

# A comparison of SAD and two-wavelength MAD phasing for radiation-damaged Se-MET crystals

Ana González

Stanford Synchrotron Radiation Laboratory, 2575 Sand Hill Road, MS 99, Menlo Park, CA 94025, USA. E-mail: ana@slac.stanford.edu

Received 15 May 2006  
Accepted 5 October 2006

Although a case has been made that single-wavelength anomalous dispersion (SAD) is the optimal strategy for data collection in the presence of radiation damage, two-wavelength MAD experiments at the inflection and a high-energy remote point of the absorption edge have been shown to be a potentially successful alternative method. In order to further investigate the performance of both data collection strategies, a comparison of SAD and MAD phasing was carried out for increasingly damaged data sets from three different selenomethionine protein samples collected under similar experimental conditions. In all but one example the MAD phases appeared to be less affected than SAD phases with increasing exposure to X-rays, and had a better overall success rate, indicating that this method should be given serious consideration when dealing with radiation-sensitive crystals. Simultaneous data collection in wedges at all wavelengths seems to be a very important factor in the success of MAD experiments; the decreased absorbed dose resulting from eschewing data collection at the maximum  $f''$  wavelength may play a less important role. Specific radiation damage to the selenium atoms is found to be a minor effect compared with the effect on the anomalous dispersion signal, although potentially large enough to be a useful contribution to phasing in both SAD and MAD experiments.

© 2007 International Union of Crystallography  
Printed in Singapore – all rights reserved

**Keywords:** SAD; MAD; radiation damage.

## 1. Introduction

When crystal samples are irradiated with X-rays, both global and specific damage to the sample take place. Global damage includes loss of resolution and expansion of the unit cell (Murray & Garman, 2002; Ravelli *et al.*, 2002); specific radiation damage typically involves reduction of atoms and eventually bond cleavage. For examples of specific radiation damage in proteins, see Ravelli & McSweeney (2000), Burmeister (2000), Weik *et al.* (2002) and Kort *et al.* (2004).

Both global and specific damage result in loss of isomorphism during data collection; this can prevent accurate measurement of the phasing signal in single- and multi-wavelength anomalous dispersion (SAD and MAD) experiments. MAD experiments in particular can be easily compromised if data at the different wavelengths are collected sequentially. In this case, sometimes better results are obtained by using only the data collected at the first wavelength (usually the maximum  $f''$  or 'peak' wavelength). This observation has led to the suggestion that a SAD experiment at the maximum  $f''$  wavelength is a better option in the presence of radiation damage (Rice *et al.*, 2000); two assumptions are made:

(i) SAD experiments are shorter than MAD experiments, therefore enough data to solve the structure can be collected before the crystal suffers significant radiation damage.

(ii) Anomalous differences are more important for the success of the experiment than dispersive differences, therefore it is best to concentrate on data collection at the maximum  $f''$  wavelength instead of measuring dispersive differences.

These assumptions are, however, not always sustained by experimental data. For example, González (2003a) demonstrated that two-wavelength MAD experiments are not necessarily longer (in terms of number of images required to solve the structure) and the dose absorbed during the experiment may therefore be the same or less than for SAD experiments. In addition, Peterson *et al.* (1996) and González *et al.* (1999) found that MAD phases are better when the dispersive differences, rather than the anomalous differences, are optimized. Based on these results, González (2003a) proposed a two-wavelength MAD experiment on the inflection and high-energy side of the absorption edge as a promising strategy to cope with radiation damage while obtaining high-quality phases. Simultaneous data collection at both wavelengths in several degree oscillation wedges can be used to preserve the dispersive differences in the event of loss

of isomorphism (Henderson, 1990; Smith, 1991). This two-wavelength MAD strategy resulted in the successful structure solution of a radiation-sensitive crystal structure (Vinculin) by Bakolitsa *et al.* (2004).

These experimental results suggest that the optimal data collection strategy for radiation-sensitive samples cannot be determined without controlled experiments comparing two-wavelength MAD and SAD phasing. Although Vinculin structure solution was also independently carried out by SAD (Borgon *et al.*, 2004), the crystal forms were not identical and the data were collected on different beamlines on different sources. Other comparisons between SAD and MAD phasing (González, 2003a) were made on undamaged (or not significantly damaged) crystals.

It is also worth investigating how the possibility of enhancing the quality of anomalous dispersion phases with the 'pseudodispersive' signal arising from specific damage to the anomalous scattering sites affects the choice of data collection strategy. Conceptually, this is an extension of the radiation-induced phasing (RIP) method proposed by Ravelli *et al.* (2003). Zwart *et al.* (2004) and Evans *et al.* (2003) have reported this approach for the case of *N*-iodosuccinimide and triiodide derivatives, respectively; Schiltz *et al.* (2004) for brominated RNA; Ramagopal *et al.* (2005) for mercury derivatives; and Ravelli *et al.* (2005) for seleno-methionine samples. If the radiation-induced signal is comparable in importance with the dispersive signal in MAD experiments, then a case could be made that data collection at a single wavelength would be a more efficient strategy for phasing structures *de novo*. This could be the case for samples in which the anomalous scatterer is cleaved very quickly (Schiltz *et al.*, 2004; Ramagopal *et al.*, 2005), but for seleno-methionine proteins it might be possible to solve the structure faster and more easily by MAD or SAD alone.

The experiments described in this paper aimed to study the performance of SAD and two-wavelength MAD experiments for seleno-methionine samples under the same (or as close as can be achieved) experimental conditions: back-to-back experiments using the same samples on the same beamlines. Quantification of the radiation damage during both types of experiment was attempted in order to investigate whether avoiding data collection at the peak wavelength has an impact on the results. To determine the potential of RIP phasing for seleno-methionine samples, the radiation-induced signal was compared with the anomalous and dispersive signal for increasing radiation dose. The results from these experiments and the implications for devising a data collection strategy will be discussed.

