (data set *A*–*C* in Table 3). The automated model-building program *ARP/wARP* (Perrakis *et al.*, 1999) was used to extend the 3.0 Å phases to 1.65 Å (available from a separate data collection at a synchrotron using λ = 1.0 Å X-rays) and refinement yielded the final model with an *R* and R_{free} of 15.7 and 18.3%, respectively (Sarma *et al.*, 2005). This model (PDB code 1zld) was subsequently used to calculate the phases that are used here as the reference 'true' phases.

2.4. Substructure solution and phase determination

For the systematic analyses reported here, the various anomalous data sets were prepared for substructure solution using the program *XPREP* (Bruker Analytical X-ray Systems). For all data sets except data set *G*, anomalous difference maps using the calculated phases showed a weak fourth peak (less than half as strong as the other sites) that corresponded to a low-occupancy sulfate ion that was not included in the original phasing model. To maximize the relevance of this analysis in mimicking a true *de novo* structure determination (which might miss weak sites), we have not used the minor site for phasing in this retrospective study. Thus, searches for three sulfur positions were carried out with the program *SHELXD* (Schneider & Sheldrick, 2002) using 1000 search cycles. An approximate number of 'correct' solutions was estimated by choosing a correlation-coefficient cutoff partway between the group of 'correct' solutions with significantly higher correlation coefficients and the group of 'incorrect' solutions with lower correlation coefficients. For

Table 4
Pairwise R factors for different data sets.

Values in parentheses are for the highest resolution bin (1.97–1.90 Å).

	R factor† (%)
A versus B	9.7 (16.4)‡
A versus C	9.2 (19.3)‡
A versus D	17.5 (28.2)
A versus E	17.9 (31.4)
A versus F	25.8 (42.1)
A versus G	26.8 (45.2)

† R factor $= \sum |F_{\mathrm{obs},A} - F_{\mathrm{obs},X}| / \sum |F_{\mathrm{obs},A}|$, where X represents one of the other data sets (B through G). ‡ The lower overall R but higher 'highest resolution bin' R for A versus C compared with A versus B is consistent with the systematic differences owing to data-collection range (larger for data set B; see text) affecting low- and high-resolution data equally, whereas the crystal decay (greater for data set C) affects the high-resolution data.

each analysis, the top three sulfur positions were refined (positions and occupancies) and phasing was carried out using the program *MLPHARE* (Otwinowski, 1991). The phase ambiguity was resolved using the solvent-flattening procedure of *DM* (Cowtan & Zhang, 1999). Thus, each run of substructure determination and phasing was carried out as a 'new' run, imitating a *de novo* structure-determination experiment, *i.e.* without *a priori* substructure and phase information. For all additional data analyses, programs from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) and *CNS* (Brünger *et al.*, 1998) were used.

3. Results

3.1. Data sets

In order to systematically study the effects of data quality on substructure solution and phasing, the 1260 oscillation images in the LAB data set were divided into seven consecutive 180-image data sets labeled A through G, each with ~20-fold redundancy (Table 2). A further 11 data sets were created to study the effect of redundancy. As an extension of
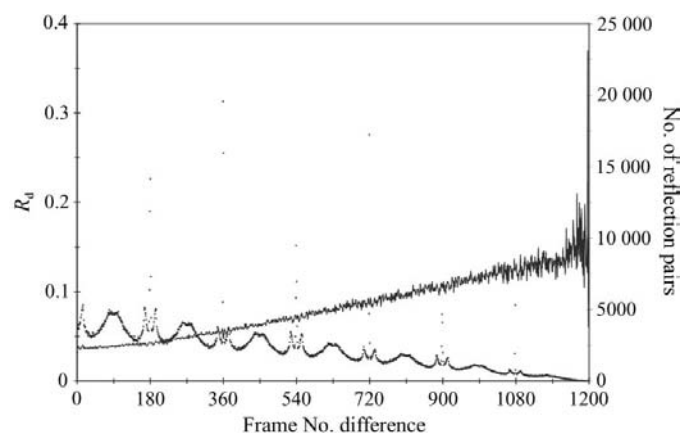

Figure 1
Quantifying crystal decay during data collection. The decay R factor (R_d) is plotted as a function of frame-number difference with the number of reflection pairs contributing to the calculation. The smooth increase in R_d is indicative of gradual crystal decay during data collection.

the above-mentioned nomenclature, six data sets were labeled A–B (360 images), A–C (540 images) and so on to A–G (1260 images) (Table 3). The five additional data sets created to test the lower limit of data quality that would allow a structure solution are designated as '30', '60', '90', '120' and '150', representing the number of images in each set (Table 3).

