

Paula S. Salgado,^a Martin A. Walsh,^b Minni R. L. Laurila,^c David I. Stuart^a and Jonathan M. Grimes^{a*}

^aDivision of Structural Biology, The Henry Wellcome Building for Genomic Medicine, Oxford University, Roosevelt Drive, Oxford OX3 7BN, England, ^bMedical Research Council France, European Synchrotron Radiation Facility, BP 220, F-38043, Grenoble CEDEX, France, and ^cInstitute of Biotechnology and Department of Biological and Environmental Sciences, Viikki Biocenter, PO Box 56 (Viikinkari 5) 00014, University of Helsinki, Finland

Correspondence e-mail:
jonathan@strubi.ox.ac.uk

Received 4 August 2004
Accepted 22 October 2004

Going soft and SAD with manganese

SAD phasing has been revisited recently, with experiments being carried out using previously unconventional sources of anomalous signal, particularly lighter atoms and softer X-rays. A case study is reported using the 75 kDa RNA-dependent RNA polymerase of the bacteriophage $\varphi 6$, which binds a Mn atom and crystallizes with three molecules in the asymmetric unit. X-ray diffraction data were collected at a wavelength of 1.89 Å and although the calculated anomalous signal from the three Mn atoms was only 1.2%, *SHELXD* and *SOLVE* were able to locate these atoms. *SOLVE/RESOLVE* used this information to obtain SAD phases and automatically build a model for the core region of the protein, which possessed the characteristic features of the right-hand polymerase motif. These results demonstrate that with modern synchrotron beamlines and software, manganese phasing is a practical tool for solving the structure of large proteins.

1. Introduction

In recent years, the possibility of systematically modifying proteins to replace sulfur atoms with selenium in methionine residues and the availability of synchrotron beamlines where the wavelength of the radiation is tuneable has rendered the multiple-wavelength anomalous dispersion (MAD) method (Hendrickson *et al.*, 1990; Hendrickson & Ogata, 1997) the method of choice for solving protein structures where molecular replacement is not possible. That being said, the use of single-wavelength anomalous dispersion (SAD) measurements for protein structure determination has seen a recent revival (Rice *et al.*, 2000; Dodson, 2003). A SAD experiment involves collection of a single data set at (or as close as possible to) the high-energy side of the absorption edge of the anomalous scatterer, thus maximizing the anomalous signal obtained (as opposed to MAD experiments, which involve the collection of two or more data sets at different wavelengths; Walsh *et al.*, 1999). The power of SAD even with very small anomalous signals was illustrated over 20 years ago by Hendrickson & Teeter (1981), who determined the structure of the small protein crambin utilizing solely the anomalous signal from the naturally occurring sulfur atoms in the protein. Furthermore, Wang (1985) postulated that very small anomalous signals in SAD experiments would be useful if these small differences could be measured accurately and described a number of test cases utilizing the method. However, the use of the weak anomalous signal from naturally occurring sulfur atoms in proteins was largely ignored until Dauter *et al.* (1999) demonstrated the utility of such signals measured with softer radiation at synchrotron sources. These authors illustrated that for hen egg-white lysozyme the small signal from sulfur at a wavelength of 1.54 Å was sufficient to determine the structure. This solution was achieved by collecting highly redundant data. Since then, the successful use of sulfur for *de novo* phase determination has been reported by a number of groups (Brown *et al.*, 2002; Gordon *et al.*, 2001; Lartigue *et al.*, 2004; Liu *et al.*, 2000; Micossi *et al.*, 2002), as has the use of other light atoms for protein structure determination, such as phosphorous and manganese (Dauter & Adamak, 2001; Ramagopal *et al.*, 2003a).

