Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure determination of VEGF-E by sulfur SAD

Armin Wagner,[a]‡ Michel Pieren,[b] Clemens Schulze-Briese,[a] Kurt Ballmer-Hofer[b] and Andrea E. Prota[b]*

[a]Swiss Light Source, Paul Scherrer Institut, Villigen-PSI, Switzerland, and [b]Biomolecular Research, Paul Scherrer Institut, Villigen-PSI, Switzerland

‡ Current address: Diamond Light Source Ltd, Chilton, United Kingdom.

Correspondence e-mail: andrea.prota@psi.ch

Received 16 May 2006
Accepted 9 September 2006

PDB Reference: VEGF-E, 2gnn, r2gnnsf.

The crystal structure of VEGF-E was solved by the sulfur single-wavelength anomalous dispersion method (S-SAD) using highly redundant low-resolution data collected at a wavelength of $\lambda \simeq 1.7$ Å with an estimated anomalous signal of 1.5%. 11 sulfur sites, nine out of 16 disulfide bonds and two out of 12 methionines could be located in the asymmetric unit using data truncated at a resolution of 4.1 Å; however, none of the common diffraction data-quality indicators for SAD allowed clear discrimination between successful and unsuccessful resolution cutoffs. The high solvent content of 75% allowed efficient density modification to be performed and an unbiased electron-density map of good quality to be generated. This study demonstrates the strength of S-SAD for phasing using accurate highly redundant data at low resolution.

1. Introduction

The multiple-wavelength anomalous dispersion method (MAD) is currently the most successful technique for determining experimental phases from protein crystals. In the absence of anomalous scatterers with absorption edges in the typical wavelength range of synchrotron beamlines, such scatterers are incorporated into the crystal by the preparation of either selenium-labelled proteins or heavy-atom derivatives. In the case of the sulfur single-wavelength anomalous dispersion method (S-SAD), no additional derivatives are needed and experimental phases can be derived directly from native protein crystals. As the absorption edge of sulfur ($\lambda = 5.02$ Å) is outside the usual range of protein crystallography beamlines, experiments are carried out at energies well above the *K* edge, typically at wavelengths in the range $\lambda = 1.5$–3.0 Å (Mueller-Dieckmann *et al.*, 2005 and references cited therein). Hence, the anomalous signals are weak and only a single-wavelength approach is chosen.

The first successful S-SAD study was performed in 1981 on the small protein crambin (Hendrickson & Teeter, 1981) and in 1985 B.-C. Wang introduced computational methods to overcome the twofold ambiguity of the SAD approach (Wang, 1985). However, it took almost 15 years before S-SAD was reconsidered as a general phasing method (Dauter *et al.*, 1999). A number of structure determinations have been published, mainly on well established protein crystals diffracting to atomic resolution, and only a small number of novel structures have been determined to date (*e.g.* Liu *et al.*, 2000; Yang & Pflugrath, 2001; Micossi *et al.*, 2002; Lemke *et al.*, 2002; Debreczeni, Bunkoczi, Girmann *et al.*, 2003; Debreczeni, Bunkoczi, Ma *et al.*, 2003; Debreczeni, Girmann *et al.*, 2003; Ramagopal *et al.*, 2003; Olsen *et al.*, 2004; Weiss *et al.*, 2004; Agarwal *et al.*, 2006). The following study presents the successful S-SAD *de novo* structure determination of vascular endothelial growth factor E (VEGF-E) and demonstrates that even low-resolution anomalous data are sufficient to yield experimental phases and hence solve the structure.

2. Materials and methods

2.1. Crystallization

The expression and purification of VEGF-E have been described in Pieren *et al.* (2006). The crystallized protein represents the mature full-length VEGF-E variant NZ2 encompassing residues 20–133 with