## 2. Methods

### 2.1. Sample preparation

Three different samples were used in this experiment. The hypothetical protein (abbreviated here as HypP) from *Agrobacterium tumefaciens* consists of a 104 residue monomer containing four methionine residues [see Protein Data Bank

(PDB ID 1VQS)]. The seleno-methionine substituted protein crystals were supplied by the Joint Center for Structural Genomics (JCSG) (Lesley *et al.*, 2002); the crystals were grown in a 2.8 M sodium acetate and 10% glycerol buffer. The crystals formed thin long plates with an approximate size of 30  $\mu\text{m}$   $\times$  10  $\mu\text{m}$   $\times$  200  $\mu\text{m}$ . The space group of the crystals used was  $P2_12_12_1$ ; eight copies of the monomer were present in the asymmetric unit; the solvent content was 52.8%.

The second sample was a putative glucosamine-fructose-6-phosphate aminotransferase (GF6PA) from *Salmonella typhimurium* (PDB ID 2A3N). The crystals were grown in 0.2 M sodium citrate, 20.0% PEG-3350, at pH 8.2; and supplied by JCSG. The seleno-methionine substituted protein contained 12 seleno-methionines in 355 residues and a solvent content of 37%. The crystals belonged to the space group  $C2$  and were approximately 100  $\mu\text{m}$   $\times$  50  $\mu\text{m}$   $\times$  100  $\mu\text{m}$  in size.

The third sample was the 184 residue seleno-methionine substituted chicken Vinculin tail (PDB ID 1QKR), containing nine seleno-methionine residues in the monomer. The crystallization conditions were 25% PEG-2000, 0.2 M ammonium sulfate and 0.1 M propanol at pH 5.0 (Bakolitsa *et al.*, 1999). The crystals used for the experiment had an approximate size of 30  $\mu\text{m}$   $\times$  10  $\mu\text{m}$   $\times$  200  $\mu\text{m}$ , and belonged to the space group  $P2_12_12_1$ , with two copies in the molecule in the asymmetric unit and a solvent content of 35%.

All the crystals used in the experiment were frozen and stored in liquid nitrogen prior to transfer to an Oxford Cryojet 100 K nitrogen cryostream for the data collection.

### 2.2. Data collection and processing

Several crystals from each sample were screened prior to the experiment using the automated sample-mounting robot developed at the Stanford Synchrotron Radiation Laboratory (SSRL) (Cohen *et al.*, 2002); the crystals were ranked and scored using the program *Web-Ice* (González, Moorhead *et al.*, 2005) to ensure consistent quality of crystals used for the experiment.

The experimental procedure was to collect a series of three consecutive MAD or SAD data sets (except for HypP for which only two data sets were collected in each series). The data sets in each series were designed *a*, *b* and *c*. All the data were collected at SSRL using the *Blu-Ice/DCS* software (McPhillips *et al.*, 2002). Auto-indexing and strategy calculation were carried out using *LABELIT* (Sauter *et al.*, 2004) and *MOSFLM* (Leslie, 1999; Collaborative Computational Project, Number 4, 1994). The two-wavelength MAD experiments used the absorption-edge inflection-point wavelength and a high remote wavelength chosen following the guidelines given by González (2003b). The two wavelengths were collected in 10° wedges. The SAD data sets were collected at the white-line wavelength on the Se absorption edge. For the HypP crystal the SAD data were collected in a single wedge; for GF6PA and Vinculin tail, the SAD data were collected in 10° wedges using the inverse beam setting.

The HypP data sets were collected on the SSRL beamline BL9-2; a long crystal was translated along the spindle axis

direction, allowing the SAD and MAD data sets to be collected from the same crystal. The SAD and MAD data sets were collected in alternation, starting with the SAD data sets. For all other samples, each data collection series was collected from individual single crystals. The GF6PA data were also collected on BL9-2, and the Vinculin tail data were collected on BL11-1; because of the higher intensity available at this beamline, data from four different crystals could be collected in the time allocated to the experiment.

The absorbed dose during the experiments was estimated using the program *RADDOSE* (Murray *et al.*, 2004, 2005). The size of the crystals was estimated visually; the anomalous scattering factors were calculated using *CHOOCH* (Evans & Pettifer, 2001), and the beam intensities were measured using a calibrated photodiode. The resultant absorbed dose for each sample is given in Table 1.

The data were integrated using *MOSFLM* (Leslie, 1999; Collaborative Computational Project, Number 4, 1994) and scaled using *SCALA* (Collaborative Computational Project, Number 4, 1994); structure factor amplitudes and the Wilson *B* factors were obtained using *TRUNCATE* (Collaborative Computational Project, Number 4, 1994); *SHELXC* and *SHELXD* (Schneider & Sheldrick, 2002) were used to locate the anomalous scatterer positions. The correct hand was determined using *SHELXE* (Sheldrick, 2002). Experimental phases were calculated using *SOLVE* (Terwilliger & Berendzen, 1999), and *RESOLVE* was used for density modification (Terwilliger, 2000) and model building (Terwilliger, 2003).

The main criteria used to determine the success of SAD and MAD phasing for each data set was the percentage of the structure traced automatically by *RESOLVE*. If no residues could be traced with confidence, the phasing was deemed to have failed. The correlation coefficient and *R*-factor after density modification were also considered as indicators of the quality of the resultant phases. Although substructure solution could be considered an integral part of anomalous phasing experiments, in practice it does not need to be determined from the same data used for phasing; therefore, whenever incorrect sites were found using *SHELXD*, phasing and model building was re-attempted supplying the correct substructure.