3.2. Crystal decay

Despite the crystals being frozen, decay was evident during data collection as a visible loss in resolution between the first and last image (not shown), as a decrease in the average signal-to-noise ratio (I/σ) of the reduced data and as increases in the R_{meas} and $R_{\mathrm{mrgd-}F}$ values (Diederichs & Karplus, 1997) for data sets A through G (Table 2). While increases in the unit-cell volume can be indicative of radiation damage (Ravelli *et al.*, 2002), for ToxA the volume increased a minimal amount (~0.3% from data set A to data set G; Table 2). Quantitatively, plotting decay R factors (R_d) as a function of image number (Diederichs, 2006) clearly shows that the crystal decay was smooth and continuous during the whole exposure time (Fig. 1). For an estimate of how crystal decay impacted on data accuracy, R factors were calculated between data sets B through G and data set A as a function of resolution (Table 4). Data set A was used as a reference since it was collected when the crystal was freshest. The R factors were lowest between data sets B and A and were highest between data sets G and A. Interestingly, the values were paired such that the B *versus* A R values are very similar to those of C *versus* A and so on (Table 4). A plausible explanation is that data sets A, C, E and G were collected over the same oscillation ranges (counting the 0.5° offset) and so have minimal systematic differences related to crystal geometry (*e.g.* absorption) or detector imperfections (*e.g.* non-homogeneity). Thus, the R values calculated using these data sets are more purely influenced by crystal decay. On the other hand, the comparison of data set A with data sets B, D and F will reflect decay plus systematic errors and this will shift the R factors to higher values for those data sets. Because systematic errors may be less resolution-dependent, this would also explain why the R factors for data sets C, E and G are higher at low resolution.

For the data sets of increasing redundancy (A through A–G), the indicators of reduced data quality, I/σ and $R_{\mathrm{mrgd-}F}$, also show the effects of decay (Table 3). I/σ increases from data set A to data set A–D, after which the values become worse despite the increase in redundancy. Similarly, $R_{\mathrm{mrgd-}F}$ (lower values of which indicate better data; Diederichs & Karplus, 1997) decreases from data set A to A–D, after which no further improvement is seen. According to these statistical measures, increased redundancy results in better quality data up to data set A–D, after which the overall data quality is negatively impacted by the merging of subsequent data.

Consistent with observations using other proteins (Burmeister, 2000; O'Neill *et al.*, 2002; Ravelli & McSweeney, 2000; Weik *et al.*, 2000), a difference map calculated between data sets A and G ($F_{\mathrm{obs, data\ set}\ A} - F_{\mathrm{obs, data\ set}\ G}$, $\alpha_{\mathrm{calc, reference}}$) showed that radiation damage involved specific loss of density

for the carboxylate groups and disulfide S atoms. Maximal local density loss was about 40%.

3.3. Assessment of anomalous signal

For SAD phasing, the quality of the anomalous signal is most important, so it is of interest to see how this is impacted by the trade-off between increasing redundancy and crystal decay. Even without phases, for any single data set it is possible to assess the strength of the anomalous signal by splitting the data set in half and calculating the anomalous correlation coefficients between the two halves (Evans, 2006; Schneider & Sheldrick, 2002; Zwart, 2005). This has been shown to be a more reliable indicator of anomalous signal than the $\langle \Delta F \rangle / \langle F \rangle$ ratios (Zwart, 2005). The individual data sets A through G show a trend of decreasing correlation, indicating that decay does impact on the quality of the anomalous signal (Fig. 2a). The correlation remains fairly high at low resolution, but the resolution at which it drops below 20% decreases from near 3.5 Å for data set A to near 4.5 Å for G. The trend in the anomalous correlations for data sets with increasing redundancies is dramatic. The anomalous signal steadily increases from data set '30' to A–D. Despite further increases in

redundancy, no further increase is seen for data sets between A–D and A–G (Fig. 2b).

The ratio of R_{anom} to $R_{p.i.m.}$ has also been used as an indicator of the level of anomalous signal (Mueller-Dieckmann et al., 2004, 2005; Weiss, 2001; Weiss et al., 2004, 2005; Weiss, Sicker, Djinovic Carugo et al., 2001; Weiss, Sicker & Hilgenfeld, 2001), as it compares the size of the anomalous differences with the precision of the intensities from which those differences are calculated (Weiss, 2001). Empirically, a ratio of >1.5 has been considered as indicating anomalous data (Mueller-Dieckmann et al., 2005). For all data sets analyzed here, the ratio is between 1.5 and 2, fairly independent of resolution and redundancy. The lack of clear trends such as were seen with the anomalous correlation coefficients is surprising and suggests that for these data the $R_{anom}/R_{p.i.m.}$ ratios are not very informative. We do not have an explanation for this.

Given known phases (from the model), a more accurate assessment of the anomalous signal present in each data set is provided by peak heights in anomalous difference Fourier maps (Fig. 3). While this could be performed as a function of resolution, we have performed this analysis at 3.0 Å to observe the trends. In all cases, the top three peaks correspond to the three sulfur sites used for the original structure solution (Sarma et al., 2005). Consistent with the other data-quality statistics, the peak heights as a function of data redundancy increase steadily from data set '30' to A–D or A–E (Fig. 3). In contrast to the other measures of data quality, the anomalous peak heights do not just level out after data set A–E, but they decrease dramatically to data set A–G (Fig. 3). By this measure, data set A–G has an even lower anomalous signal than data set A alone! This suggests that the anomalous correlation coefficients reflect accurately the increase in
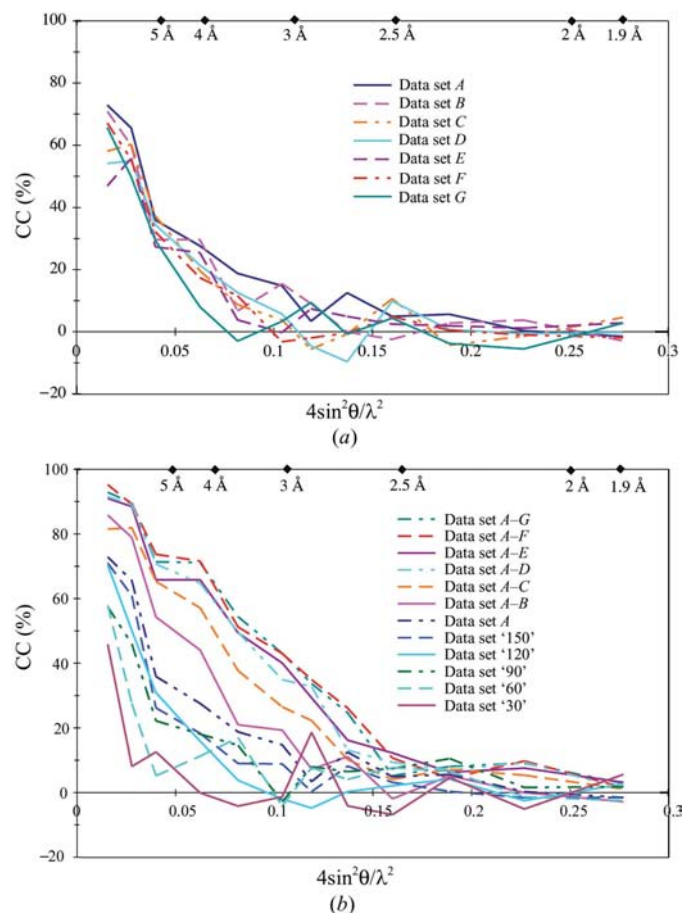