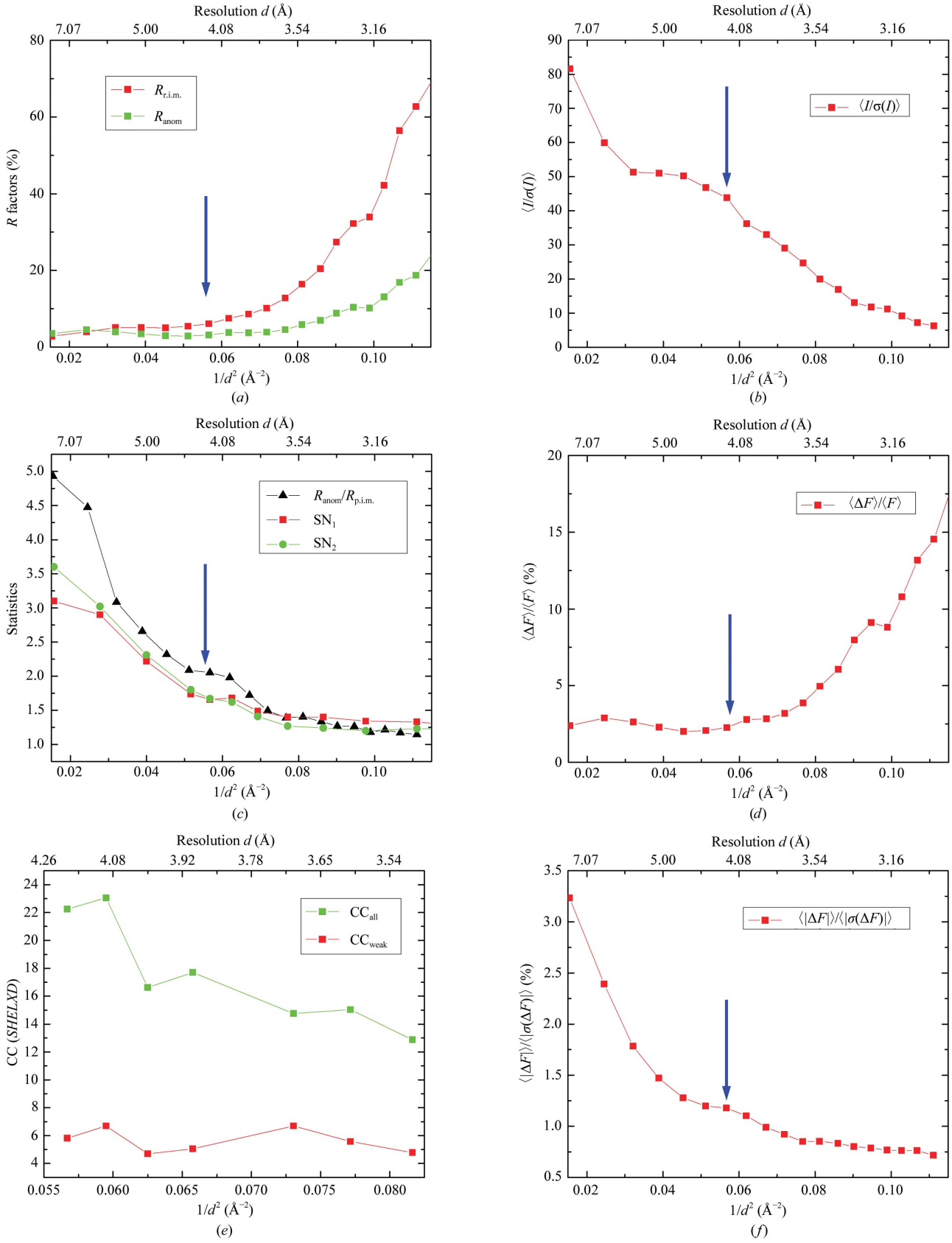

Figure 1
Statistical values for the S-SAD data set plotted as a function of resolution. The data points are set at the high-resolution limits of the corresponding bins. The resolution cutoff for the successful sulfur-substructure determination is indicated by a blue arrow in each panel. (*a*) $R_{\rm r.i.m.}$ and $R_{\rm anom}$. (*b*) $\langle I/\sigma(I)\rangle$. (*c*) Signal-to-noise ratios. Black: $R_{\rm anom}/R_{\rm p.i.m.}$ ratio. Green: SN_1, a signal-to-noise ratio based on input σ values from *XPREP*. Red: SN_2, a signal-to-noise ratio based on variances of F^+ and F^- from *XPREP*. (*d*) The Bijvoet ratio $\langle|\Delta F^{\rm anom}|/|F|\rangle$. (*e*) Correlation coefficients $CC_{\rm all}/CC_{\rm weak}$ of the top solutions from *SHELXD*. (*f*) $\langle|\Delta F^{\rm anom}|\rangle/\langle\sigma(\Delta F)\rangle$.

