

Phasing in the presence of radiation damage

Raimond B. G. Ravelli,^{a*} Max H. Nanao,^a Andy Lovering,^b Scott White^b
and Sean McSweeney^{c*}

^aEMBL, 6 rue Jules Horowitz, BP 181, 38042 Grenoble CEDEX 9, France, ^bSchool of Biosciences, University of Birmingham, UK, and ^cESRF, 6 rue Jules Horowitz, BP 220, 38043 Grenoble CEDEX 9, France. E-mail: ravelli@embl-grenoble.fr, seanmcs@esrf.fr

In the accurate estimation of small signals, redundancy of observations is often seen as an essential tool for the experimenter. This is particularly true during macromolecular structure determination by single-wavelength anomalous dispersion (SAD), where the exploitable signal can be less than a few percent. At the most intense undulator synchrotron beamlines, the effect of radiation damage can be such that all usable signal is obscured. Here the magnitude of this effect in experiments performed at the Se *K*-edge is quantified. Six successive data sets were collected on the same crystal, interspersed with two exposures to the X-ray beam during which data were not collected. It is shown that the very first data set has excellent phasing statistics, whereas these statistics degrade for the later data sets. Merging several data sets into one, highly redundant, data set only gave moderate improvements as a result of the presence of radiation damage. Part of the damage could be corrected for using a linear interpolation scheme. Interpolation of the data to a low-dose as well as to a high-dose data set allowed us to combine the SAD method with the radiation-damage induced phasing (RIP) technique, which further improved the experimental phases, especially after density modification. Some recommendations are given on how to mitigate the effect of radiation damage during structure determination.

Keywords: radiation damage; data redundancy; radiation-damage induced phasing (RIP); multiwavelength anomalous dispersion (MAD); single-wavelength anomalous dispersion (SAD).

1. Introduction

The availability of rapidly tuneable synchrotron radiation beamlines has brought the technique of multiwavelength anomalous dispersion (MAD) to the forefront of techniques available for experimental phase determination. The MAD method requires the collection of complete data sets at several wavelengths (often three) from, ideally, the same crystal. A distinct advantage of the MAD method over the traditional alternative, multiple isomorphous replacement (MIR), is that of (near) perfect isomorphism. The requirement of multiple data sets at wavelengths that are often very close together poses many experimental and technological challenges. However, as a result of work in the past decade, these problems have been overcome to such an extent that MAD has become the method of choice for *de novo* structure determination. Selenium is now the heavy atom of choice, thanks to elegant protocols that allow the substitution of methionine with selenomethionine (Doublié, 1997; Hendrickson *et al.*, 1990; Van Duyne *et al.*, 1993). The availability and increasing ease of MAD experiments has stimu-

lated a surge in the presence of software designed to exploit the possibilities of the technique.

Successful completion of MAD experiments requires careful monitoring and control of the sources of experimental errors. This need comes about because the MAD signal that is used to determine phases tends to be much smaller than that available in MIR. The signal available is typically governed by the kind of absorption edge of the heavy atom of interest rather than the atomic number of the atom. In general, the MAD signal will be about one-tenth that of the MIR signal, and only in exceptional cases, such as data collection at the *M* absorption edges of U (Liu *et al.*, 2001), will the signal be comparable with or greater than that of MIR. Therefore, careful experimental design and data collection is a prerequisite for a successful experiment. Such complications have led to renewed interest in the related technique of single-wavelength anomalous dispersion (SAD). In a SAD experiment one must employ additional information, typically available from density modification and/or non-crystallographic symmetry, to help overcome the phase ambiguity inherent to the SAD technique. With the exception of

experiments utilizing the anomalous signal from sulfur and phosphorous, most experimental phasing using SAD has been undertaken at the absorption edge of the anomalous scatterer, since data collection at these wavelengths maximizes the anomalous signal within the data set. This choice, whilst logical, is not without difficulties, in particular those due to the inherent increase of X-ray absorption cross section around the absorption edge, which limits the lifetime of the crystal (Murray *et al.*, 2004) and enhances the background of the diffraction image.

It has long been recognized that exposure to X-rays has a deleterious effect on the quality of diffraction data obtained from macromolecular crystal samples. The revolution brought about by the adoption of cryogenic cooling of samples has caused the misconception that such cooling confers 'immortality' on the crystal sample. Studies at high-energy third-generation synchrotron radiation sources have demonstrated that, at the photon fluxes available, crystals suffer rapidly from appreciable radiation damage (Perrakis *et al.*, 1999). The conventional thought was that this damage would result in a gradual deterioration of diffraction resolution by disruption of the crystalline lattice, thus affecting the detail of the structure of the macromolecule studied, though not the actual structure (Nave, 1995). Later studies have shown that in fact highly specific changes are induced within the sample, even before significant deterioration of diffraction quality is observed (Burmeister, 2000; Ravelli & McSweeney, 2000; Weik *et al.*, 2000). These systematic studies have demonstrated that specific, cumulative, conformational changes are induced that are believed to be linked to local redox potentials and the X-ray absorption cross section.

Rice *et al.* (2000) have investigated a series of MAD experiments in terms of radiation damage. These authors re-analysed the phasing of seven structures, all determined using anomalous scattering collected on crystals of selenomethionine-labelled proteins. All structures could be solved using the data of the peak data set alone (SAD). In many cases, MAD gave somewhat better correlations with the final σ_a -weighted $2F^o - F^c$ maps, both before and after solvent flattening. However, in some cases it was not possible to solve the structure through MAD, whereas SAD gave a solution. This situation was attributed to radiation damage, although only a first step was made towards a comprehensive study to assess the resulting effects on SAD and MAD experiments.

While our work has focused on phasing and radiation damage in the absence of anomalous scattering (Ravelli *et al.*, 2003), a number of other comprehensive studies have been presented that address radiation damage occurring during MAD or SAD experiments (Ennifar *et al.*, 2002; Evans *et al.*, 2003; Schiltz *et al.*, 2004; Weiss *et al.*, 2004; Zwart *et al.*, 2004). The heavy atoms used in these studies were bromine ($\times 2$), iodine ($\times 2$) and sulfur, respectively.