Figure 2
Phase-independent assessment of the anomalous signal. Correlation coefficients (%) between the anomalous differences for (a) data sets A through G and (b) data sets A through A–G are plotted against resolution.
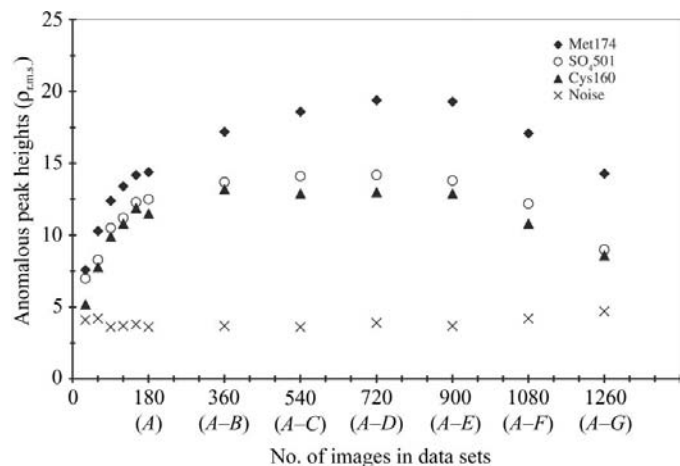

Figure 3
Effect of redundancy on the anomalous peak heights at 3.0 Å resolution. The anomalous peak heights (in $\rho_{r.m.s.}$) of the top three peaks, Met174 (closed diamonds), Sulfate501 (open circles), Cys160 (closed triangles), and the largest noise peak plotted as a function of number of images in each data set. The fourth largest anomalous peak is a second sulfate ion (not shown) that is weak but consistently present. For data sets B through G, the respective anomalous peak heights are data set B, 14.3, 10.1, 11.4, 4.0; data set C, 14.9, 10.6, 6.8, 4.1; data set D, 14.0, 10.1, 9.2, 3.7; data set E, 12.9, 8.5, 8.4, 3.9; data set F, 10.9, 9.0, 7.4, 4.0; data set G, 11.5, 8.0, 7.2, 3.6.

Table 5
Substructure solution for increasing redundancy data sets.

The first line indicates the approximate number of 'correct' solutions out of 1000 search trials, with the correlation coefficient of all reflections of the best solution (as output by *SHELXD*) in parentheses. The second line of each entry is the number of correct sulfur positions (out of the three) located by *SHELXD* (see §2).

	'30'	'60'	'90'	'120'	'150'	A	A–B	A–C	A–D	A–E	A–F	A–G
2.0 Å	0 (9.1)	0 (8.3)	0 (9.2)	0 (8.8)	0 (8.4)	0 (8.5)	0 (8.7)	12 (16.4)	20 (16.9)	27 (16.9)	24 (14.0)	0 (8.4)
	0	0	0	1	1	2	3	3	3	3	3	2
2.5 Å	0 (11.8)	0 (11.1)	0 (12.0)	0 (11.9)	22 (15.3)	9 (20.1)	121 (24.4)	214 (29.5)	236 (28.7)	203 (28.7)	115 (23.9)	4 (19.1)
	0	0	3	3	3	3	3	3	3	3	3	3
3.0 Å	0 (17.4)	0 (18.4)	3 (19.6)	44 (23.5)	48 (24.5)	80 (27.7)	411 (34.7)	408 (39.7)	435 (38.6)	348 (37.6)	239 (30.6)	74 (22.5)
	0	0	3	3	3	3	3	3	3	3	3	3
3.5 Å	0 (20.4)	0 (22.7)	52 (27.3)	218 (30.6)	79 (31.8)	165 (33.4)	502 (41.4)	523 (45.4)	525 (44.9)	332 (42.8)	267 (36.7)	120 (29.9)
	0	2	3	3	3	3	3	3	3	3	3	3

anomalous signal, but are not sensitive reporters of the loss in anomalous signal arising from the addition of decayed data. As a measure of the effects of crystal decay independent of redundancy changes, the sulfur anomalous peak heights for data sets *A* to *G* show a steady decrease with increased crystal decay (see legend to Fig. 3).