Macroscopic indicators of radiation damage such as unit-cell volume, mosaicity (calculated using *MOSFLM* during post-refinement), Wilson *B* factors, and maximum resolution of the diffraction [defined in terms of  $I/\sigma(I)$ ] were examined to try to determine the relative amount of damage inflicted during the SAD and MAD experiments. Specific damage to atoms in the samples over the data collection series was detected with difference Fourier maps [ $F_b - F_a \times \exp(i\varphi)$ ] and [ $F_c - F_a \times \exp(i\varphi)$ ] using the density modified phases  $\varphi$  derived from the first data set *a*. For MAD experiments, the structure factors derived from the inflection wavelength data were used. The resolution of the data used for the map calculation was truncated in order to avoid using reflections

**Table 1**

Estimated absorbed dose in Gy calculated for all the data collection series for each type of crystal.

Series	Time per image (s)	Number of images	$f''$ ( $e^-$ )	Data set		
				<i>a</i>	<i>b</i>	<i>c</i>
HypP SAD	30	180	6.8	$1.9 \times 10^7$	$3.8 \times 10^7$	–
HypP MAD				$1.5 \times 10^7$	$3.1 \times 10^7$	–
Remote	38.5	90	3.2	$0.8 \times 10^7$	$1.6 \times 10^7$	–
Inflection	30	90	4.2	$0.73 \times 10^7$	$1.5 \times 10^7$	–
GF6PA SAD	10	360	7.4	$1.0 \times 10^7$	$2.0 \times 10^7$	$3.0 \times 10^7$
GF6PA MAD				$0.75 \times 10^7$	$1.5 \times 10^7$	$2.7 \times 10^7$
Remote	10	180	3.3	$0.35 \times 10^7$	$0.7 \times 10^7$	$1.1 \times 10^7$
Inflection	10	180	4.1	$0.4 \times 10^7$	$0.8 \times 10^7$	$1.6 \times 10^7$
Vinc SAD	4	180	5.5	$1.0 \times 10^7$	$2.0 \times 10^7$	$3.0 \times 10^7$
Vinc MAD				$0.75 \times 10^7$	$1.5 \times 10^7$	$2.3 \times 10^7$
Remote	6	90	3.4	$0.45 \times 10^7$	$0.9 \times 10^7$	$1.4 \times 10^7$
Inflection	4	90	3.2	$0.3 \times 10^7$	$0.6 \times 10^7$	$0.9 \times 10^7$

present in the *a* data sets but absent or very weak in the more damaged *b* and *c* sets. Phased difference anomalous and dispersive maps were also calculated to compare the anomalous signal with the radiation-induced signal. The electron density maps were calculated using *FFT* (Collaborative Computational Project, Number 4, 1994) and inspected using *COOT* (Emsley & Cowtan, 2004).

### 3. Results

#### 3.1. SAD and MAD phasing with increasing radiation damage

Table 2 lists the results of MAD and SAD data processing and phasing using the *a*, *b* and *c* data sets for each of the samples.

All three structures could be solved either by SAD or MAD when using the *a* data set. For GF6PA the map correlation and the number of residues traced automatically is similar; this is also the case for HypP, although the MAD data were worse in terms of the resolution, *R*-merge and mosaicity. For Vinculin tail, the percentage of the structure traced varies from case to case and appears to depend more on the resolution of the data than the method used to solve the structure.

When using the *b* data set, the HypP structure could neither be solved by SAD nor by MAD. GF6PA could be solved both by SAD and MAD, but the SAD phasing results were significantly worse. Structure solution also failed for one of the Vinculin tail SAD data sets (labeled SAD-1 in Table 2*c*). For the other Vinculin tail MAD and SAD data sets the final results are comparable with those obtained for the *a* data set, although the quality of the phases before density modification (evidenced by the figure of merit) decreased in all cases. None of the structures could be solved by SAD using the *c* (most damaged) data set; one of the Vinculin tail cases (MAD-2 in Table 2*c*) and GF6PA could still be solved by MAD.

Because incorrect selenium sites were found by *SHELXD* for all the failed experiments, the phasing was repeated in these cases after supplying *SOLVE* with the correct substructure (*e.g.* the sites found using the *a* data set). A percentage of the chain could then be traced for some of the

**Table 2**

Data processing and structure solution statistics for the (a) HypP, (b) GF6PA and (c) Vinculin data collection series.

The resolution and  $R$ -sym are given for the data with an average  $I/\sigma I$  equal to or greater than 5; data to higher resolution were also processed and used for phasing when available. For the MAD data sets, the mosaicity and Wilson temperature factors are given for the inflection wavelength;  $R$ -sym is calculated after scaling (but not merging) the inflection and remote wavelengths data together. N.A. = not applicable.

(a)	SAD		MAD	
	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>
Unit-cell axes (Å)	82.44 114.86 117.98	83.02 115.17 118.75	82.56 114.94 118.24	82.73 115.13 118.50
Mosaicity (°)	0.37	0.32	0.40	0.50
5σ resolution (Å)	2.80	3.23	2.99	3.95
$R$ -sym	0.075	0.078	0.092	0.079
Multiplicity	7.0	7.0	7.0	7.0
Completeness (%)	99.8	99.8	99.9	99.9
$B$ factor (Å <sup>2</sup> )	42.2	58.1	44.8	57.0
Substructure solved	Yes	No	Yes	No
Figure of merit	0.32	0.16	0.38	0.31
(experimental phases)				
Correlation	0.84	0.51	0.84	0.57
$R$ -factor	0.22	0.35	0.23	0.34
Sites traced (%)	78	0	77	0
Sites traced with correct substructure† (%)	N.A.	53	N.A.	0