a carboxyterminal hexahistidine tag. Eight cysteines and three methionine residues are present in the polypeptide chain, which has a total molecular weight of 14.2 kDa. Crystals with average dimensions of 250 × 250 × 700 µm were grown at 293 K within 10 d using the hanging-drop method by mixing 0.8 µl reservoir buffer (0.6 *M* ammonium sulfate, 3% PEG 4K, 0.1 *M* sodium citrate buffer pH 5, 0.3% benzamidine) and 1.6 µl protein solution at 40 mg ml^{-1} (50 m*M* Tris–HCl pH 7.5, 0.1 *M* NaCl). For data collection, crystals were sequentially soaked for 1 min in 2 µl drops containing reservoir buffer supplemented with 0.3% benzamidine and 5–30% glycerol and were flash-cooled directly in the nitrogen cryostream at the beamline.

2.2. Data collection and processing

VEGF-E crystallizes in space group $P4_122$ ($a = b = 98.7$, $c = 240.2$ Å) and contains two dimers in the asymmetric unit. For the S-SAD phasing experiment, highly redundant data were collected at 100 K to a resolution of 3.0 Å (λ = 1.7 Å) at beamline X06SA of the Swiss Light Source (SLS, PSI-Villigen, Switzerland) using a MAR 165 CCD detector. Because of the relatively small expected anomalous signal from sulfur ($\Delta F/F \simeq 1.5\%$; Hendrickson & Teeter, 1981), special care was taken to choose the data-collection parameters. In order to balance undesired absorption effects against the strength of the anomalous signal, a wavelength of λ = 1.7 Å was chosen. The choice was based on in-house experience that an expected anomalous signal of 1.5% should be sufficient for a successful structure solution. Furthermore, to minimize the detrimental effect of possible vibrations introduced by the cryostream and beam fluctuations, the exposure time was adjusted to 6 s per frame and the beam was vertically defocused to an approximate beam size of 80 × 120 µm (horizontal × vertical, FWHM). The primary beam intensity was attenuated to 2.7% to reduce radiation damage and to avoid saturation of the low-resolution data. A total of 360° of data were collected as a consecutive series of 1° rotation images, giving an overall redundancy of 15 with separate Friedel pairs. During the data collection, the crystal was horizontally translated by 100 µm every 90° in order to reduce the effect of radiation damage on the data.

High-resolution data to 2.3 Å were assembled from a high- and a low-resolution run collected from a second crystal of similar size. For both runs, 180 images with a rotation range of 0.5° and an exposure time of 5 s were collected at a wavelength of λ = 1.0 Å using a focused beam (80 × 15 µm FWHM). Because of the limited dynamic range of the CCD detector, a second run with the beam attenuated to 2.4% was taken to obtain complete data.

All data were integrated and scaled using *XDS* and *XSCALE* (Kabsch, 1993). The program *XPREP* (Sheldrick, 2002*a*) was used to merge symmetry-equivalent reflections and to calculate quality indicators and anomalous differences. Diffraction data statistics are given in Table 1.

Table 1
Data-collection and refinement statistics.

Values in parentheses refer to the outermost 0.1 Å resolution shell.

	S-SAD	High resolution
Space group	$P4_122$	$P4_122$
Unit-cell parameters (Å)	$a = b = 98.2$, $c = 240.4$	$a = b = 98.7$, $c = 240.1$
Source	SLS-X06SA	SLS-X06SA
Wavelength (Å)	1.6984	1.0003
Resolution range (Å)	50–3.0 (3.1–3.0)	50–2.3 (2.4–2.3)
Total reflections	685907	391093
Unique reflections	44644†	52820‡
Redundancy	15.4	7.4
Completeness (%)	99.2 (99.1)	98.7 (98.4)
R_{sym}§ (%)	8.2 (47.3)	8.3 (93.7)
$\langle I/\sigma(I)\rangle$	29.8 (7.6)	13.8 (2.6)
No. of S sites	11	
R_{cryst}¶, working set (%)		22.3
R_{cryst}¶, free set (%)		24.7
R.m.s.d. bond lengths (Å)		0.011
R.m.s.d. bond angles (°)		1.2

‡ Friedel pairs merged. † Friedel pairs treated as independent reflections. § $R_{sym} = \sum|I_{hkl} - \langle I\rangle|/\sum I_{hkl}$, where I is the intensity of a reflection hkl and $\langle I\rangle$ is the average over symmetry-related reflections of hkl. ¶ $R_{cryst} = \sum|F_o - F_c|/\sum|F_o|$, in which F_o and F_c are the observed and calculated structure-factor amplitudes, respectively. The free set (Brünger, 1992) is calculated from 5% of the reflections not used in the model refinement.