Manganese, with a *K* edge at 1.89 Å, which is within the softer X-ray range, is a natural cofactor for many enzymes and would seem to be a perfect candidate for SAD phasing of proteins. The RNA-dependent RNA polymerase (RdRP) of bacteriophage $\varphi 6$ (75 kDa) catalyses both replication and transcription of the viral RNA (Makeyev & Bamford, 2000b). Polymerase activity is stimulated up to

fivefold *in vitro* by the presence of manganese ions (optimal at 2 mM; Ojala & Bamford, 1995; van Dijk *et al.*, 1995; Yang *et al.*, 2001; Makeyev & Bamford, 2000*b*). The structure has been described previously both in isolation and in a number of functionally relevant complexes (Butcher *et al.*, 2001; Salgado *et al.*, 2004). It has the characteristic ‘right-hand’ polymerase motif, with fingers, palm and thumb subdomains. The crystal structures identified a high-affinity binding site for Mn²⁺ in the palm domain of the protein, providing a potential source of phase information. Here, we report phasing of the 225 kDa protein asymmetric unit (three φ 6 polymerase molecules) using the anomalous scattering signal from three manganese ions recorded near the manganese *K* edge.

2. Methods and results

2.1. Expression, purification and crystallization

The *Escherichia coli* expression, purification and crystallization of recombinant full-length polymerase has been described previously (Makeyev & Bamford, 2000*a*; Butcher *et al.*, 2000).

2.2. Data collection and processing

X-ray diffraction data were collected at the high-energy side of the manganese *K* edge, 6.552 keV (wavelength 1.89 Å), on the UK MAD beamline BM14 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France using a MAR Research 165 mm CCD detector. At this wavelength the expected f'' signal for an Mn²⁺ ion is 3.95 e⁻, providing an expected anomalous scattering signal of approximately 1.2%. All data were collected as 1° oscillation images with the crystal maintained at 100 K in a nitrogen-gas stream. Crystals were cryoprotected with 25%(*v/v*) glycerol.

Data were processed and scaled using HKL-2000 (Otwinowski & Minor, 1997). Crystals were isomorphous with those reported previously (Butcher *et al.*, 2000, 2001; Salgado *et al.*, 2004), belonging to space group $P2_1$, with unit-cell parameters $a = 105.5$, $b = 93.6$, $c = 141.1$ Å, $\beta = 101.1^\circ$. As the anomalous signal expected was small, highly redundant data were collected in order to provide highly accurate measured diffraction intensities. Although restricted to a single rotation axis, 647° of data were collected, providing an overall redundancy of 20. The anomalous signal-to-noise ratio, as calculated by XPREP (Bruker, Madison, WI, USA), is 2.2 in the low-resolution shell (30–8 Å) and 1.3 at 3.5 Å, reflecting the small anomalous signal. Data-processing statistics are given in Table 1. Note that whilst the diffraction is strong at low resolution [at 4 Å $\langle I \rangle / \langle \sigma(I) \rangle = 39$], the diffraction limit is little better than 2.9 Å [$\langle I \rangle / \langle \sigma(I) \rangle = 25$ at 3.5 Å, falling to 6.5 in the 3–2.9 Å shell], *i.e.* the quality of diffraction from the crystal is mediocre and representative of that observed for many interesting proteins.

2.3. SAD phasing

Phase determination requires solving the substructure of the anomalous scatterers; in this case, the positions of the three manganese ions present in the asymmetric unit (Fig. 1). Calculated phases and derived electron-density maps were compared with reference data derived from the previously refined 2.0 Å model (Butcher *et al.*, 2001) by calculating mean phase errors with PHISTATS (Collaborative Computational Project, Number 4, 1994) and map correlation coefficients with the program GAP (D. I. Stuart, J. M. Grimes and J. M. Diprose, unpublished work) (Table 2). On the basis of the anomalous signal-to-noise ratio calculated by XPREP, data were initially truncated to 3.5 Å and the bulk of the discussion in this paper refers to results to this resolution. Subsequently, we repeated the

Table 1
Data-collection and processing statistics.

Values in parentheses refer to the highest resolution bin.