3.4. Substructure determination

A systematic study of the effects of redundancy, crystal decay and resolution on substructure determination was carried out for data sets '30' through *A–G* at resolution cutoffs of 2.0, 2.5, 3.0 and 3.5 Å. Representative results are given by histograms of correlation coefficients at 3.0 Å resolution (Fig. 4*a*) and the results at all four resolutions are summarized in terms of the correlation coefficients and number of correct sulfur positions found for the best trial and the number of trials significantly above the noise level (Table 5). At all four resolutions, the indicators show that the ability to find the correct sites is maximal for data sets *A–C* or *A–D* (or *A–E* at 2.0 Å resolution) and steadily decreases both toward lower or higher redundancy. Interestingly, as was seen for the anomalous difference map peak heights, data set *A–G* seems roughly similar in phasing power to data set *A*. The best data sets (*A–C* and *A–D*) led to correct solutions at all resolutions, but for any given data set the success rate strongly increased with decreasing resolution, for example from 3% at 2.0 Å to 53% at 3.5 Å resolution for data set *A–D* (Table 5). Also, the searches carried out at 2.0 Å resolution only yielded a distinct population of correct solutions for data sets *A–C* through *A–F*, whereas those carried out using the 3.5 Å resolution cutoff yielded distinctly correct solutions for a broader set of data sets: '90' through *A–G* (Table 5). Finally, the separation of the correlation coefficient of the 'correct' trials from the 'incorrect' trials is increasingly pronounced at the lower resolutions (data not shown).

To separately assess the effects of crystal decay on substructure solution, with no influence of redundancy, substructure solution attempts were made using data sets *A* through *G*, each of which is based on 180° of data. We only tested the behavior using 3.0 and 3.5 Å resolution cutoffs, since we expected solution clarity to decrease with decay, and only these resolutions gave sufficiently strong success rates for data set *A* that there would be room for monitoring a decrease

in quality. The strong trend at both 3.0 and 3.5 Å resolution is a steady decrease in the number of correct trials with increased decay to the point that for data set *G*, even though a correct solution was found, its correlation coefficient was not
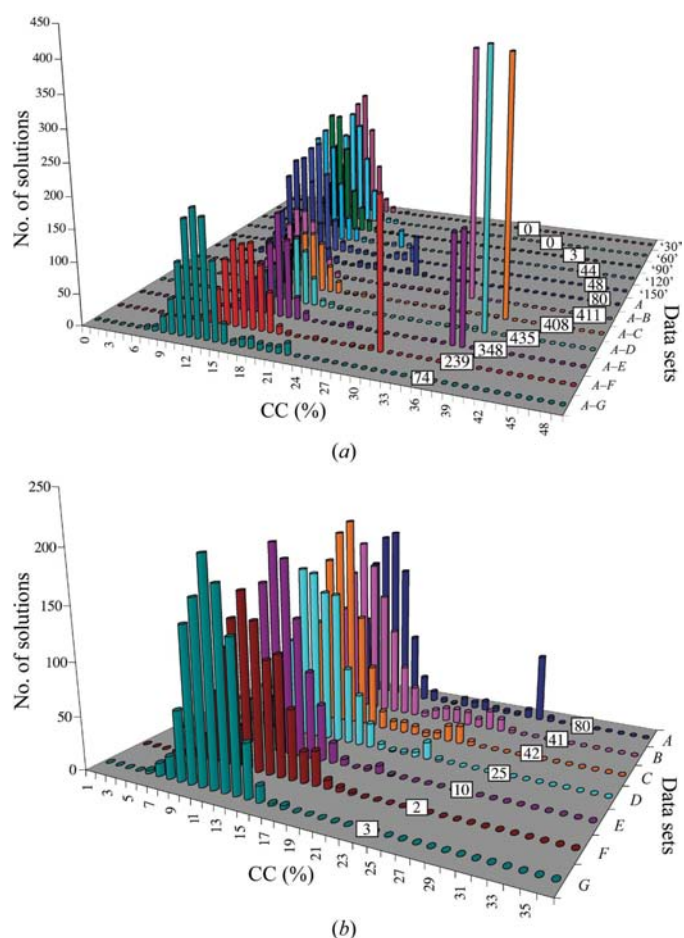

Figure 4
Effects of redundancy and radiation damage on substructure solution results using a 3.0 Å resolution cutoff. Histograms are shown of the correlation coefficients for the 1000 substructure-solution trials carried out at 3.0 Å resolution for (*a*) data sets '30' through *A–G* and (*b*) data sets *A* through *G*. Each histogram has two major peaks, one at higher correlation coefficients corresponding to correct solutions and one at lower correlation coefficient corresponding to non-solutions. The printed number gives the total number of trials in the high correlation coefficient group. The higher the number of correct solutions and the greater the separation of the high and low groups, the clearer the solution.

Table 6
Substructure solution for the 180° data sets.