2.3. Substructure determination, phasing and phase improvement

For the sulfur-substructure determination, the dual-space recycling program *SHELXD* (Schneider & Sheldrick, 2002) was used. All attempts with resolution cutoffs $d < 4.1$ Å and $d > 4.2$ Å failed; correct solutions could only be obtained for cutoffs of 4.1 and 4.2 Å, with a success rate of one or two correct solutions in 1000 trials (Fig. 1*e*). The solution with the highest correlation coefficient was directly fed into *SHELXE* (Sheldrick, 2002*b*) for phasing and density modification without further refinement. Owing to the high solvent content of 75%, the correct enantiomorph could be identified easily from the contrast of 1.255 compared with 0.787 for the inverted structure. The 11 positions with the highest occupancies were identified as nine 'super-sulfurs' consisting of unresolved disulfide bridges (nine out of a total of 16) and two methionine sulfurs (out of a total of 12). In Fig. 2, the 11 sites from *SHELXD* are superimposed with the anomalous difference Fourier map after phase extension to 3.3 Å with *SHELXE*. The resulting electron-density map was already interpretable. The map was further improved by phase extension to 2.5 Å resolution using *SHELXE* (Sheldrick, 2002*b*). We also tried to further refine the 11 sulfur positions with *SHARP*, where all parameters, including those modelling the residual error, were refined using a fully

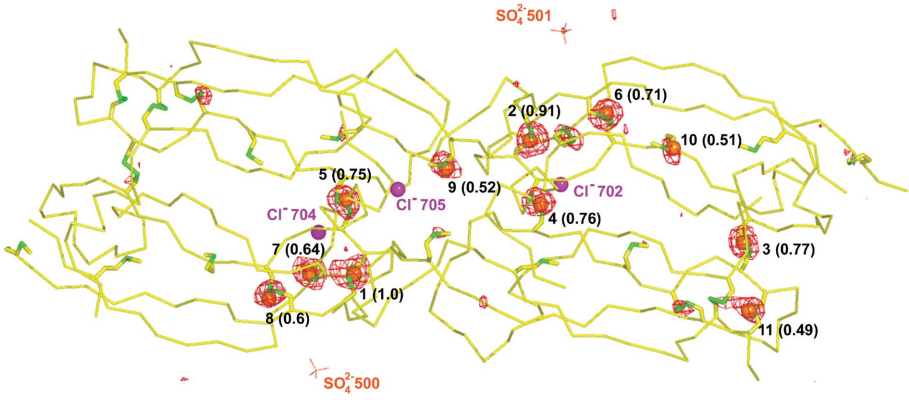

Figure 2
C^α trace of the final VEGF-E model superimposed with the sulfur sites located by *SHELXD* and the anomalous difference Fourier map calculated after phase extension to 3.3 Å resolution with *SHELXE*. The map is contoured at 3.5σ. Cysteine and methionine residues are shown in stick representation. S atoms are highlighted in green. The sulfur sites from *SHELXD* are shown as orange spheres. The numbers correspond to the site number, with the corresponding occupancy in parentheses. Three chlorides and two sulfate ions are shown as magenta spheres and orange sticks, respectively. This figure was prepared with *PyMOL* (DeLano, 2002).

fledged maximum-likelihood approach (de La Fortelle & Bricogne, 1997). Solvent flattening was subsequently performed by the program *SOLOMON* (Abrahams & Leslie, 1996). Both experimental electron-density maps from *SHELXE* and *SHARP* were of comparable overall quality, differing only in some regions. Therefore, both maps were used for model building.