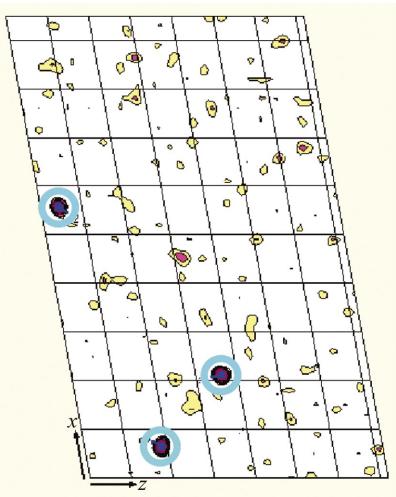
X-ray source	ESRF BM14
Wavelength (Å)	1.89
Space group	$P2_1$
Unit-cell parameters (Å, °)	$a = 105.5$, $b = 93.6$, $c = 141.1$, $\beta = 101.1$ 30.0–2.9 (3.0–2.9)
Resolution range (Å)	647
No. of images (1° oscillations)	1230575
Observations	59598
Unique reflections	99.2 (97.9)
Completeness (%)	29.6 (6.5)
$I/\sigma(I)$	9.5 (30.5)
R_{merge} (%)	
Anomalous $I/\sigma(I)$	
∞–8.0 Å	2.2
8.0–6.0 Å	2.1
6.0–5.0 Å	1.9
5.0–4.5 Å	1.8
4.5–4.3 Å	1.7
4.3–4.1 Å	1.7
4.1–3.9 Å	1.5
3.9–3.7 Å	1.4
3.7–3.5 Å	1.3
3.5–3.3 Å	1.1
3.3–3.1 Å	1.0
3.1–2.9 Å	1.0
Solvent content (%)	60

calculations using data to 2.9 Å (these results are summarized in Table 2); the process still worked well, with improved side-chain resolution compared with the maps at 3.5 Å resolution.

2.3.1. Mn-substructure solution. Both *SHELXD* (Schneider & Sheldrick, 2002) and *SOLVE* (Terwilliger & Berendzen, 1999) correctly identified the three manganese-ion sites. The correlation coefficients for the correct solution were 0.31 (all data) and 0.19 (weak data) from *SHELXD*. The mean figure of merit for the solutions found by *SOLVE* is 0.32, with an overall Z-score value of 35.2. In both cases, there was a clear step in the occupancies between correct and incorrect atomic sites. The solutions found by each program were verified by visual inspection of the self- and cross-vectors in the difference Patterson maps (Fig. 1) and when brought to a common origin the positions are identical within 0.1 Å. The *SHELXD* and *SOLVE* manganese substructures agree with those described previously (Butcher *et al.*, 2001). The hand of the substructure could be determined in the usual manner by visual comparison of electron-density maps calculated (by *SOLVE*) using both alternatives.

2.3.2. Phase calculation. *SOLVE* was used to refine the positions, occupancies and *B* factors of the manganese ions and to derive estimates of the protein phases. The overall figure of merit calculated by *SOLVE* was 0.25 and statistics of the quality of the phases are given in Table 2. The mean phase error between the phases derived from *SOLVE* and those of the reference set is 67°, which corresponds to a figure of merit of 0.39. Visual inspection of the resulting maps showed interpretable regions and secondary-structure features could readily be identified (Fig. 2a). The high real-space correlation coefficient with the model-derived maps (0.69) indicates that those features already correspond to the protein model, although the noise level in the map is high.

2.3.3. Automated model building. Phases calculated by *SOLVE* were improved by solvent flattening and density modification using *RESOLVE* (Terwilliger, 2000) (Table 2). *RESOLVE* dramatically improved the phases (mean phase error 45°) and the maps are very similar to those derived from the reference model (correlation coefficient = 0.80).

**Figure 1**

Anomalous difference Patterson map. Harker section ($v = 1/2$). Self-vectors are shown as cyan circles. Heavy-atom positions found by *SHELXD* were 0.610, 0.797, 0.126 with occupancy 1, 0.210, 0.710, 0.485 with occupancy 0.9 and 0.033, 0.333, 0.065 with occupancy 0.9. *SOLVE* positions were 0.610, 0.796, 0.126 with occupancy 0.44 and $B = 34 \text{ \AA}^2$, 0.210, 0.710, 0.485 with occupancy 0.40 and $B = 28 \text{ \AA}^2$ and 0.033, 0.333, 0.065 with occupancy 0.40 and $B = 31 \text{ \AA}^2$.