(b)	SAD			MAD		
	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>
Unit cell axes (Å)	81.21 102.64 49.32	81.51 102.67 49.42	82.58 102.56 49.46	81.47 102.66 49.34	81.94 102.78 49.54	82.12 102.62 49.62
$\beta$ (°)	123.19	123.29	123.28	123.25	123.43	123.57
Mosaicity (°)	0.58	0.79	1.02	1.01	1.14	1.24
5σ resolution (Å)	1.76	2.08	2.32	1.75	2.07	2.32
$R$ -sym	0.070	0.065	0.076	0.094	0.072	0.064
Multiplicity	7.4	7.4	7.2	7.3	7.2	7.2
Completeness (%)	99.8	99.9	99.9	99.9	100	100
$B$ factor (Å <sup>2</sup> )	14.5	20.8	27.4	14.3	20.5	25.2
Substructure solved	Yes	Yes	No	Yes	Yes	Yes
Figure of merit	0.49	0.29	0.16	0.49	0.31	0.21
(experimental phases)						
Correlation	0.80	0.60	0.38	0.83	0.71	0.53
$R$ -factor	0.27	0.31	0.35	0.22	0.24	0.24
Sites traced (%)	75	33	0	77	84	62

(c)	SAD-1			SAD-2		
	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>
Unit-cell axes (Å)	35.02 87.95 118.94	35.03 88.03 119.38	35.04 88.06 119.63	35.05 88.35 119.00	35.05 88.39 119.23	35.04 88.42 119.39
Mosaicity (°)	0.58	0.62	0.63	0.80	0.81	1.00
5σ resolution (Å)	2.87	3.39	3.79	2.58	2.83	3.16
$R$ -sym	0.080	0.079	0.078	0.064	0.067	0.066
Multiplicity	6.6	6.8	6.7	6.6	6.6	6.5
Completeness (%)	99.8	99.5	99.5	98.1	97.7	97.2
$B$ factor (Å <sup>2</sup> )	32.8	37.4	43.3	24.5	28.1	31.9
Substructure solved	Yes	No	No	Yes	Yes	No
Figure of merit	0.35	0.19	0.18	0.31	0.22	0.17
(experimental phases)						
Correlation	0.79	0.44	0.48	0.80	0.77	0.47
$R$ -factor	0.27	0.41	0.41	0.26	0.27	0.38
Sites traced (%)	56	0	0	76	76	0
Sites traced with correct substructure† (%)	N.A.	56	0	N.A.	N.A.	39

previously failed experiments, as shown in Tables 2(a) and 2(c). The HypP *b* MAD structure and the *c* SAD structures for GF6PA and Vinculin tail (SAD-1) remained unsolved.

### 3.2. Radiation damage during the experiments

Consistent increases in the unit-cell volume, Wilson  $B$  factor and a decrease of the maximum data resolution with the absorbed dose were observed for all samples. The mosaicity also shows a tendency to increase, although in a more irregular fashion: the increase is not linear between all the data sets, and for HypP it was found to decrease between the *a* and *b* data sets. The difference in unit-cell volume increases for GP6PA, and the Vinculin tail MAD and SAD experiment is within the variation observed for different crystals of the same protein by Murray & Garman (2002). The Wilson  $B$  factor correlates the closest (but not perfectly) with the calculated absorbed dose. The  $B$  increase is not very different for MAD and SAD experiments, although slightly more pronounced for SAD in most cases.

The peak heights of the phased difference Fourier maps about the selenium and sulfur sites for the SAD and MAD data collection series are shown in Figs. 1, 2 and Fig. 3.<sup>1</sup> These RIP peaks were typically lower for the MAD experiments, although there is a large variation in the decay rate between different sites, with a few sites showing more damage during the MAD data collection series. The average ratio of the site occupancy loss between SAD and MAD experiments is  $1.32 \pm 0.04$  for HypP;  $1.12 \pm 0.07$  and  $1.09 \pm 0.06$  for Vinculin tail (calculated using the *b* and *c* data, respectively), and  $1.12 \pm 0.03$  and  $1.25 \pm 0.02$  for GF6PA. The ratio between the absorbed dose during the SAD and MAD experiments is 1.23 for HypP, and 1.33 for Vinculin tail and GF6PA. No clear evidence of a rate-dose effect was observed in these experiments, as expected from the flux densities (less than  $5 \times 10^{12}$  photons  $s^{-1} mm^{-2}$ ) used during the collection (Müller *et al.*, 2002).

### 3.3. Comparison of radiation-induced signal and anomalous signal

Despite the specific site decay shown above, structure solution by RIP using the SAD data sets (data set *a* combined with either *b* or *c*) was not possible, even when using the correct sites calculated using the *a* data.

Another possible way to use the radiation-induced signal is to carry out MAD phasing using

<sup>1</sup> Specific radiation damage to other sites was also observed, particularly for the better diffracting GF6PA crystals, but was not examined in detail.

Table 2 (continued)

	MAD-1			MAD-2		
	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>
Unit-cell axes (Å)	35.07	35.08	35.10	35.05	35.05	35.04
	88.17	88.30	88.42	88.35	88.39	88.42
	118.68	118.95	119.22	119.00	119.23	119.39
Mosaicity (°)	0.24	0.29	0.33	0.80	0.81	1.00
5σ resolution (Å)	2.39	2.58	2.83	2.75	3.01	3.37
<i>R</i> -sym	0.085	0.081	0.078	0.086	0.081	0.077
Multiplicity	7.0	7.0	7.0	6.6	6.6	6.6
Completeness (%)	100	100	100	99.9	99.9	99.9
<i>B</i> factor (Å <sup>2</sup> )	21.4	24.6	29.0	28.0	31.9	36.2
Substructure solved	Yes	Yes	No	Yes	Yes	Yes
Figure of merit (experimental phases)	0.44	0.35	0.29	0.38	0.31	0.26
Correlation	0.83	0.79	0.43	0.76	0.71	0.70
<i>R</i> -factor	0.24	0.25	0.39	0.26	0.26	0.27
Sites traced (%)	82	79	0	69	73	60
Sites traced with correct substructure† (%)	N.A.	N.A.	76	N.A.	N.A.	N.A.