In the work described here, our aim is to expand on the report of Rice *et al.* (2000) by analysing the effect of radiation damage on phasing. To this end, we use the anomalous scattering from the most commonly used heavy atom nowadays: selenium. We have performed a controlled set of experiments

at the peak wavelength of the *K*-absorption edge of selenium and investigated the correlation between anomalous signal variation, phasing power and local structural changes. We examine the viability of the anomalous scattering experiment with respect to a number of criteria: in terms of peak heights in the Harker sections of the anomalous difference Patterson syntheses, the efficiency of determination of the anomalous scattering substructure, the quality of the phases produced and the effect on local atomistic errors. We discuss possible causes of the observed susceptibility of the anomalous scatterers, and we evaluate the widely heard call for redundancy in SAD experiments (Usón *et al.*, 2003), as well as the possible advantages of zero-dose extrapolation (Diederichs *et al.*, 2003).

2. Methods

As a test sample we used crystals of the flavoprotein nitroreductase from *Escherichia coli*, a protein containing 217 amino acid, which crystallizes in the space groups $P2_1$, $P2_12_12_1$ and $P4_12_12$ at 291 K, in the presence of 10% polyethylene glycol 4000, 25% ethylene glycol, 15 mM nicotinic acid and 100 mM sodium acetate (pH 4.6) (Lovering *et al.*, 2001). We selected a tetragonal crystal of external dimensions of about 0.3 mm \times 0.3 mm \times 0.1 mm. An energy scan around the selenium *K*-edge was performed on the crystal prior to data collection, using a highly attenuated beam. Data were collected on the peak of the white line of the fluorescence scan. Initial images were indexed with *DENZO* (Otwinowski & Minor, 1997) to confirm the space group. A strategy for data collection was determined to ensure greater than 95% completeness of individual data sets while treating Friedel-related reflections separately (Ravelli *et al.*, 1997). Whereas the crystal diffracted to high resolution (better than 2 Å), not all higher-order reflections could be collected, since we wanted to prevent possible overlap of reflections due to the long *c* axis of more than 260 Å. The data were collected in fine slices of 0.15° per frame and the attenuation was adjusted to ensure that data quality would be good to a reasonable resolution (2.0 Å), while preventing serious problems with overloads at low resolution. In order to prevent convolution of radiation-damage effects with systematic errors such as crystal absorption and detector non-uniformity, we collected several data sets over an identical total angular range of 48°. The data sets are labelled *A*, *B*, *C*, *D*, *E* and *F*. In between the data sets *D* and *E*, as well as in between data sets *E* and *F*, the rotating crystal was exposed to a single X-ray exposure (a 'burn') with an exposure time that equalled the total exposure time used for one entire data set (Fig. 1). The data sets were collected in Grenoble, France, on the European Synchrotron Radiation Facility (ESRF) undulator MAD beamline ID14-4 using an ADSC Q4R detector and an attenuated beam of about 3×10^{11} photons s⁻¹ through a 0.15 mm-diameter pinhole. The individual data sets were integrated and scaled separately using *DENZO* and *SCALEPACK* (option anomalous on). Both merged and unmerged data sets were produced.

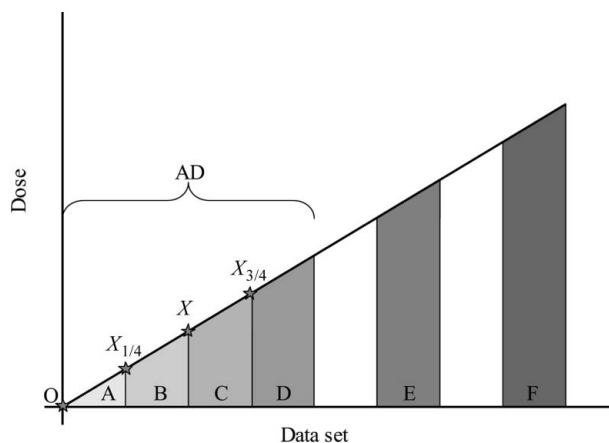


Figure 1

Data collection diagram. All data sets A–F were collected with a similar dose per data set. In between data sets D and E, as well as in between data sets E and F, the crystal received an X-ray dose equal to that used for each individual data set. Data sets A–D were merged into one data set, called AD. This highly redundant data set was used to extrapolate to zero dose (data set O) as well as to interpolate to doses X, $X_{1/4}$ and $X_{3/4}$.

Despite a careful selection of the attenuator and exposure times, a few overloads could still be observed for the initial data sets. Direct-methods-based programs such as *SHELXD* (Schneider & Sheldrick, 2002) and *SnB* (Weeks & Miller, 1999) can be very sensitive to a few missing low-resolution reflections, possibly resulting in better statistics for later data sets, by which time the crystal had lost enough scattering power to convert initially overloaded reflections into observable (non-overloaded) ones. By contrast, the spot size and mosaicity increased with X-ray dose, resulting in more (partly overlapped) rejected reflections, possibly having deleterious effects on the statistics for later data sets. In order to prevent these pitfalls, the common subset of reflections was identified and used to create data sets (A^{sub} , B^{sub} , C^{sub} , D^{sub} , E^{sub} and F^{sub}). The overall (anomalous) completeness of the common subset of reflections was 93.5% (93.2%).

A highly redundant data set was manufactured by merging data sets A, B, C and D, hereafter referred to as AD. In addition, a zero-dose extrapolated data set, O, was reconstructed from the redundant data set AD using an early *DENZO*-compatible version of the program presented by Diederichs *et al.* (2003).

The analysis of the data was performed in a two-tier process; each data set was first considered independently and then in comparison with the other data sets. During the analysis of each data set we applied the same criteria and ran the same tests. This mechanism allows one to identify specific statistical markers that may be useful in the identification of encroaching, but significant, radiation damage. For each data set the anomalous-difference Patterson maps were computed and compared with those calculated both on the absolute ($e^{-\text{Å}^{-3}}$) and relative (σ) scales. The anomalous-scatterer substructures were sought using *SHELXD* and *SOLVE* (Terwilliger, 2003). The quality of the results obtained with *SHELXD* were compared by examination of (i) the number of correct solutions per 500 trials, (ii) the best correlation coef-

Table 1

Theoretical diffraction ratios (%) for SeMet-substituted flavoprotein nitroreductase from *E. coli*.