2.4. Model building and refinement

We also tried molecular replacement using the core structure of VEGF-A (41.5% sequence identity) as a search model (by mutating all nonconserved residues to alanine and by deleting all flexible loop regions), but did not obtain a clear solution. However, one solution matched the experimental electron density from *SHELXE* very well and was used as a scaffold to build the VEGF-E model. All residues were manually fitted into the electron-density map using the computer-graphics program *MOLOC* (Gerber & Muller, 1995). After an initial rigid-body refinement, the manually built model was refined against the 50–2.3 Å high-resolution data set using *REFMAC* v.5.2 (Murshudov *et al.*, 1997). It was completed by the addition of 177 water molecules, three chloride ions, three benzamidine molecules and two *N*-acetylglucosamine units (attached at Asn75 of two of the four crystallographically independent VEGF monomers) and further improved through iterative rounds of model building and refinement. Restrained refinements using TLS parameters and individual isotropic *B* factors were very effective in improving the quality of the fit and reduced the *R* factors by more than 4% from the initial round of refinement. The final model converged at an *R* factor of 22.3% ($R_{\rm free}$ = 24.7%) with very good stereochemistry (Table 1). Quality checks of the final model were performed using *PROCHECK* (Laskowski *et al.*, 1993), *WHATIF* (Vriend, 1990) and *ROTAMER* (Collaborative Computational Project, Number 4, 1994).

No residues were observed in the generous and in the disallowed regions of the Ramachandran plot. The final statistics of the structure solution and refinement are given in Table 1. The biological implications of the crystal structure of VEGF-E are discussed in Pieren *et al.* (2006).

3. Discussion

Several statistical parameters, such as $R_{\rm sym}$, $R_{\rm r.i.m.}$, $R_{\rm mrgd\text{-}\mathit{F}}$, $R_{\rm p.i.m.}$ or $\langle I/\sigma(I)\rangle$, are generally used to assess the quality of X-ray data (Diederichs & Karplus, 1997; Drenth, 1994; Weiss, 2001; Weiss & Hilgenfeld, 1997; Weiss *et al.*, 1998). However, none of these parameters have a direct correlation with the anomalous signal in an X-ray diffraction data set. Additional statistical figures, such as the Bijvoet difference ratio ($R_{\rm anom}$; Hendrickson & Teeter, 1981), the normal probability plot (δR plot; Howell & Smith, 1992) and the χ^2 statistic (Otwinowski & Minor, 1997), have been proposed for estimation of the anomalous signal level, but do not provide a quantitative indication. Two sets of anomalous signal-to-noise ratios (SN_1 and SN_2) implemented in the program *XPREP* (Sheldrick, 2002*a*), the $R_{\rm anom}/R_{\rm p.i.m.}$ ratio (Mueller-Dieckmann *et al.*, 2005) and the $R_{\rm as}$ index implemented in the program 3*DSCALE* (Fu *et al.*, 2004) have been proposed as indicators to judge the usefulness of a data set for solving the anomalous scatterer substructure. As a rule of thumb, to obtain phases of sufficient quality a value of 1.5 has been proposed for all of the three indicators; however, none of them is predictive of the success of a phasing experiment. Therefore, data accuracy remains the critical quality parameter for the success of a SAD phasing experiment. We plotted several statistical values for our S-SAD data set against resolution (Fig. 1). As judged by $R_{\rm r.i.m.}$, $R_{\rm anom}$ and $\langle I/\sigma(I)\rangle$, the S-SAD data set is of excellent quality (Figs. 1*a* and 1*b*). The $R_{\rm anom}/R_{\rm p.i.m.}$ ratio and the signal-to-noise ratios from *XPREP* (SN_1 and SN_2) fall below the value of 1.5 in the 3.8–3.7 and 4–3.8 Å resolution bins, respectively (Fig. 1*c*). The ratio of ΔF to its estimated standard deviation $[\langle\Delta F/\sigma(\Delta F)\rangle]$ as a function of resolution drops below the 1.3 limit given by Schneider & Sheldrick (2002) in the 5.07–4.7 Å resolution shell (Fig. 1*f*). These statistical figures indicate a significant anomalous signal; however, none of them provides a clear indication why the sulfur-substructure determination was only successful with resolution cutoffs of 4.1 and 4.2 Å. All plots show little 'bumps' in this resolution range, but no drastic changes. These data show that the correct resolution cutoff cannot be determined automatically and therefore has to be carefully and systematically evaluated.

4. Conclusions

The structure determination of VEGF-E demonstrates that in the case of a protein crystal with high sulfur content and high solvent content, low-resolution anomalous difference data below 4.0 Å resolution can be sufficient to locate the correct sulfur substructure and deduce experimental phases from it. The examined diffraction data-quality indicators could not predict the correct resolution cutoff for the sulfur-substructure determination. Therefore, a wide range of cutoffs should be taken into consideration.