† This is only relevant in the cases where the model could not be traced because the anomalous substructure could not be solved with the input data.

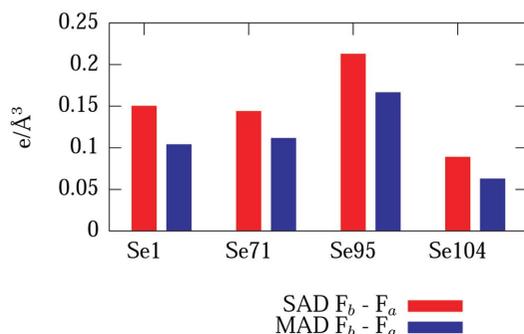


Figure 1

Peak height in difference maps  $F_b - F_a$  for SAD and MAD experiments for HypP. The maps were calculated to 3 Å resolution, and the peak heights for sites in the asymmetric unit related by non-crystallographic symmetry were averaged.

the data collected at the *a* data set remote wavelength in conjunction with the inflection-point wavelength data from the *b* or *c* data sets. This only worked for HypP and Vinculin tail, and only by supplying the correct sites.<sup>2</sup>

Fig. 4 shows the peak heights for the maximum radiation-induced signal (*i.e.* between the first and last data sets in the series) and the phased Fourier dispersive and anomalous differences for the same data sets. The anomalous and dispersive difference maps are less noisy and the sites appear at an overall higher contour level, even for the most damaged *c* data set. Only for Vinculin tail is the RIP signal at a similar contrast level as the anomalous differences for the *c* SAD data set; it is worth noting that SAD phasing did not work for this particular data set. The smaller contour level of the peaks suggest that the amount of site damage in these cases was not enough to provide a RIP signal comparable with the anomalous or dispersive signal, or it is not large enough to be easily measured from non-isomorphous data sets. This explains why the attempts to use the radiation-induced signal for phasing gave poor results compared with MAD (using data collected in wedges) and SAD.

alous or dispersive signal, or it is not large enough to be easily measured from non-isomorphous data sets. This explains why the attempts to use the radiation-induced signal for phasing gave poor results compared with MAD (using data collected in wedges) and SAD.

#### 4. Discussion

The premise that the data used for MAD and SAD phasing were of similar quality could be evaluated by comparing quantities such as resolution, *R*-merge,  $I/\sigma(I)$  and completeness for the *a* data sets, for which the absorbed dose is the least. In this respect it can be concluded that the comparison results are very reliable for the GF6PA crystals, and are reasonably adequate for Vinculin tail. A larger difference was observed for HypP, despite collecting all the data from the same crystal; both *a* and *b* SAD data sets were better than the corre-

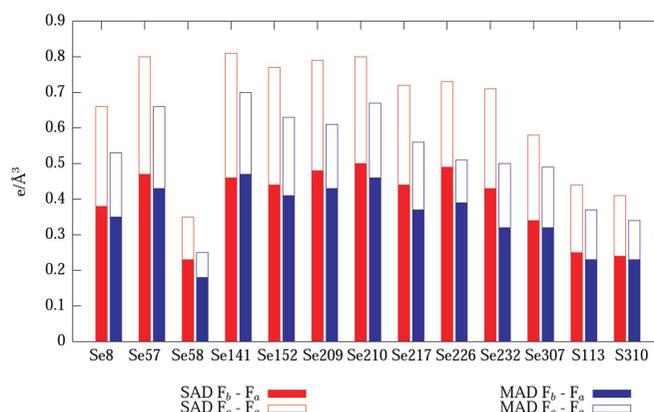


Figure 2

Peak height in difference maps  $F_b - F_a$  and  $F_c - F_a$  for SAD and MAD data collection series for GF6PA. The maps were calculated to 2.3 Å resolution.