Three (of three) correct sulfur positions are located using the data sets at both resolutions. The top line indicates the approximate number of 'correct' solutions out of 1000 search trials with the correlation coefficient of all reflections of the best solution (as output by *SHELXD*) in parentheses.
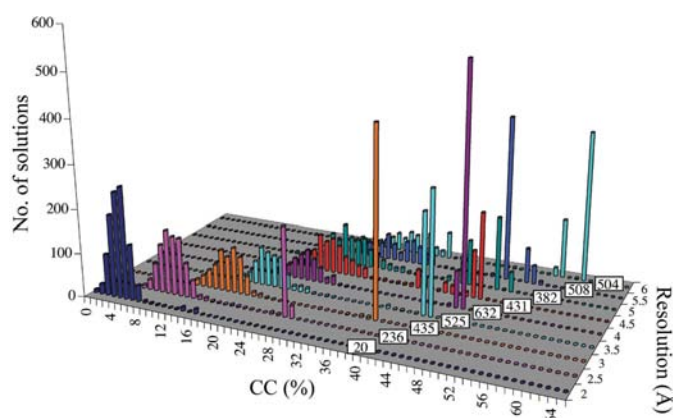
	A	B	C	D	E	F	G
3.0 Å	80 (27.7)	41 (25.2)	42 (23.6)	25 (22.2)	10 (20.1)	2 (19.5)	3 (16.5)
3.5 Å	165 (33.4)	110 (31.6)	77 (31.0)	49 (30.2)	100 (26.0)	19 (26.3)	22 (22.8)

very distinct from those of the non-solutions (Fig. 4*b* and Table 6).

Faced with the result that the lower resolution cutoffs yielded better success, we used data set *A–D*, the best performing data set at all resolutions thus far tested, to attempt substructure solution at five additional resolution cutoffs: 4.0, 4.5, 5.0, 5.5 and 6.0 Å. Surprisingly, clear solutions were obtained all the way up to 6.0 Å with ever-increasing correlation coefficients (Fig. 5). The maximal success rate was 63% using the 4.0 Å resolution cutoff.

3.5. Phase determination

Phases were calculated for each data set and resolution combination using the sulfur substructure determined for that combination. Phase quality was evaluated by calculating the differences between the experimental phases (both before and after density modification) and the final refined 1.65 Å calculated phases (see §2). Fig. 6(*a*) shows the set of phase differences calculated for data set *A–D* using three sulfur positions at various resolutions. At all resolutions between 2.0 and 4.5 Å, phasing succeeded with comparable accuracy (while phase determination below 4.5 Å worked well, the

density-modification routine did not work at these resolutions). These results are representative, as all data sets at all resolutions led to similar quality phase sets that would easily be accurate enough to guide modeling. This even extends to data sets that did not allow substructure solution; for instance, when given the correct sulfur sites, even data set '30' yielded phases at 3.0 Å resolution equivalent to those seen in Fig. 6(*a*) ($\Delta\varphi_{\text{pre-DM}}/\Delta\varphi_{\text{post-DM}} = 66.3/51.2°$).

In order to find out the phasing power of the 'unsuccessful' substructure results, phases were calculated using the incomplete sulfur models found using certain data sets at 2.0 Å resolution (Fig. 6*b*). The phases calculated using two and three

Figure 6
Quality of experimental phases. Phase differences between experimental phases [before (thin lines) and after (thick lines) density modification] and final refined phases are plotted as a function of resolution. (*a*) Experimental phases were calculated using three sulfur positions and data set *A–D* at six different resolutions: 2.0 Å (blue), 2.5 Å (violet), 3.0 Å (orange), 3.5 Å (cyan), 4.0 Å (purple) and 4.5 Å (red). These differences are representative of the quality of phases calculated using different resolution cutoffs. (*b*) Experimental phases calculated at 2.0 Å resolution using data set *A–D* (three sites, blue), data set *A* (two sites, violet) and data set '150' (one site, orange) are shown. They are representative of phase errors from any data set using three, two or one sulfur site(s) for phasing. Phases calculated from substructures having no correct sites were random (90° phase difference), as expected.
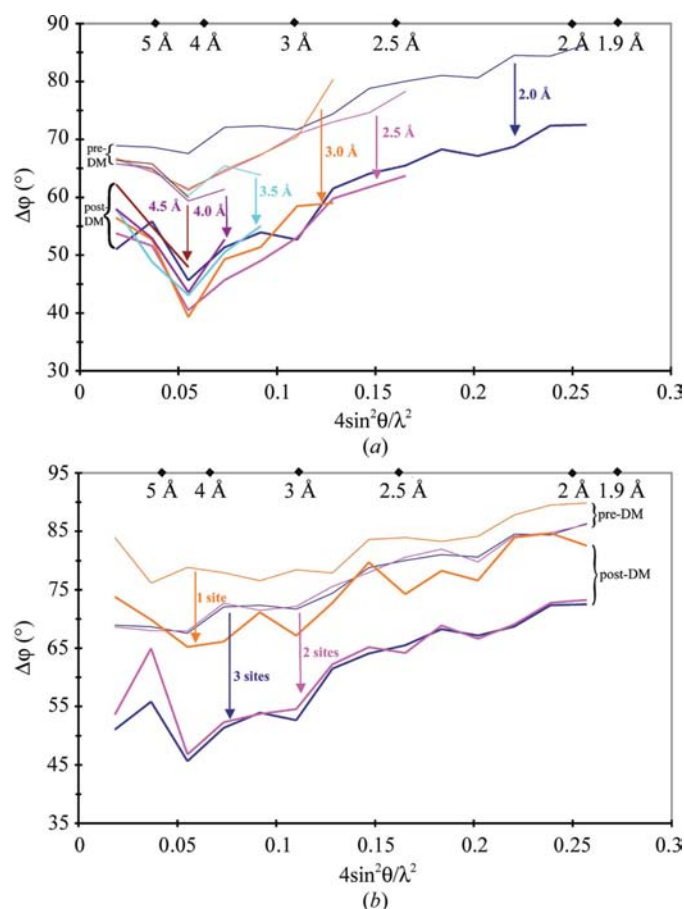