	Inflection point	Peak	Remote
Inflection point	4.1	1.5	3.6
Peak		7.2	2.1
Remote			3.6

ficient and (iii) the best Patterson figure of merit. For *SOLVE*, the number of peaks found, their peak heights and the overall Z scores were compared. Protein phases were obtained using *MLPHARE*, *SOLVE*, *SHARP* (La Fortelle & Bricogne, 1997) and *SHELXE*, all using identical *SHELXD* sites as input. Density modification was carried out using *RESOLVE*, *SHELXE* and *DM*. Although the asymmetric unit of the tetragonal unit cell contains two copies of the protein, no NCS averaging was performed, in order that this system could be used as a general model for an average SeMet SAD experiment. The solvent content of the unit cell is low (32%), which, despite the good diffraction from these crystals and an average Met occurrence (Hendrickson *et al.*, 1990) in this protein (four methionines plus the amino-terminal methionine in 217 residues), increases the difficulty of successful SAD phasing.

The structure of nitroreductase was refined against all data sets using the program *REFMAC5* (Collaborative Computational Project, Number 4, 1994). Unfortunately, this program does not allow the refinement of occupancies of (heavy) atoms, so only estimates could be made, based on the inspection of σ_a -weighted $F^o - F^c$ maps around the Se atoms. Those estimated occupancies could be compared with the occupancies of the Se atoms based on the anomalous signals and as refined by the programs *SHELXD*, *MLPHARE* and *SOLVE*. $F^o - F^c$ difference Fourier maps were used to rank the different X-ray susceptible sites in the protein.

The dose per data set, as well as the dose for the X-ray burn, was estimated using the program *RADDOSE* (Murray *et al.*, 2004) assuming a uniform beam profile. The increase in absorption of Se around its K-edge was taken into account using a reference experimental fluorescence scan (Murray *et al.*, 2004), as well as the escape of Se K-edge fluorescence photons.

3. Results and discussion

In order to consider the results within the context of a MAD/SAD experiment it is useful to calculate the signal one might expect from an ideal experiment. Hendrickson introduced the concept of the diffraction ratios to describe the variation of anomalous and dispersive differences between data sets collected at different wavelengths. We calculate the diffraction ratios for nitroreductase according to the standard formulae (Hendrickson & Teeter, 1981). Values for f' and f''' were calculated by the program *CHOOCH* (Evans & Pettifer, 2001) from absorption-edge measurements made on the crystal. A remote wavelength of 0.939 Å (this is usual for Se MAD on ID14-4) was used for the calculation of the theoretical signal (Table 1). It can be seen that the exploitable differences are of

Table 2

Data collection statistics.

Six data sets, labelled *A* to *F*, were collected on a single crystal of nitroreductase. The crystal received an X-ray dose in between data sets *D* and *E*, and *E* and *F*.

	Resolution (Å) (high-res)	R_{sym} (%)	Mosaicity (°)	Wilson <i>B</i> factor (Å ²)	Unit-cell axis <i>a</i> (Å)	Unit-cell axis <i>c</i> (Å)
<i>A</i>	2.0 (2.07–2.0)	4.3 (5.9)	0.18	12.9	57.35	261.10
<i>B</i>	2.0 (2.07–2.0)	3.9 (5.7)	0.20	14.4	57.36	261.21
<i>C</i>	2.0 (2.07–2.0)	3.9 (6.3)	0.25	15.8	57.35	261.25
<i>D</i>	2.0 (2.07–2.0)	4.0 (7.9)	0.30	17.3	57.34	261.24
<i>E</i>	2.0 (2.07–2.0)	5.3 (13.2)	0.44	21.1	57.34	261.22
<i>F</i>	2.0 (2.07–2.0)	7.5 (21.3)	0.50	20.7	57.40	261.39

the order of a few percent; only for the anomalous signal at the peak wavelength is the signal much larger (7.2%) than the noise level one might expect within an average data set. This result clearly indicates why this wavelength is the favoured value for the collection of SAD data and why it was used in our study. The maximum dispersive signal (3.6%), between the remote and inflection point, is smaller than the anomalous signal available at the peak of the absorption edge.

3.1. Data quality

After data processing and scaling, the intensities were converted into structure factors using the CCP4 program *TRUNCATE* (Collaborative Computational Project, Number 4, 1994). Table 2 shows that X-ray radiation resulted in an increase of the Wilson-plot *B* factor as well as of the unit-cell volume. The unit-cell parameters changed anisotropically, causing general non-isomorphism between the data sets. The sample mosaicity, as refined with *SCALEPACK*, increases, indicating that the longer-range crystalline order is also compromised. The most significant jump in mosaicity and Wilson *B* factor is observed after the first 'burn': this will be further discussed in the next section. It has been previously observed that neither the Wilson *B* factor nor the mosaicity always increase proportionally with X-ray dose (Ravelli & McSweeney, 2000). The *R* factors for the data sets are all acceptable; for most purposes, *F* still represents a useful 2.0 Å data set.

The data sets were scaled together on a common scale using *SCALEIT* (Collaborative Computational Project, Number 4, 1994), allowing a more detailed comparison of the inter-data set variations. Table 3, which compares the merging statistics on structure-factor amplitudes ($|F|$), shows that significant differences have emerged between the respective data sets; the *R* factor between *D* and *A* is 11.7%. If data sets *A* and *D* had been part of a four-wavelength MAD experiment, it is clear that the dispersive signal would have been entirely swamped by radiation-damage effects. However, the anomalous signal within one, highly redundant, SAD data set could also be compromised by radiation damage, as is shown in the following paragraphs by comparing the individual data sets with combined ones.

Table 3Merging statistics on *F* (%) for SeMet-substituted nitroreductase.

Data set *AD* is obtained by combining data sets *A*, *B*, *C* and *D*. Data set *AD* was corrected for radiation damage through zero-dose extrapolation (data set *O*) as well as constant-dose interpolation (data set *X*). See also Fig. 1.

	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>AD</i>	<i>O</i>	<i>X</i>
<i>A</i>	4.9	8.7	11.7	18.5	22.0	6.9	2.5	6.5
<i>B</i>		4.4	7.6	15.0	19.2	2.9	6.3	2.4
<i>C</i>			3.8	11.5	16.4	2.2	10.2	2.4
<i>D</i>				8.5	14.0	4.9	13.3	5.4
<i>E</i>					7.8	12.1	19.7	12.6
<i>F</i>						16.2	23.0	16.7
<i>AD</i>							14.5	1.1
<i>O</i>								8.2

3.2. Heavy-atom substructure determination

Most SAD/MAD phasing programs explicitly determine the position of the anomalous scattering atoms prior to phasing the full macromolecular structure. Some programs make explicit use of the anomalous difference Patterson synthesis for the heavy-atom substructure determination. It is therefore interesting to investigate how the increasing dose absorbed by the sample affects both the anomalous difference Patterson synthesis and the possibility of determining the heavy-atom substructure.