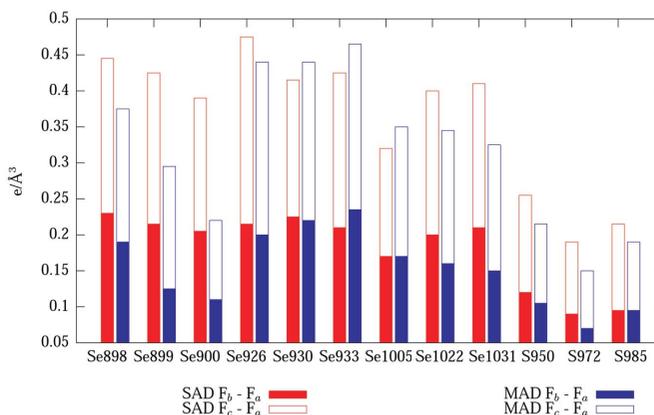
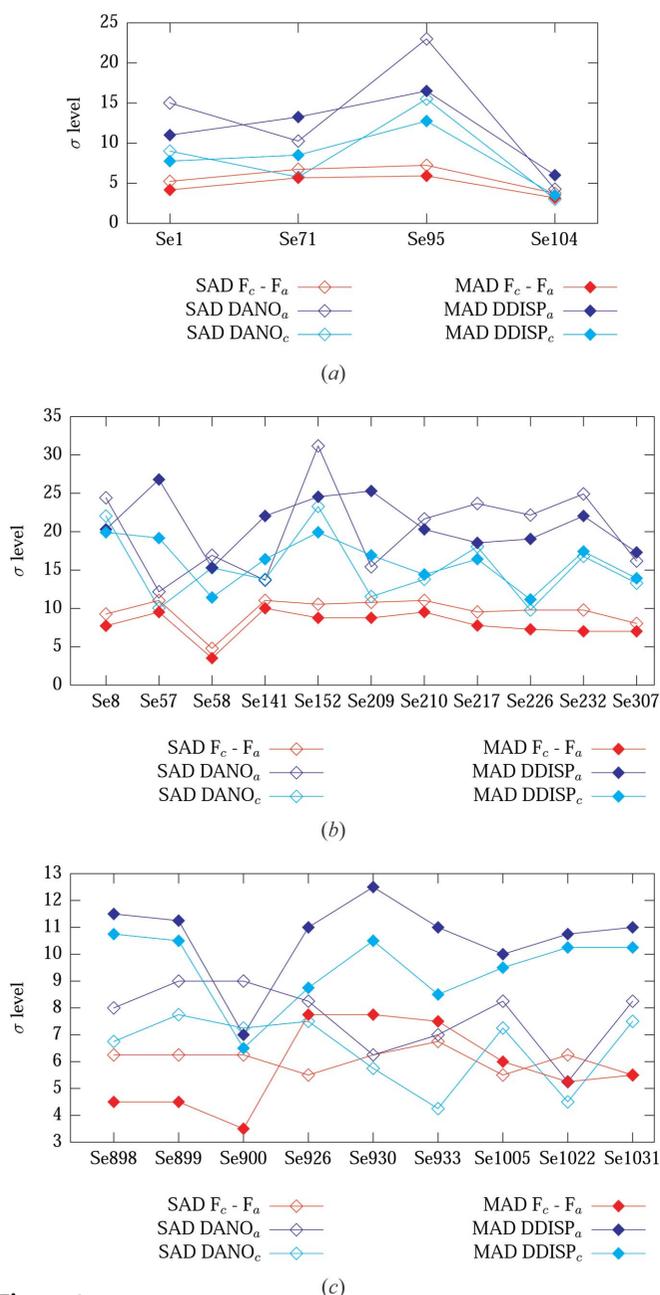


Figure 3

Peak height in difference maps  $F_b - F_a$  and  $F_c - F_a$  for SAD and MAD data collection series for Vinculin tail. The maps were calculated to a resolution of 3 Å, and the peak height averaged for sites related by non-crystallographic symmetry and for maps calculated from the two different crystals used for the MAD and SAD experiments.



**Figure 4** Comparison of the maximum radiation-damage-induced differences in SAD (SAD  $F_{c-b} - F_a$ ) and MAD experiments (MAD  $F_{c-b} - F_a$ ) and the SAD anomalous and MAD dispersive differences (SAD DANO and MAD DDISP, respectively) in the least damaged (a) and most damaged (b or c) data sets. The differences given correspond to the peaks at the selenium sites in the corresponding Fourier difference maps. For HypP, the peak heights were averaged for all NCS-related sites. For Vinculin tail, they were averaged for the different crystals and NCS-related sites.

sponding MAD data sets. This was not expected from an inspection of the initial screening shots from both areas in the fresh crystal and has to be taken into account when interpreting the results obtained in this case.

## 4.1. Substructure solution

Two-wavelength MAD had a somewhat higher rate of success than SAD when phasing in the presence of radiation

damage in all cases except for HypP. Often, the explanation for poorer performance of SAD phasing is a relatively low solvent content (Dodson, 2003). However, this cannot be the only explanation here, as unsuccessful experiments always failed at the substructure solution stage. In a previous study comparing SAD and MAD (González, 2003a), it was also found that substructure solution was the critical step when many anomalous scatterers were present (this was not the case when the substructure was smaller than ten atoms) and that the correct substructure could be found more easily combining the inflection and the remote wavelengths than using the peak SAD data, at least when *SHELXC* or *XPREP* (Bruker, 2001) are used to calculate the total anomalous contribution ( $F_A$ ) from the unmerged data from both wavelengths (González, von Delft *et al.*, 2005). In the current study, the smallest substructure (GGF6PA) contained 12 selenium atoms. This implies that for substructures larger than the low tens (or less if the anomalous signal is small) it could be far more advantageous to collect MAD data, even if the solvent content is high.

## 4.2. The importance of data collection in wedges

Even when substructure solution was decoupled from phasing by using the correct substructure, the MAD experiments tended to result in more complete structures. This is in agreement with the results obtained in previous comparisons of SAD and MAD phasing for non-damaged crystals (González, 2003a), and shows that MAD phasing is no more affected by radiation damage than SAD phasing is, as long as the reflections used to measure the dispersive differences are collected close in time to avoid losing the dispersive signal to radiation-induced loss of isomorphism. Data collection in wedges allows the effect of radiation damage to cancel out when computing the dispersive signal, as long as the anomalous scatterers do not extensively dissociate or become disordered.

SAD experiments are also likely to be sensitive to lack of isomorphism, although this may not be so apparent in cases where the crystal symmetry and alignment allow collection of Bijvoet pairs on the same image or relatively close in time. Under less favorable conditions, collecting Friedel pairs in inverse mode in wedges might also be the optimal strategy for SAD experiments.

## 4.3. The role of density modification

The presence of abundant non-crystallographic symmetry (NCS) contributes to the high number of residues traced for HypP (both for SAD and MAD experiments) using data set a. If the option to use NCS was turned off in the *RESOLVE* script, only 50–60% of the structure could be traced. GF6PA had both low solvent content and only one molecule in the asymmetric unit, and the SAD maps deteriorate most quickly in this example; however, the SAD and MAD results are comparable for the least-damaged a data set. Vinculin tail was a slightly more favorable case, with twofold NCS. In this example the density modification resulted in similar results for

the *a* and *b* data sets, while, for the *c* data set, SAD maps were again more affected than the MAD ones. These results suggest that density modification can help mitigate the effects of radiation damage on phasing (probably because the additional phasing information used at this stage of structure determination is not affected by radiation damage). However, it is not certain why phase improvement is less successful for SAD phases as radiation damage increases. Perhaps the importance of unimodal experimental phases (even poor ones) defining an initial molecular envelope increases as the data quality decreases.