As specific radiation damage is primarily located at the sulfur sites in the absence of other anomalous scatterers, it is better to collect highly redundant data of limited resolution to obtain as accurate data as possible instead of pushing the diffraction to the crystal's resolution limit. X-ray doses that are too high would directly alter the sulfur substructure and therefore change the anomalous signal as a function of dose. This site-specific radiation damage can be exploited for phasing *via* the RIPAS method (Ravelli *et al.*, 2003), an approach which was not considered in this study. Highest redundancy does not necessarily yield the best anomalous differences (Sarma & Karplus, 2006). Specific and non-specific radiation damage degrade the anomalous signal and therefore at the brightest synchrotron beamlines the X-ray beam should be attenuated for S-SAD experiments. Additionally, for crystals that are larger than the X-ray beam it is recommended to vertically defocus the beam and to translate the crystal horizontally to minimize radiation-damage effects.

Currently, several synchrotron beamlines optimized for low-energy phasing are in the planning phase (Djinovic Carugo *et al.*, 2005). They include specific hardware devices such as He beam paths. In combination with improved algorithms to correct the diffraction data for crystal and solvent absorption, experiments at longer wavelengths will be possible and will extend the S-SAD method to proteins with lower sulfur content.

We would like to thank Dr Dirk Kostrewa for his valuable help and for useful discussions. AW acknowledges the NCCR Structural Biology for financial support.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Agarwal, R., Bonanno, J. B., Burley, S. K. & Swaminathan, S. (2006). *Acta Cryst.* D**62**, 383–391.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.

Debreczeni, J. É., Bunkoczi, G., Girmann, B. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 393–395.
Debreczeni, J. É., Bunkoczi, 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.
DeLano, W. L. (2002). *The Pymol Molecular Graphics System.* DeLano Scientific, San Carlos, CA, USA.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
Djinovic Carugo, K., Helliwell, J. R., Stuhrmann, H. & Weiss, M. S. (2005). *J. Synchrotron Rad.* **12**, 410–419.
Drenth, J. (1994). *Priciples of X-ray Crystallography.* New York: Springer–Verlag.
Fu, Z.-Q., Rose, J. P. & Wang, B.-C. (2004). *Acta Cryst.* D**60**, 499–506.
Gerber, P. R. & Muller, K. (1995). *J. Comput. Aided Mol. Des.* **9**, 251–268.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Howell, P. L. & Smith, G. D. (1992). *J. Appl. Cryst.* **25**, 81–86.
Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Laskowski, R. A., MacArthur, M., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lemke, C. T., Smith, G. D. & Howell, P. L. (2002). *Acta Cryst.* D**58**, 2096–2101.
Liu, Z.-J., Vysotski, E. S., Chen, C.-J., Rose, J. P., Lee, J. & Wang, B.-C. (2000). *Protein Sci.* **9**, 2085–2093.
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.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Olsen, J. G., Flensburg, C., Olsen, O., Bricogne, G. & Henriksen, A. (2004). *Acta Cryst.* D**60**, 250–255.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pieren, M., Prota, A. E., Ruch, C., Kostrewa, D., Wagner, A., Biedermann, K., Winkler, F. K. & Ballmer-Hofer, K. (2006). *J. Biol. Chem.* **281**, 19578–19587.
Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 1020–1027.
Ravelli, R. B. G., Leiros, H.-K. S., Pan, B., Caffrey, M. & McSweeney, S. (2003). *Structure*, **11**, 217–224.
Sarma, G. N. & Karplus, P. A. (2006). *Acta Cryst.* D**62**, 707–716.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Sheldrick, G. M. (2002*a*). *XPREP* v.6.10. Bruker-AXS, Madison, Wisconsin, USA.
Sheldrick, G. M. (2002*b*). *Z. Kristallogr.* **217**, 644–650.
Vriend, G. (1990). *J. Mol. Graph.* **8**, 52–56.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Weiss, M. (2001). *J. Appl. Cryst.* **34**, 130–135.
Weiss, M. S. & Hilgenfeld, R. (1997). *J. Appl. Cryst.* **30**, 203–205.
Weiss, M. S., Mander, G., Hedderich, R., Diederichs, K., Ermler, U. & Warkentin, E. (2004). *Acta Cryst.* D**60**, 686–695.
Weiss, M. S., Metzner, H. J. & Hilgenfeld, R. (1998). *FEBS Lett.* **423**, 291–296.
Yang, C. & Pflugrath, J. W. (2001). *Acta Cryst.* D**57**, 1480–1490.