Anomalous difference Patterson maps were calculated for each data set and compared for different Harker sections. Fig. 2 shows the $w = \frac{1}{4}$ section for the common subset of reflections of each data set, on an absolute scale. One observes a definite and progressive diminution of the quality of the Patterson synthesis as a function of dose; peaks decrease in height and the definition and separation of individual peaks is reduced for the series *A* to *D*, although the position of most of the peaks seems to be preserved. Contouring the same Harker sections from Fig. 2 on a relative scale shows a less drastic effect on the reduction in peak height, although the increased noise level is more readily apparent as well (figure not shown). The effects of the X-ray burn are dramatic; the Patterson synthesis for *E* clearly demonstrates that much of the signal has been erased. Furthermore, false new peaks start to appear on special positions in *F*.

The trend observed in the Harker sections correlate well with the success of the programs *SHELXD* and *SOLVE* in determining the heavy-atom substructure, as shown in Table 4. A clear dependence of the number of good solutions on absorbed dose is evident in *A–D* for the program *SHELXD*, and *SOLVE* also shows degraded statistics on going from *A* to *D*. As expected from Fig. 2, a large change is observed after the X-ray burns. No interpretable density maps are found for data sets *E* and *F*, although *SHELXD/E* still produces a non-random phase set for *E*.

In the absence of radiation damage, one would combine all the data sets into one highly redundant data set. In our case, a trade-off occurs between increased redundancy and increased radiation damage. It does not seem to be profitable to use data sets *E* and *F* at all. These data sets show very weak anomalous signal (Table 4 and Fig. 2); it appears as if the X-ray burns after

radiation damage

data set *D* and *E* have enhanced the radiation damage, disproportionately to the dose (Fig. 1). While the exact cause of this observation remains elusive, we can only speculate that it could be due to the increased mosaicity and spot shape which resulted in partially overlapping reflections, to the X-ray burn where the total dose was received over a much shorter time scale than during normal data collection, or to nonlinear effects in later stages of radiation damage (Teng & Moffat, 2000). Indeed, a very puzzling situation could occur if users collected data on a crystal that had previously been exposed on another beamline with a dose similar to that used for data sets *A–D*. As shown for data sets *E* and *F*, it could be possible to collect good data on such a crystal, while failing to observe significant anomalous signal. It is not exceptional that such a situation (*i.e.* good crystal, good data, poor anomalous signal) occurs on the beamline. A good record of previous experiments (as well as verification of the presence of the anomalous scatterers and the wavelength of the X-ray beam) could aid in finding an explanation for this problem.

A highly redundant data set (*AD*) was created by scaling the data sets *A*, *B*, *C* and *D* into one data set (Table 2 and Fig. 1). The degradation in the quality of the anomalous signal that was acquired during the data sets *B–D* is partially recovered by this combination, as the *SHELXD/E* statistics are similar in data sets *AD* and *A* (Table 4). The Harker section shows a somewhat cleaner map compared with the individual *A/B/C/D* data sets (Fig. 2). However, the peak heights are not as high as those observed for data set *A*. The phase errors after solvent flattening are somewhat lower for the combined data set *AD* compared with those obtained for data set *A*, especially from the programs *SOLVE/RESOLVE*. This fact seems to corroborate the widely heard call for redundancy in SAD experiments. However, we feel that the improvements of the merged data set *AD* over *A* are rather disappointing in view of the fourfold increase in time spent on data collection and what phase information could, theoretically, have been obtained from a MAD experiment taking the same amount of time. Thus the call for redundancy in SAD