#### 4.4. The role of the absorbed dose

The results obtained for the Vinculin tail and GF6PA samples could sustain the hypothesis that the lower dose received during the two-wavelength MAD experiments has an impact on the phasing results. The quantitative estimates of the specific radiation damage to heavy-atom sites (higher for SAD) are also in agreement with lower radiation damage during MAD experiments, although the advantage for MAD experiments is less than expected from the absorbed dose. This discrepancy could be caused by different rates of radiation damage at different wavelengths around the absorption edge. For example, photoelectrons excited at a remote wavelength on the high-energy side of the absorption edge are more energetic than photoelectrons generated near or at the absorption edge and are less likely to recombine with the atom. The effects of different modes and rates of decay are not sufficiently well understood yet; it would be useful to carry out experiments to investigate these effects, as it would be possible to decrease even more the absorbed dose during MAD experiments without affecting significantly the final maps by decreasing the exposure time at one of the wavelengths. Knowing whether the remote is more likely to contribute to radiation damage than the edge wavelength is therefore important.

The HypP data do not entirely support the same conclusions as the other two examples: MAD phasing did not give better results in this case while, on the other hand, the MAD data sets showed less damage than expected from the absorbed dose. The most likely explanation for this is that the comparison was rendered less meaningful because of the lower quality of the MAD data sets compared with the corresponding SAD data sets. A possibility is that the crystal was accidentally translated too near to the edge of the crystals for the MAD data collection and that a smaller area was exposed to the beam. An alternative explanation is that secondary radiation damage affected the unexposed area of the crystal prior to data collection (the SAD data were collected before the MAD data). Theoretical calculations show that photoelectrons generated by hard X-rays can diffuse over distances consistent with the separation of the two exposed zones of the crystal, which was only of the order of a few micrometers (Nave & Hill, 2005). The amount of site-specific loss of occupancy and other radiation damage indicators appear to be more consistent with an underestimation

of the absorbed dose for the MAD experiment on this sample, which would support the first explanation. However, a likely case of propagation of radiation damage to unexposed parts of a cryo-cooled crystal has been described before (Brodersen *et al.*, 2003).

#### 4.5. Specific radiation damage in SAD and MAD phasing

An important observation for the SAD and MAD experiments is that the anomalous (and dispersive differences for MAD) are far higher above the noise level than the radiation-induced differences, even for high doses. This suggests that RIP is not likely to succeed as an alternative to MAD and SAD phasing under experimental conditions similar to the ones reported here. A data collection strategy aiming to minimize both the effects of loss of isomorphism on anomalous and dispersive differences and the dose received by the crystal appears to be the most reasonable and safest strategy for experimental phasing in these cases, regardless of whether one or more wavelengths are used.

The above does not mean that specific radiation damage cannot be a potentially valuable source of improvement for the phases, even when not adequate as the sole means to solve the structure. The efforts to model the specific radiation damage during the anomalous scatterer parameter refinement and phasing [as done, for example, in *SHARP* (Evans *et al.*, 2003; Schiltz *et al.*, 2004)] are likely to have a general positive impact on structure solution regardless of the data collection strategy used.

### 5. Conclusions: what is the optimal data collection strategy?

The main advantage of using MAD rather than SAD as the method for *de novo* structure solution is the overall higher quality of density maps achievable by the former method, the availability of experimental phases and less dependency on favorable sample characteristics (*e.g.* high solvent content). The results presented here show that MAD can also be successfully used for phasing for absorbed doses during the experiment exceeding the Henderson limit ( $2.2 \times 10^7$  Gy). MAD can give better results than SAD, both in overall success of phasing and map quality; the results obtained suggest that MAD should be the method of choice particularly for low-solvent and NCS content cases, or when these conditions cannot be assessed with certainty before the experiment.

The success of two-wavelength MAD experiments with radiation damage depends critically on the data collection approach. The following guidelines can be summarized for seleno-methionine samples:

(i) Collecting data in wedges can be critical for MAD experiments, independent of the wavelengths used for data collection. This result is by no means new, as data collection in wedges was habitual in MAD experiments before sample cryocooling became a standard procedure to collect data. This procedure should become the norm again for MAD experiments in order to extract the maximum amount of information

from anomalous and dispersive differences. Good wavelength stability, highly reproducible  $\varphi$  and monochromator motors and fast or automated changes of wavelength are important requisites for MAD-dedicated beamlines, as well as data collection software that allows this strategy; if these requisites are not fulfilled, SAD experiments may be a better choice for high-dose experiments.

(ii) The order of wavelengths matters. The steady decrease of the occupancy of the heavy-atom scatterers is equivalent to decreasing the  $f'$  value at the second wavelength. This is something to consider when collecting wavelength consecutively or when using very large oscillation wedges (this is an acceptable trade-off at beamlines where changes of wavelengths are not totally automated). This implies that the remote high-energy wavelength should be collected first.

(iii) Since a lower total absorbed dose could play a role in the success of a two-wavelength MAD experiment, the exposure time at one of the wavelengths can be reduced while relying on phase extension to achieve higher resolution maps. The high-energy remote wavelength may be the best candidate for reduction of the exposure in general, since diffracted intensities are already weaker at higher energy; this would, however, not apply to MAD experiments on low-energy absorption edges (e.g. iron); in these cases the remote wavelength is the optimal one to collect higher-resolution data because of the higher beam flux at most beamlines and less absorption by material surrounding the sample. The rate of decay of the crystals at different wavelengths, if proven a significant effect, may also need to be taken into consideration.