Figure 5
Effect of resolution cutoff on substructure-solution success. Histograms for the correlation coefficients for the 1000 substructure solutions are shown for data set *A–D* using nine resolution cutoffs in the range 2.0–6.0 Å at 0.5 Å intervals. See the legend of Fig. 4 for a description of the plot. The closest distances between the three sulfurs, including symmetry-related atoms, are 25.0 Å for Met174–Met174, 23.8 Å for Cys160–Cys160, 8.9 Å for Sulfate501–Sulfate501, 14.7 Å for Met174–Sulfate501, 18.8 Å for Sulfate501–Cys160 and 13.8 Å for Met174–Cys160, so that for locating S atoms in this crystal structure even 6 Å can be considered atomic resolution.

sulfur sites are of similar quality and the automated model-building program *ARP/wARP* was able to build ∼90% of the model. Phases calculated with only one site were much less accurate and experimental maps calculated from them might be interpretable with difficulty, but *ARP/wARP* failed to build a model.

4. Discussion

This study was undertaken to understand the effects of redundancy, crystal decay and resolution limits on the ability of a given data set to lead to substructure solution and subsequent phasing by S-SAD. Our analyses of the ToxA data reinforce a number of well known principles: that data quality is of great importance and that increased multiplicity can improve data quality immensely. Furthermore, our analyses allow us to draw two additional conclusions of practical importance. Firstly, for data collected from a single crystal, as multiplicity increases the data improve and then become worse, so that there is an optimal amount of data to include for successful substructure solution. Secondly, substructure solution success is a sensitive function of resolution and there is an optimal resolution cutoff for successful substructure solution that will be structure and data-quality dependent.

4.1. Higher multiplicity does not necessarily mean better data

It has been pointed out that crystal decay can limit data quality, but in published S-SAD structure determinations data quality and phasing power always increase monotonically with increasing redundancy (Dauter *et al.*, 2002; Dauter & Nagem, 2002; Debreczeni, Bunkóczi, Girmann *et al.*, 2003; Olsen *et al.*, 2004; Ramagopal *et al.*, 2003; Roeser *et al.*, 2005; Zwart, 2005; Dauter & Adamiak, 2001; Stevenson *et al.*, 2004; Debreczeni, Bunkóczi, Ma *et al.*, 2003). Here, this is not the case.

The behavior of decay *R* factors (Fig. 1), *R* factors calculated between data sets *B* to *G versus A* (Table 4), the anomalous signal for data sets *A* through *G* (Fig. 2*a*) and substructure-solution success using these data sets (Fig. 4*b* and Table 6) all show that crystal decay is a process that continuously decreases data quality with increased exposure.

Despite the fact that decay begins with the first image, the properties of the data sets with increasing redundancies ('30' through *A–G*) indicate that for a certain period (in this case through data set *A–D*) adding more data improves overall data quality, as seen in the I/σ, $R_{\mathrm{mrgd}\text{-}F}$ (Table 3), the anomalous signal (Fig. 2*b*) and the substructure-solution success (Fig. 4*a* and Table 5). However, there is then a change with the merging of data sets *E*, *F* and *G*. While the data-quality statistical indicators continue to increase or stay the same, the anomalous signal quality is decreased as shown by the dramatic drop in anomalous difference map peak heights (Fig. 3) and in substructure-solution success rates (Fig. 4*a*).

This disparate behavior between the data quality statistical measures and the actual data quality of the anomalous signal suggests that the standard statistics are not sensitive indicators of the anomalous signal quality. Among the statistics that were calculated from the data alone, it seems the anomalous correlation coefficient is the most informative; it shows dramatic increases from data set '30' through *A–D*, matching well the true level of anomalous signal. Even this statistic, though, does not accurately capture the deleterious nature of including data sets *E*, *F* and *G*. These results emphasize the important point that more data are not always better, but this may not be obvious from the statistics.

4.2. Substructure olution success is highly dependent on the resolution cutoff chosen

Our analyses have shown that for data sets with increasing redundancies, the anomalous correlation coefficients are consistently higher in lower resolution bins compared with the higher resolution bins (Fig. 2). The greater anomalous signal in the lower resolution bins is directly reflected in the ability to locate the S atoms using lower resolution cutoffs. For data sets '90' to *A–G*, the correlation coefficients and the number of correct solutions increase with decreasing resolution for each data set (Fig. 4*a* and Table 5). Furthermore, the distinction in the correlation-coefficient values between the correct and incorrect solutions is much more pronounced using the 3.5 Å cutoff compared with using the 2.0 Å cutoff (not shown). Our data analyses show that anomalous data from as few as 60 images are sufficient to solve the structure if the search is performed using the low resolution of 3.5 Å (Table 5).

In fact, when the data are truncated to unprecedented lower resolutions (up to 6.0 Å), the sulfur positions are still located (Fig. 5). Not surprisingly, the correlation coefficients of the solutions mirror the anomalous correlation coefficients by becoming even higher as resolution becomes lower (Figs. 2*b* and 5). However, the success rate does not simply increase with decreasing resolution, but is maximal using the 4.0 Å cutoff and decreases with increasing or decreasing resolution within data set *A–D* (Fig. 5). This clearly shows that there exists an optimal resolution cutoff at which the likelihood of structure solution is maximized.