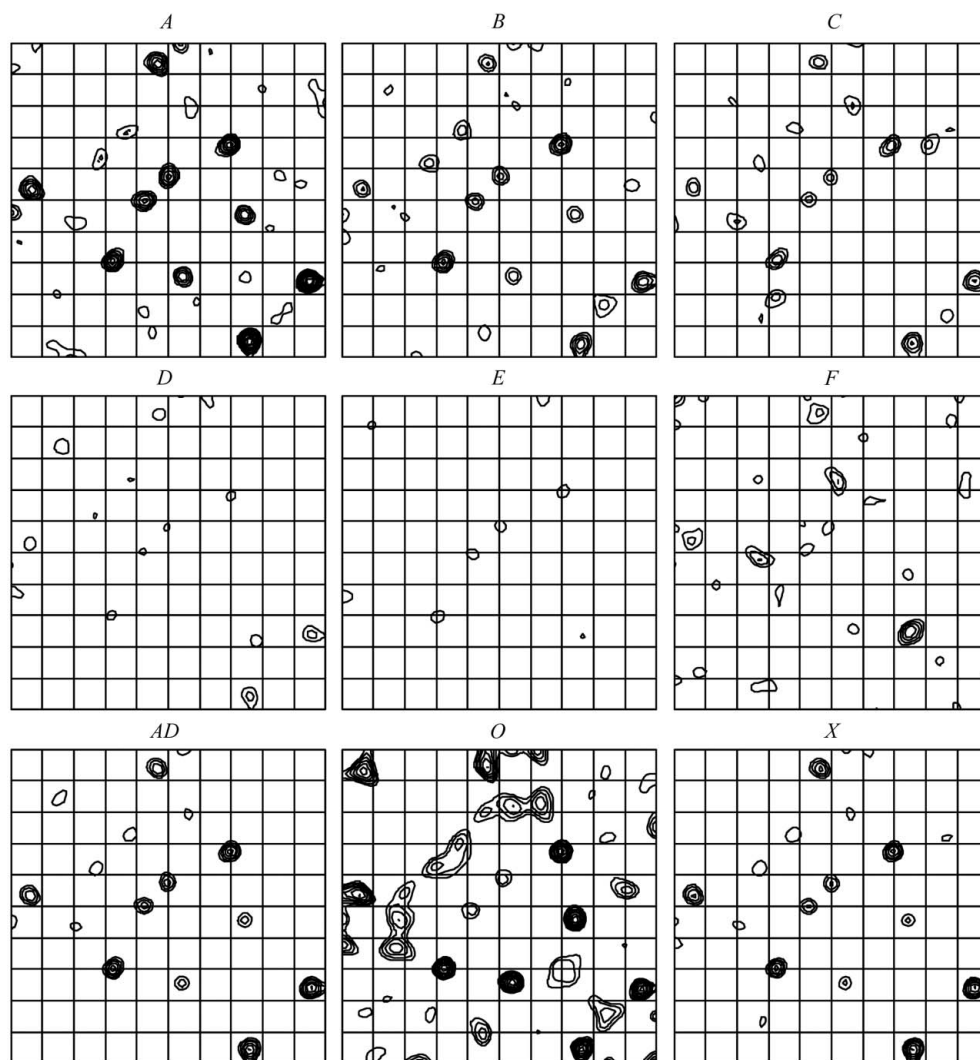


Figure 2

Harker sections $w = \frac{1}{4}$ anomalous difference Patterson, contoured on an absolute scale. Only those reflections that were measured for all the data sets *A–F* were used in the anomalous difference Patterson calculation.

Table 4
Substructure determination statistics.

	A^{sub}	B^{sub}	C^{sub}	D^{sub}	E^{sub}	F^{sub}	AD^{sub}	O^{sub}	X^{sub}
<i>SHELXD</i>									
No. correct†	262	244	233	239	0	0	257	244	277
Best cc	54.3/	51.6/	48.5/	43.5/	27.2/	9.0/	53.0/	53.7/	54.2/
All/weak	33.6	33.0	31.4	28.5	19.3	4.5	35.1	33.9	34.6
Best	18.2	16.8	15.8	14.1	6.6	1.4	18.9	17.8	19.8
PATFOM									
<i>SHELXE</i>									
Contrast	0.354	0.359	0.365	0.370	0.336	0.121	0.366	0.344	0.361
Connect.	0.925	0.925	0.921	0.918	0.901	0.834	0.922	0.923	0.924
Pseudo-freeCC (%)	65.6	65.5	64.5	66.4	62.4	31.0	66.0	65.3	67.7
wMPE‡ (°)	50.9	52.7	53.7	56.3	67.5	89.8	49.9	51.7	48.3
<hr/>									
	A^{sub}	B^{sub}	C^{sub}	D^{sub}	E^{sub}	F^{sub}	AD^{sub}	O^{sub}	X^{sub}
<i>SOLVE</i>									
No. of sites found (fom)	9	9	9	9	4	6	9	8	9
Overall Z score	47.1	39.0	36.9	31.2	4.4	15.5	43.9	27.5	41.8
wMPE (°)	61.3	61.6	63.5	65.8	89.6	89.8	61.1	61.7	60.3
<i>RESOLVE</i>									
wMPE (°)	53.5	51.2	56.1	61.3	89.5	89.3	48.1	56.4	46.0

† Out of 500 trials. Solution is marked 'correct' if $cc(\text{all}) > 40\%$. ‡ Weighted mean phase error as calculated using *PHISTATS* (Collaborative Computational Project, Number 4, 1994).

should be nuanced in the case where radiation damage becomes an issue.

3.3. Zero-dose extrapolation

Diederichs *et al.* (2003) have introduced a simple scheme to correct for radiation damage and have demonstrated its potential benefits for Se-SAD phasing. We applied a similar procedure for the *AD* data sets, and cross-compared the zero-dose extrapolated data set (called *O*) in the same way as for the other data sets (Tables 3–5 and Fig. 2). In addition to a zero-dose extrapolated data set, we reconstructed an interpolated data set, thus preventing possible extrapolation errors. The interpolation scheme used was linear and Friedel pairs were treated separately, although some reflections clearly seemed to be better described by quadratic or exponential models. So far, we have not been able to devise a general extrapolation scheme that proved to be highly robust for a wide range of data sets. Different schemes will give less divergent results for interpolated than for extrapolated data sets. The dose used for interpolation was half the dose the crystal received for the collection of the first four data sets (Fig. 1). The interpolated data set is called *X* and has been cross-compared with all other data sets (Tables 3–5 and Fig. 2).

The merging statistics in Table 3 clearly show the difference between sets *O* and *X*, where *O* gives the lowest-merging *R* factor with data set *A* and *X* is closer to *AD*. In absolute terms, the peak heights in the $w = \frac{1}{4}$ Harker section (Fig. 2) are comparable for data sets *O* and *A*, demonstrating the success of zero-dose extrapolation. However, data set *O* seems to

Table 5
Phasing statistics using sites of data set *A*.

	A^{sub}	B^{sub}	C^{sub}	D^{sub}	E^{sub}	F^{sub}	AD^{sub}	O^{sub}	X^{sub}
<i>SHELXE</i>									
Contrast	0.351	0.355	0.366	0.373	0.329	0.116	0.362	0.344	0.361
Connect.	0.922	0.923	0.918	0.919	0.897	0.827	0.921	0.924	0.923
Pseudo-freeCC (%)	65.7	65.3	63.8	64.9	61.9	30.4	66.5	65.5	66.8
wMPE (°)	51.1	52.5	54.8	57.4	68.2	78.0	49.9	51.2	48.5
<hr/>									
	A^{sub}	B^{sub}	C^{sub}	D^{sub}	E^{sub}	F^{sub}	AD^{sub}	O^{sub}	X^{sub}
<i>SOLVE</i>									
(fom)	0.39	0.39	0.37	0.34	0.20	0.06	0.43	0.38	0.44
Overall Z score	309.6	291.5	259.5	229.3	120.8	49.3	304.8	308.7	320.6
wMPE (°)	61.3	61.6	63.5	65.8	72.7	83.5	61.2	61.6	60.3
<i>RESOLVE</i>									
wMPE (°)	53.7	52.0	56.4	61.4	82.1	89.4	48.6	55.9	45.4
<hr/>									
	A^{sub}	B^{sub}	C^{sub}	D^{sub}	E^{sub}	F^{sub}	AD^{sub}	O^{sub}	X^{sub}
<i>MLPHARE</i>									
<i>R</i> _{cutis} (ano)	0.75	0.78	0.83	0.87	0.99	1.00	0.70	0.80	0.67
<i>DM</i>									
wMPE (°)	62.1	62.4	64.1	66.5	73.2	82.2	60.8	61.5	60.4
<hr/>									
	A^{sub}	B^{sub}	C^{sub}	D^{sub}	E^{sub}	F^{sub}	AD^{sub}	O^{sub}	X^{sub}
<i>SHARP</i>									
(fom) (acen)	0.52	0.48	0.48	0.45	0.37	0.10	0.53	0.49	0.58
Phasing power	2.16	1.86	1.89	1.64	1.12	0.25	2.24	1.83	2.87
wMPE (°)	59.2	59.2	60.6	63.4	70.8	89.4	57.9	57.9	58.4
<i>DM</i>									
wMPE (°)	54.2	53.5	56.0	57.8	71.4	89.5	50.8	53.0	51.3