The author thanks Dr Constantina Bakolitsa for preparing the Vinculin tail crystals, and JCSG staff for helping to select suitable crystals for this study. This research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences.

## References

Bakolitsa, C., Cohen, D. M., Bankston, L. A., Bobkov, A. A., Cadwell, G. W., Critchley, D. R., Carig, S. W. & Liddington, R. C. (2004). *Nature (London)*, **430**, 583–586.

Bakolitsa, C., de Pereda, J. M., Bagshaw, C. R., Critchley, D. R. & Liddington, R. C. (1999). *Cell*, **99**, 603–613.

Borgon, R. A., Vornrhein, C., Bricogne, G., Bois, P. R. & Izard, T. (2004). *Structure*, **12**, 1189–1197.

Brodersen, D. E., Clemons, W. M., Carter, A. P., Wimberly, B. T. & Ramakrishnan, V. (2003). *Acta Cryst.* **D59**, 2044–2050.

Bruker (2001). Bruker AXS Inc., Madison, Wisconsin, USA.

Burmeister, W. P. (2000). *Acta Cryst.* **D56**, 328–341.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.

Cohen, A. E., Ellis, P. J., Miller, M. D., Deacon, A. M. & Phizackerley, R. P. (2002). *J. Appl. Cryst.* **35**, 720–726.

Dodson, E. (2003). *Acta Cryst.* **D59**, 1958–1965.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.

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.

González, A. (2003a). *Acta Cryst.* **D59**, 315–322.

González, A. (2003b). *Acta Cryst.* **D59**, 1935–1942.

González, A., Moorhead, P., McPhillips, S. & Sauter, N. (2005). *Acta Cryst.* **A61**, C486.

González, A., Pédelacq, J.-D., Solà, M., Gomis-Rüth, F. X., Collect, M., Samama, J.-P. & Benini, S. (1999). *Acta Cryst.* **D55**, 1449–1458.

González, A., von Delft, F., Liddington, R. C. & Bakolitsa, C. (2005). *J. Synchrotron Rad.* **12**, 285–291.

Henderson, R. (1990). *Proc. R. Soc. London Ser. B*, **241**, 6–8.

Kort, R., Komori, H., Adachi, S., Miki, K. & Ecker, A. (2004). *Acta Cryst.* **D60**, 1205–1213.

Lesley, S. A. *et al.* (2002). *Proc. Natl. Acad. Sci. USA*, **99**, 11664–11669.

Leslie, A. G. (1999). *Acta Cryst.* **D55**, 1696–1702.

McPhillips, T. M., McPhillips, S. E., Chiu, H.-J., Cohen, A. E., Deacon, A. M., Ellis, P., Garman, E., González, A., Sauter, N. K., Phizackerley, R. P., Soltis, S. M. & Kuhn, P. (2002). *J. Synchrotron Rad.* **9**, 401–406.

Müller, R., Weckert, E., Zellner, J. & Drakopoulos, M. (2002). *J. Synchrotron Rad.* **9**, 368–374.

Murray, J. & Garman, E. (2002). *J. Synchrotron Rad.* **9**, 347–354.

Murray, J. W., Garman, E. F. & Ravelli, R. B. G. (2004). *J. Appl. Cryst.* **37**, 513–522.

Murray, J. W., Rudiño-Piñera, E., Owen, R. L., Grininger, M., Ravelli, R. B. G. & Garman, E. F. (2005). *J. Synchrotron Rad.* **12**, 268–275.

Nave, C. & Hill, M. A. (2005). *J. Synchrotron Rad.* **12**, 299–303.

Peterson, M. R., Harrop, S. J., McSweeney, S. M., Leonard, G. A., Thompson, A. W., Hunter, W. N. & Helliwell, J. R. (1996). *J. Synchrotron Rad.* **3**, 24–34.

Ramagopal, U. A., Dauter, Z., Thirumuruhan, R., Fedorov, E. & Almo, S. C. (2005). *Acta Cryst.* **D61**, 1289–1298.

Ravelli, R. B. G. & McSweeney, S. M. (2000). *Struct. Fold. Des.* **8**, 315–328.

Ravelli, R. B. G., Nanao, M. H., Lovering, A., White, S. & McSweeney, S. (2005). *J. Synchrotron Rad.* **12**, 276–284.

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

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

Rice, L. M., Earnest, T. N. & Brünger, A. T. (2000). *Acta Cryst.* **D56**, 1413–1420.

Sauter, N. K., Grosse-Kunstleve, R. W. & Adams, P. D. (2004). *J. Appl. Cryst.* **37**, 399–409.

Schiltz, M., Dumas, P., Ennifar, E., Flensburg, C., Paciorek, W., Vornrhein, C. & Bricogne, G. (2004). *Acta Cryst.* **D60**, 1024–1031.

Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* **D58**, 1772–1779.

Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.

Smith, J. L. (1991). *Curr. Opin. Struct. Biol.* **1**, 1002–1011.

Terwilliger, T. C. (2000). *Acta Cryst.* **D56**, 965–972.

Terwilliger, T. C. (2003). *Acta Cryst.* **D59**, 38–44.

Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* **D55**, 849–861.

Weik, M., Bergès, J., Raves, M. L., Gros, P., McSweeney, S., Silman, I., Sussman, J. L., Houée-Levin, C. & Ravelli, R. B. G. (2002). *J. Synchrotron Rad.* **9**, 342–346.

Zwart, P. H., Banumathi, S., Dauter, M. & Dauter, Z. (2004). *Acta Cryst.* **D60**, 1958–1963