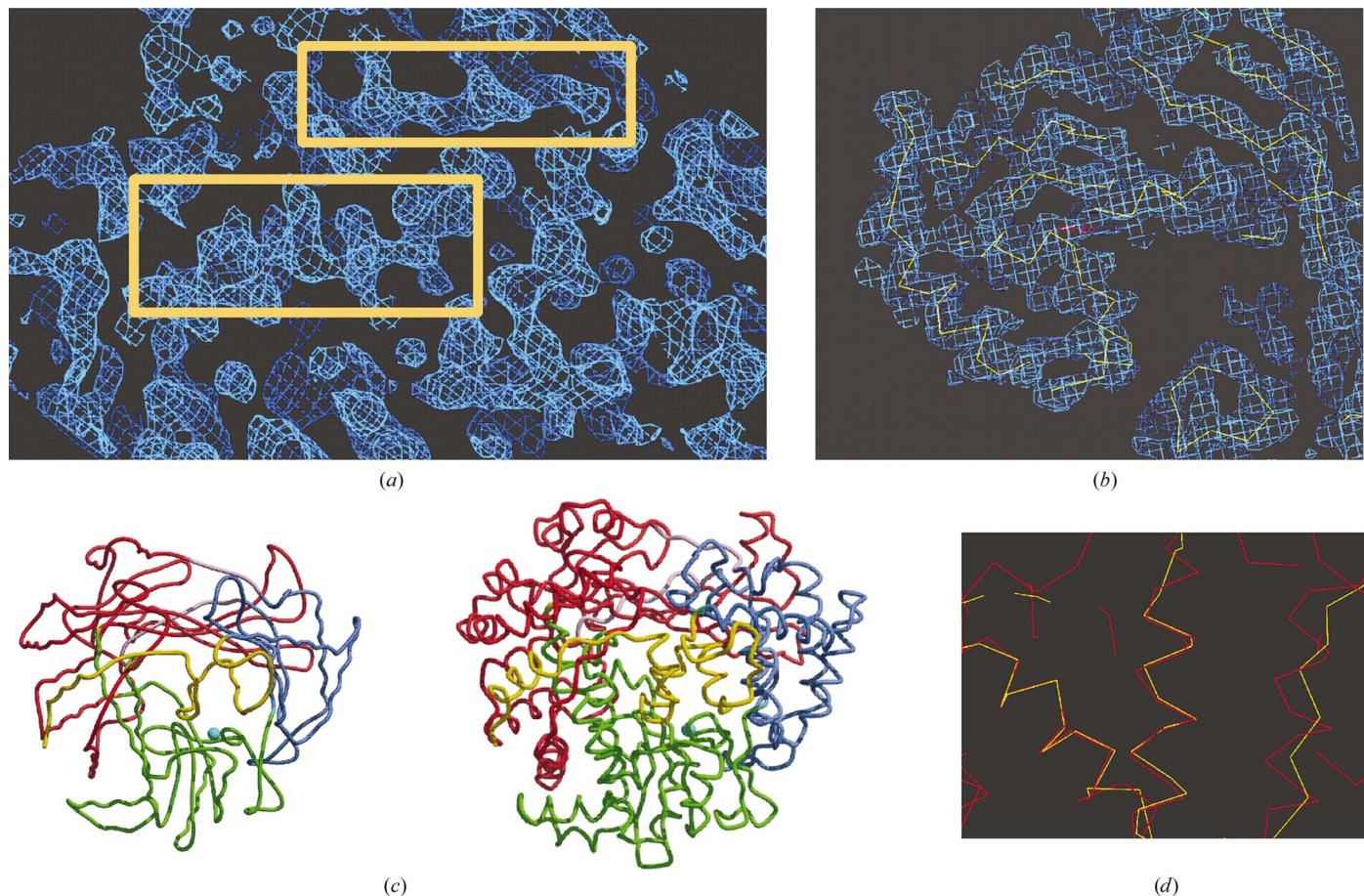
Without recourse to any other external information, *RESOLVE* built 1163 out of 1992 residues in the asymmetric unit, primarily

Table 2
Phase and map analyses.