These results contradict a widespread expectation, recently stated by Nagem *et al.* (2005), that for data with low theoretical anomalous signal successful substructure solution using S-SAD 'requires high or even atomic resolution data'. This expectation may be a carry-over from the legendary 1.5 Å resolution S-SAD phasing of crambin (Hendrickson & Teeter, 1981) and the fact that the requirement of very high resolution data is commonly associated with the use of direct methods (Debreczeni, Girmann *et al.*, 2003; Ramagopal *et al.*, 2003). The power of low-resolution data for S-SAD is actually not a new observation, but was first pointed out by Liu *et al.* (2000), who solved their substructure at 4.5 Å resolution and pointed out that for solving a substructure of isolated S-atom scatterers, 4.5 Å does correspond to atomic resolution. Also, Schneider & Sheldrick (2002) promoted the concept of only using resolution ranges having anomalous correlation coefficients greater that 30% (Dauter & Nagem, 2002; Debreczeni, Bunkoczi, Girmann *et al.*, 2003), because including higher

resolution data that have poor anomalous signal will only add noise and thus hinder substructure solution. Finally, Dauter & Nagem (2002) were surprised to note from their study that 'the solution of the partial structure of anomalous scatterers can be obtained at high and as well as low resolution, although at low resolution the chance of success is higher'. Our work emphasizes this point that not only might substructure solution attempts at resolutions lower than the limit of data collected succeed, they may even generally be better. The only real downside of lower resolution substructure solution is the possible inability to resolve the two sulfurs of a disulfide at resolutions lower than ~2.5 Å.

The result of better substructure solutions using low-resolution cutoffs raises the intriguing question of whether previous studies that reported the failure or limited success of S-SAD (Lemke *et al.*, 2002; Usón *et al.*, 2003) would have had better success if lower resolution cutoffs were used. In the work of Lemke *et al.* (2002) on *Escherichia coli* arginino-succinate synthetase (EAS; a 51 kDa protein with 19 S atoms and 0.86% theoretical anomalous signal) substructure solution barely succeeded (a success rate of 0.17%) using a 2.0 Å resolution cutoff. Based on extrapolation from our analyses, it is quite possible the success rate would have been much higher with a lower resolution cutoff. In the work of Usón *et al.* (2003) on R11, substructure-solution attempts failed using a resolution cutoff of 2.5 Å and no other cutoffs were tried. However, the 30% anomalous correlation coefficient cutoff for that data set was at 3.3 Å resolution, making it quite plausible that the data were sufficient to drive structure solution if a lower resolution cutoff had been used.

4.3. Strategy for in-house sulfur SAD structure solution

The results of our analyses allow us to propose a concrete strategy for data collection, assessment of the anomalous signal and substructure solution for S-SAD using the in-house Cu $K\alpha$ X-ray source. The approach would be to start collecting data on a single crystal, monitoring statistics that are indicators of general reduced data quality (I/σ, $R_{\mathrm{mrgd-}F}$, $R_{\mathrm{p.i.m.}}$; Diederichs & Karplus, 1997; Weiss, 2001), as well as monitoring the anomalous signal *via* the anomalous correlation coefficients. Data collection should be continued as long as these parameters improve with increasing redundancy. When these parameters reach a plateau, data collection on a fresh crystal should be started in order to avoid decreasing the data quality. This strategy is an important adjustment from the strategy proposed by Debreczeni, Girmann, Ma *et al.* (2003), where it was suggested 'simply to continue to collect data, increasing the redundancy, until the structure can be solved'.

Since the anomalous correlation coefficient is one of the most useful indicators for determining the amount of anomalous signal (Fig. 2; Evans, 2006; Schneider & Sheldrick, 2002; Zwart, 2005), it should used to monitor the increase or decrease in the anomalous signal, which in turn can be used to guide the crucial decision of resolution cutoffs for substructure solution. For the original ToxA substructure solution, the data were truncated at a resolution (3.0 Å) where the anomalous

correlation was ~20%. This is slightly lower than the range of 25–30% recommended by Schneider & Sheldrick (2002). While this range can be used as a guide for determining resolution cutoffs for initial substructure solution attempts, it should not be used as a rule. Lower resolution cutoffs, all the way up to 6.0 Å, should be tried for locating sulfur positions; then, if need be, the identified substructure can be used with higher resolution data for phasing and density modification.

Given the slow nature of in-house data collection, it makes sense to attempt structure solution even while data collection is in progress. In hindsight, the ToxA structure could be solved using only data from 60 images, but we did not try until much later. We suspect merging data from fresh isomorphous crystals should allow data quality to be improved even for crystals that decay rapidly (Diederichs & Karplus, 1997).

The analyses presented here underscore suggestions made by others that in-house Cu $K\alpha$ S-SAD is a powerful technique (Liu *et al.*, 2000; Dauter *et al.*, 1999; Dauter, 2002; Dauter & Nagem, 2002) and provides evidence that the mindset of requiring high-resolution data has led to an underappreciation of just how powerful it is. Taking into account the natural presence of S atoms in proteins (Ramagopal *et al.*, 2003; Nagem *et al.*, 2005), it should find wide applicability in high-throughput structure determination.