suffer from over-extrapolation, as false peaks were introduced along the diagonal. In contrast, data sets *AD* and *X* give moderate absolute peak heights in the Harker section. In sigma levels, the difference between both *AD* and *X* versus *A* is less striking since the general noise level is lower for the redundant data sets.

The heavy-atom substructure determination is best using the interpolated data set *X* (Table 4). Most statistics in *SHELXD* and *SOLVE* are better for *X* than for *A* or *AD*. This difference is maintained during solvent flattening, where the largest improvement is observed while using *RESOLVE*; the weighted mean phase errors (wMPEs) between experimental and model phases are 53.5, 48.1 and 46.0° for data sets *A*, *AD* and *X*, respectively (Table 4).

3.4. SAD phasing

Protein phases were calculated using one set of *SHELXD* sites and the phasing programs *SHELXE*, *SOLVE* (mode analyze_solve), *SHARP*, *MLPHARE* and *SHELXE*. Density modification was performed using *SHELXE* (after *SHELXD*), *RESOLVE* (after *SOLVE*) and *DM* (after *MLPHARE* and *SHARP*). Table 5 gives overall phase errors before and after solvent flattening, using calculated model phases for a model that was refined against data set *A* as a

radiation damage

reference. Statistics were always calculated against the common subset of reflections, possibly compromising the absolute statistics due to the lower anomalous completeness (93.2%) of the common subset. The different programs have some important differences in the way in which initial phases are calculated, data and model errors are treated, and the density modification is carried out. Table 5 solely aims to compare the data sets, not the software that was used.

Surprisingly, data set *B* gives the smallest wMPE after *RESOLVE* and *DM*, whereas according to all other statistics, data set *A* gives the best results among the individual data sets. Data set *B* has slightly better statistics than data set *A* (Table 2), which possibly compensates for the reduced anomalous signal in the second data set. Data sets *C* and *D* give gradually worse statistics, as expected from Fig. 2. While starting with the correct sites, most programs could still produce non-random phase sets for data set *E*, despite the absence of strong peaks in the Harker sections. The corresponding maps are, however, not interpretable and did not improve after *RESOLVE* or *DM*.

The combined data set *AD* gives smaller phase errors than data set *A*. The applied zero-dose correction did not give better results, most likely as a result of over-fitting (see also §3.3). In contrast, interpolation (data set *X*) shows the best statistics. The improvement was most noticeable after *RESOLVE*, both in terms of phase errors and upon visual inspection of the map (Fig. 3).

3.5. Specific structural changes

The difference Fourier analysis between data sets *A* and *D* shows eight selenium atoms, at peak heights between 17 and 10σ , directly followed by peaks on $S\gamma$ for the single cysteine in each molecule (9σ), water molecules (8σ and less), carboxyl groups (7σ and less), and the N_Z of Lys179 and the O_{G1} of Thr184 (both 7.5σ). Whereas the Harker sections seem to indicate highly significant damage to the Se atoms, it is somewhat surprising to find such a small contrast between the Se atoms and other atoms in the difference Fourier map. We have investigated local structural differences between structures refined against data sets with different total absorbed dose. The Se atoms remain well defined in the σ_a -weighted $2F^o - F^c$ maps, even for data sets *E* and *F*, with an estimated average loss of occupancy of 20% rather than 100%.

How does the dramatic loss of anomalous signal with dose relate to the moderate susceptibility of the Se atoms as judged by the $2F^o - F^c$, $F^o - F^c$ and $F^o - F^o$ Fourier syntheses? Disulphide bonds are highly susceptible to X-ray damage and it was hypothesized that this fact is due to (partial) reduction (Weik *et al.*, 2002). X-ray radiation damage also results in the loss of definition of carboxyl groups. Sometimes the effect is only noticeable on the O atoms, whereas occasionally the carboxyl group and $C\beta$ in Asp or $C\gamma$ in Glu is lost as well. These effects are all referred to as 'decarboxylation' based on radiation chemistry (see Ravelli & McSweeney, 2000, and references therein) studies. Methionine is known to be susceptible to oxidative damage, although this has, to our

knowledge, so far not been observed by X-ray crystallography. Any cleavage of the Se— $C\gamma$ bond should show a correlated reduction of the electron density of the SeMet $C\delta$ atom, but we could not observe this in our study. Thus the underlying radiation chemistry responsible for the observed susceptibility of selenomethionine remains elusive to the authors.

In terms of absorption, the relative contribution of the Se atoms to the total protein absorption is very large. Fig. 4 shows the relative photoelectric cross sections of the C, N, O, S and Se atoms at 13.1 keV, expressed in barns per atom (1 barn =

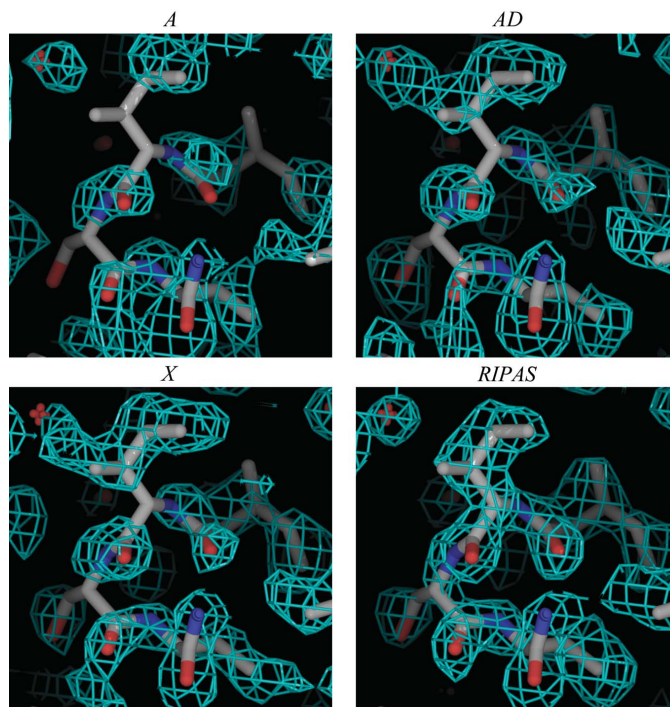


Figure 3 Part of the experimental electron density map of nitroreductase for *A*, *AD*, *X* and *RIPAS*, contoured at 1.3σ . The weighted mean phase errors compared with a model refined against data set *A* are 53.7, 48.6, 45.4 and 39.9°, respectively.