	Resolution	Mean phase error (Å)	Map correlation coefficient
<i>SOLVE</i>	30–3.5	67.0	0.69
<i>RESOLVE</i>	30–3.5	44.8	0.80
<i>RESOLVE</i> with NCS	30–3.5	34.7	0.85
<i>SOLVE</i>	30–2.9	71.2	0.68
<i>RESOLVE</i>	30–2.9	51.1	0.79
<i>RESOLVE</i> with NCS	30–2.9	37.4	0.89

polyalanine chains (only 37 side chains identified) in the core region of the protein (Fig. 2b). Despite being incomplete, the polypeptide built shows characteristic features of the polymerase and thus the fingers, palm and thumb domains are clearly identifiable (Fig. 2c); although incorrect in some regions (Fig. 2d), the errors could be easily corrected by manual rebuilding.

RESOLVE was also run imposing the known non-crystallographic symmetry (NCS) operators. These were not applied initially because determination of NCS operators from the manganese substructure is not possible; with only one Mn atom per molecule, the orientation of the protein molecules cannot be determined. However, the model built by *RESOLVE* displays structural features from which appropriate non-crystallographic relationships could be derived. Running *RESOLVE* with these NCS operators imposed improved the phases further (mean phase error 35°) and produced a map of excellent

**Figure 2**

(a) Map calculated by *SOLVE* contoured at 1.3σ ; orange boxes indicate interpretable secondary-structure features. (b) *RESOLVE* model: C^α trace of *RESOLVE*-built model (yellow) and corresponding map contoured at 1σ (blue chicken wire). (c) *RESOLVE*-built (left) model and previously solved (right) models showing polymerase structural features: fingers, red; palm, green; thumb, blue. (d) C^α superposition of *RESOLVE* (yellow) and reference (red) models.

quality (correlation coefficient with the reference map = 0.85). With NCS imposed, more residues were built (1302 out of 1992) and additional side chains were identified (88). It is important to note, however, that knowledge of the NCS, whilst facilitating model building and refinement, was not necessary to produce interpretable electron-density maps.

3. Conclusions

Divalent cations such as manganese and calcium are common as protein cofactors, making these cations attractive candidates for *de novo* protein structure determination. Inevitably, because of their small mass, these cations provide only weak anomalous signals, which until recently have been ignored. However, advances in crystallographic hardware and software now make it possible to measure these small differences sufficiently accurately that they can be used for phase determination, obviating the need for the preparation of selenomethionine-labelled proteins or the exhaustive search and preparation of heavy-atom derivatives. The polymerase of bacteriophage $\varphi 6$ provides an excellent test case for this method. The results clearly show that if a highly accurate data set is collected, an anomalous signal as low as 1.2% is sufficient to provide phase information for a crystallographic asymmetric unit containing 225 kDa of protein for a crystal that diffracts only weakly beyond 3 Å. These phases (which use no non-crystallographic symmetry information) are accurate enough to automatically build a model that delineates the characteristic structural features of the polymerase. This is the second report of successful manganese phasing (Ramagopal *et al.*, 2003a) and demonstrates that it can effectively phase large proteins crystallized in low-symmetry space groups with only moderate-resolution X-ray data.

The success of experiments such as that described here and others (Weiss *et al.*, 2001; Ramagopal *et al.*, 2003b) has provided the impetus for optimizing current beamlines for experiments at longer X-ray wavelengths (for example, the protein crystallography beamline XRD1 at the ELETTRA synchrotron, Trieste; Polentarutti *et al.*, 2004) and for the construction of dedicated long-wavelength protein crystallography beamlines (as proposed for the new third-generation UK synchrotron DIAMOND). We expect technical advances at these beamlines to further enhance the power of such methods so that they may take their place alongside SeMet labelling and molecular replacement in the front line of phasing methods.