References

Blundell, T. L., Jhoti, H. & Abell, C. (2002). *Nature Rev. Drug Discov.* **1**, 45–54.
Blundell, T. L. & Patel, S. (2004). *Curr. Opin. Pharmacol.* **4**, 490–496.
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.
Brunzelle, J. S., Shafaee, P., Yang, X., Weigand, S., Ren, Z. & Anderson, W. F. (2003). *Acta Cryst.* D**59**, 1138–1144.
Burmeister, W. P. (2000). *Acta Cryst.* D**56**, 328–341.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. D. & Zhang, K. Y. (1999). *Prog. Biophys. Mol. Biol.* **72**, 245–270.
Creagh, D. C. (2004). In *International Tables for Crystallography*, Vol. C, edited by A. J. C. Wilson & E. Prince. Dordrecht: Kluwer Academic Publishers.
Cromer, D. T. & Liberman, D. A. (1970). *J. Chem. Phys.* **53**, 1891–1898.

research papers

Cromer, D. T. & Liberman, D. A. (1981). *Acta Cryst.* A**37**, 267–268.

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.* **289**, 83–92.

Dauter, Z., Dauter, M. & Dodson, E. (2002). *Acta Cryst.* D**58**, 494–506.

Dauter, Z. & Nagem, R. A. P. (2002). *Z. Kristallogr.* **217**, 694–702.

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.

Debreczeni, J. É., Girmann, B., Zeeck, A., Kratzner, R. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 2125–2132.

Diederichs, K. (2006). *Acta Cryst.* D**62**, 96–101.

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

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

Djinovic Carugo, K., Helliwell, J. R., Stuhrmann, H. & Weiss, M. S. (2005). *J. Synchrotron Rad.* **12**, 410–419.

Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.

Gordon, E. J., Leonard, G. A., McSweeney, S. & Zagalsky, P. F. (2001). *Acta Cryst.* D**57**, 1230–1237.

Heinemann, U., Illing, G. & Oschkinat, H. (2001). *Curr. Opin. Biotechnol.* **12**, 348–354.

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

Kitago, Y., Watanabe, N. & Tanaka, I. (2005). *Acta Cryst.* D**61**, 1013–1021.

Kuhn, P., Wilson, K., Patch, M. G. & Stevens, R. C. (2002). *Curr. Opin. Chem. Biol.* **6**, 704–710.

Lartigue, A., Gruez, A., Briand, L., Blon, F., Bezirard, V., Walsh, M., Pernollet, J. C., Tegoni, M. & Cambillau, C. (2004). *J. Biol. Chem.* **279**, 4459–4464.

Lemke, C. T., Smith, G. D. & Howell, P. L. (2002). *Acta Cryst.* D**58**, 2096–2101.

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.

Loll, P. J. (2003). *J. Struct. Biol.* **142**, 144–153.

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

Mueller-Dieckmann, C., Panjikar, S., Tucker, P. A. & Weiss, M. S. (2005). *Acta Cryst.* D**61**, 1263–1272.

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

Nagem, R. A. P., Ambrosio, A. L. B., Rojas, A. L., Navarro, M. V. A. S., Golubev, A. M., Garratt, R. C. & Polikarpov, I. (2005). *Acta Cryst.* D**61**, 1022–1030.

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

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*, 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.

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

Pusey, M. L., Liu, Z. J., Tempel, W., Praissman, J., Lin, D., Wang, B.-C., Gavira, J. A. & Ng, J. D. (2005). *Prog. Biophys. Mol. Biol.* **88**, 359–386.

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

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

Ravelli, R. B. G., Theveneau, P., McSweeney, S. & Caffery, M. (2002). *J. Synchrotron Rad.* **9**, 355–360.

Roeser, D., Dickmanns, A., Gasow, K. & Rudolph, M. G. (2005). *Acta Cryst.* D**61**, 1057–1066.

Rupp, B. (2003). *Acc. Chem. Res.* **36**, 173–181.

Sarma, G. N., Manning, V. A., Ciuffetti, L. M. & Karplus, P. A. (2005). *Plant Cell*, **17**, 3190–3202.

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

Sharff, A. & Jhoti, H. (2003). *Curr. Opin. Chem. Biol.* **7**, 340–345.

Stevenson, C. E., Tanner, A., Bowater, L., Bornemann, S. & Lawson, D. M. (2004). *Acta Cryst.* D**60**, 2403–2406.

Tuori, R. P., Wolpert, T. J. & Ciuffetti, L. M. (1995). *Mol. Plant-Microbe Interact.* **8**, 41–48.

Usón, I., Schmidt, B., von Bulow, R., Grimme, S., von Figura, K., Dauter, M., Rajashankar, K. R., Dauter, Z. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 57–66.

Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.

Watanabe, N., Kitago, Y., Tanaka, I., Wang, J., Gu, Y., Zheng, C. & Fan, H. (2005). *Acta Cryst.* D**61**, 1533–1540.

Weik, M., Ravelli, R. B., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J. & Sussman, J. L. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 623–628.

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., Panjikar, S., Mueller-Dieckmann, C. & Tucker, P. A. (2005). *J. Synchrotron Rad.* **12**, 304–309.

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.

Yang, C. & Pflugrath, J. W. (2001). *Acta Cryst.* D**57**, 1480–1490.

Yang, C., Pflugrath, J. W., Courville, D. A., Stence, C. N. & Ferrara, J. D. (2003). *Acta Cryst.* D**59**, 1943–1957.

Zwart, P. H. (2005). *Acta Cryst.* D**61**, 1437–1448.