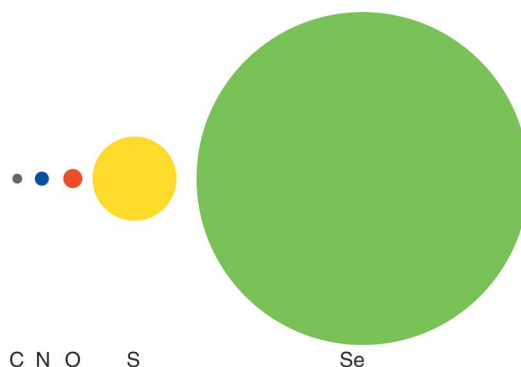


Figure 4 Relative photoelectric cross sections of carbon (grey), nitrogen (blue), oxygen (red), sulfur (yellow) and selenium (green) at 13.114 keV. The values are obtained through DABAX (http://www.esrf.fr/computing/scientific/dabax/tmp_file/FileDesc.html) and are 15.3, 31.8, 58.5, 1153.9 and 18120.0 barns per atom for C, N, O, S and Se, respectively.

10^{-24} cm²). The dose absorbed was of the order of 3×10^6 Gy per data set. The number of X-ray photons absorbed per unit cell per data set is estimated to be between 1 and 2; about 50% of them have been absorbed by the Se atoms (as calculated with *RADDOSE*). There are 80 Se atoms per unit cell; thus by the end of data set *D* a small but potentially observable percentage of Se atoms (< 5%) will have suffered from X-ray absorption.

XANES (X-ray absorption near-edge structure) and EXAFS (extended X-ray absorption fine structure) studies have shown photoreduction of heavy atoms due to the X-ray beam (Peisach *et al.*, 1982; Stroppolo *et al.*, 1998). A MAD experiment is normally preceded by an XANES scan, and the peak wavelength is chosen to give the maximum absorption and f'' . Photoreduction would change the appearance of the XANES scan (Stroppolo *et al.*, 1998) and thus modify the values of f' and f'' , which are traditionally thought to stay constant during the SAD experiment. For selenium, two oxidation states are known – oxidized and reduced – but as we started with the reduced state it is unclear how photoreduction could explain the differences we observe. In addition to photoreduction, the white line in the XANES scan can disappear upon radiation damage, indicating structural changes in the local environment of the heavy atom. Those changes are not necessarily directly observable by X-ray crystallography. In our case, all changes we observe around the SeMet residues are located on the Se atoms alone. Degradation of the white line could explain part of our observations.

An energy drift could also cause a reduction in anomalous signal. However, the energy is in general extremely stable on ID14-4 and it would be unlikely that such a putative drift would correlate exactly with the dose applied to the sample. Furthermore, later studies have shown that similar results are obtained at energies far remote from the edge.

3.6. The use of radiation damage to improve SAD phases

The theoretical anomalous signals for crystals of nitroreductase are shown in Table 1. The strongest signal is obtained for the peak, where the anomalous differences can lead to the structure determination as shown in this paper. However, powerful density modification techniques are required to overcome the phase ambiguity that is inherent to the SAD experiment. In our case, the solvent content of the nitroreductase crystals is only 32%, which limits the success of density-modification programs, resulting in relatively large phase errors (Tables 4 and 5) and poor electron density maps.

The combination of remote and inflection point data sets produces a weak dispersive signal (Table 1), which is nevertheless crucial for resolving the SAD phase ambiguity. Radiation damage could provide a similar benefit. An important difference between radiation damage and the dispersive signal is that the latter tends to be much more specific. The large number of weak radiation-damage sites will, in general, give poorer phases than carefully measured dispersive signals. In our case, phasing solely on the radiation-damage differences (no anomalous signal, no density modifi-

cation) between data sets *A* and *D* gave a phase error of 81° with *SHARP*.

The program *SOLVE* allows one to define two separate SAD data sets, and to subsequently combine the phases (option *combine*) while treating the second data set as a derivative of the first one. We tried this scheme using two interpolated data sets. The interpolation scheme was identical to that used to construct *O* and *X*, though now two data sets were created at 1/4 and 3/4 of the total dose used for the data sets *A*, *B*, *C* and *D*. The use of these two interpolated data sets (called $X_{1/4}$ and $X_{3/4}$, see Fig. 1) gave better results than, for example, using data sets *A* and *D*. The method is called RIPAS (Zwart *et al.*, 2004). After *SOLVE*, the RIPAS phases had a wMPE of 57.2° compared with the calculated model phases. This value improved to 39.9° after *RESOLVE*, which is a much lower wMPE than for any of the SAD data sets. A part of the corresponding electron density map is shown in Fig. 3.

4. Conclusions

One must sound several notes of caution for the experimenter undertaking an anomalous scattering experiment. The anomalous signal is highly susceptible to radiation damage; thus extreme care must be taken to account for the total absorbed dose experienced by the sample. It seems wisest to collect a phasing data set to a relatively moderate resolution using a dose that is only a fraction of that needed to give clear decay of the crystalline diffractive power. Radiation damage is especially detrimental for the MAD experiment, as dispersive signals are easily masked by general non-isomorphism introduced by the X-ray beam.

The usual prescription that increased redundancy equates to improved accuracy in the measurements is not necessarily true in the presence of radiation damage. In our case merging four data sets *A–D* resulted in the highly redundant data set *AD*, which was comparable with or only slightly better than the very first data set *A*. Merging *AD* with *E* and *F* would have led to an inferior overall signal-to-noise ratio.