E. Mancini helped with synchrotron data collection and R. Esnouf and J. Diprose helped with computing. The authors thank the staff at the UK MAD beamline BM14, ESRF, Grenoble. BM14 is supported by the UK Research Councils, the BBSRC, the EPSRC and the

MRC. The work was supported by the Human Frontier Science Project (RGP0320/2001-M), the Academy of Finland [Finnish Centres of Excellence Program 2000–2005 (1202855)] and the Medical Research Council, UK. JMG is supported by the Royal Society and DIS by the Medical Research Council, UK.

References

- Brown, J., Esnouf, R. M., Jones, M. A., Linnell, J., Harlos, K., Hassan, A. B. & Jones, E. Y. (2002). *EMBO J.* **21**, 1054–1062.
- Butcher, S. J., Grimes, J. M., Makeyev, E. V., Bamford, D. H. & Stuart, D. I. (2001). *Nature (London)*, **410**, 235–240.
- Butcher, S. J., Makeyev, E. V., Grimes, J. M., Stuart, D. I. & Bamford, D. H. (2000). *Acta Cryst. D56*, 1473–1475.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst. D50*, 760–763.
- Dauter, Z. & Adamia, D. A. (2001). *Acta Cryst. D57*, 990–995.
- Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
- Dodson, E. (2003). *Acta Cryst. D59*, 1958–1965.
- Gordon, E. J., Leonard, G. A., McSweeney, S. & Zagalsky, P. F. (2001). *Acta Cryst. D57*, 1230–1237.
- Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
- Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
- Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
- Lartigue, A., Gruez, A., Briand, L., Blon, F., Bezzirard, V., Walsh, M., Pernollet, J. C., Tegoni, M. & Cambillau, C. (2004). *J. Biol. Chem.* **279**, 4459–4464.
- Liu, Z.-J., Vysotski, E. S., Chen, C.-J., Rose, J. P., Lee, J. & Wang, B.-C. (2000). *Protein Sci.* **9**, 2085–2093.
- Makeyev, E. V. & Bamford, D. H. (2000a). *EMBO J.* **19**, 124–133.
- Makeyev, E. V. & Bamford, D. H. (2000b). *EMBO J.* **19**, 6275–6284.
- Micossi, E., Hunter, W. N. & Leonard, G. A. (2002). *Acta Cryst. D58*, 21–28.
- Ojala, P. M. & Bamford, D. H. (1995). *Virology*, **207**, 400–408.
- Otwinski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Polentarutti, M., Glazer, R. & Djinovic Carugo, K. (2004). *J. Appl. Cryst.* **37**, 319–324.
- Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003a). *Acta Cryst. D59*, 868–875.
- Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003b). *Acta Cryst. D59*, 1020–1027.
- Rice, L. M., Earnest, T. N. & Brunger, A. T. (2000). *Acta Cryst. D56*, 1413–1420.
- Salgado, P. S., Makeyev, E. V., Butcher, S. J., Bamford, D. H., Stuart, D. I. & Grimes, J. M. (2004). *Structure*, **12**, 307–316.
- Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst. D58*, 1772–1779.
- Terwilliger, T. C. (2000). *Acta Cryst. D56*, 965–972.
- Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst. D55*, 849–861.
- Van Dijk, A. A., Frilander, M. & Bamford, D. H. (1995). *Virology*, **211**, 320–323.
- Walsh, M. A., Evans, G., Sanishvili, R., Dementieva, I. & Joachimiak, A. (1999). *Acta Cryst. D55*, 1726–1732.
- Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
- Weiss, M. S., Sicker, T., Djinovic Carugo, K. & Hilgenfeld, R. (2001). *Acta Cryst. D57*, 689–695.
- Yang, H., Makeyev, E. V. & Bamford, D. H. (2001). *J. Virol.* **75**, 11088–11095.