While more and more beamlines are being automated and equipped with sample changers (Arzt *et al.*, 2005), the experimenter will face a new choice. Should one select the ‘best crystal’ on which to collect a MAD or highly redundant SAD data set, or should one examine the ‘top *N*’ crystals and collect complete low-dose low-redundancy SAD data sets in order to avoid radiation damage? The answer to this question cannot be answered yet, as it will depend on the variation of the amount of anomalous signal measured from different crystals of a particular protein *versus* the extra phase information one can extract from a carefully measured highly redundant data set. The answer also depends on the way ‘best crystal’ is defined. Currently, the ranking is mainly made on the basis of diffraction limit, spot shape and mosaicity (Deacon *et al.*, 2002), which does not necessarily relate to the size of measurable anomalous signal.

There are several ways to improve SAD phases if one has a highly redundant data set. Inspection of the correlation of anomalous signal between the different data sets will help to

evaluate the resolution to which good anomalous signal is available. Examination of the Harker sections in the anomalous difference Patterson maps will allow one to select and reject that part of the data where radiation damage has become too severe. Ideally, future zero-dose extrapolation schemes will reliably identify these parts and automate this process. Those algorithms could be used not only to create zero-dose data sets but also to reconstruct constant-dose data sets that are less susceptible to fitting errors (K. Diederichs, personal communication). Low- and high-dose data sets could be reconstructed and the specific radiation-damage induced structural differences between the corresponding two structures can aid in breaking the phase ambiguity of the SAD experiment prior to density modification. Unimodal phase distributions are normally seen as one of the major advantages of MAD. If successful, careful SAD phasing in the presence of radiation damage could also produce such distributions.

This work was supported by the FP6 EU BioXhit grant, under contract number LHS-G-CT-2003-503420.

References

- Arzt, S., Beteva, A., Cipriani, F., Delageniere, S., Felisaz, F., Forstner, G., Gordon, E., Launer, L., Lavault, B. & Leonard, G. (2005). *Prog. Biophys. Mol. Biol.* In the press.
- Burmeister, W. P. (2000). *Acta Cryst.* **D56**, 328–341.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Deacon, A., Brinen, L., Kuhn, P., McPhillips, S., McPhillips, T., Miller, M., van den Bedem, H., Wolf, G., Zhong, J. & Zhang, Z. (2002). *Acta Cryst.* **A58**, C299.
- Diederichs, K., McSweeney, S. & Ravelli, R. B. (2003). *Acta Cryst.* **D59**, 903–909.
- Doublet, S. (1997). *Methods Enzymol.* **276**, 523–530.
- Ennifar, E., Carpentier, P., Ferrer, J. L., Walter, P. & Dumas, P. (2002). *Acta Cryst.* **D58**, 1262–1268.
- Evans, G. & Pettifer, R. F. (2001). *J. Appl. Cryst.* **34**, 82–86.
- Evans, G., Polentarutti, M., Djinovic Carugo, K. & Bricogne, G. (2003). *Acta Cryst.* **D59**, 1429–1434.
- Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
- Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
- La Fortelle, E. de & Bricogne, G. (1997). *SHARP: A Maximum-Likelihood Heavy-Atom Parameter Refinement Program for the MIR and MAD Methods*. Orlando, FL: Academic Press.
- Liu, Y., Ogata, C. M. & Hendrickson, W. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10648–10653.
- Lovering, A. L., Hyde, E. I., Searle, P. F. & White, S. A. (2001). *J. Mol. Biol.* **309**, 203–213.
- Murray, J. W., Garman, E. F. & Ravelli, R. B. G. (2004). *J. Appl. Cryst.* **37**, 513–522.
- Nave, C. (1995). *Radiat. Phys. Chem.* **45**, 483–490.
- Otwinowski, Z. & Minor, W. (1997). *Methods in Enzymology*, Vol. 276, *Macromolecular Crystallography*, Part A, edited by C. W. Carter Jr & R. M. Sweet, pp. 307–326. New York: Academic Press.
- Peisach, J., Powers, L., Blumberg, W. E. & Chance, B. (1982). *Biophys. J.* **38**, 277–285.
- Perrakis, A., Cipriani, F., Castagna, J. C., Claustre, L., Burghammer, M., Riek, C. & Cusack, S. (1999). *Acta Cryst.* **D55**, 1765–1770.
- Ravelli, R. B. G., Leiros, H. K., Pan, B., Caffrey, M. & McSweeney, S. (2003). *Structure (Camb.)*, **11**, 217–224.
- Ravelli, R. B. G. & McSweeney, S. M. (2000). *Structure Fold. Des.* **8**, 315–328.
- Ravelli, R. B. G., Sweet, R. M., Skinner, J. M., Duisenberg, A. J. M. & Kroon, J. (1997). *J. Appl. Cryst.* **30**, 551–554.
- Rice, L. M., Earnest, T. N. & Brunger, A. T. (2000). *Acta Cryst.* **D56**, 1413–1420.
- Schiltz, M., Dumas, P., Ennifar, E., Flensburg, C., Paciorek, W., Vonrhein, C. & Bricogne, G. (2004). *Acta Cryst.* **D60**, 1024–1031.
- Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* **D58**, 1772–1779.
- Stroppolo, M. E., Nuzzo, S., Pesce, A., Rosano, C., Battistoni, A., Bolognesi, M., Mobilio, S. & Desideri, A. (1998). *Biochem. Biophys. Res. Commun.* **249**, 579–582.
- Teng, T. & Moffat, K. (2000). *J. Synchrotron Rad.* **7**, 313–317.
- Terwilliger, T. C. (2003). *Methods Enzymol.* **374**, 22–37.
- 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.* **D59**, 57–66.
- Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.
- Weeks, C. M. & Miller, R. (1999). *Acta Cryst.* **D55**, 492–500.
- Weik, M., Berges, J., Raves, M. L., Gros, P., McSweeney, S., Silman, I., Sussman, J. L., Houee-Levin, C. & Ravelli, R. B. (2002). *J. Synchrotron Rad.* **9**, 342–346.
- 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., Mander, G., Hedderich, R., Diederichs, K., Ermler, U. & Warkentin, E. (2004). *Acta Cryst.* **D60**, 686–695.
- Zwart, P. H., Banumathi, S., Dauter, M. & Dauter, Z. (2004). *Acta Cryst.* **D60**, 1958–1963.