

# SAD at home: solving the structure of oxalate decarboxylase with the anomalous signal from manganese using X-ray data collected on a home source

Clare E. M. Stevenson,\* Adam Tanner, Laura Bowater, Stephen Bornemann and David M. Lawson

Department of Biological Chemistry, John Innes Centre, Norwich NR4 7UH, England

Correspondence e-mail:  
clare.stevenson@bbsrc.ac.uk

Oxalate decarboxylase (OxdC) from *Bacillus subtilis* is a hexamer containing two manganese ions per 43.6 kDa subunit. A single highly redundant data set collected at a medium resolution of 2 Å on an in-house X-ray source was sufficient to solve the structure by the single-wavelength anomalous diffraction (SAD) method using the anomalous signal from the manganese ions. The experimentally phased electron-density map was of high quality, enabling 96% of the amino-acid sequence to be automatically traced using *ARP/wARP*. Further analysis showed that only half of the original raw data were required for successful structure solution. Manganese currently occurs in approximately 2% of PDB entries. A brief survey suggests that several of these structures could also have been determined using manganese SAD. Moreover, the ability of manganese to substitute for other more commonly occurring divalent metal ions may indicate that the use of Mn SAD could have much wider application.

Received 6 August 2004

Accepted 24 September 2004

## 1. Introduction

Oxalate decarboxylase (OxdC) catalyses the conversion of oxalate to carbon dioxide and formate using molecular oxygen as a cofactor. The crystal structure of the *Bacillus subtilis* enzyme was originally solved using the multiple-wavelength anomalous diffraction (MAD) method with selenomethionine-substituted protein (Anand *et al.*, 2002); more recently, the structure of an alternative conformer was resolved using molecular replacement (Just *et al.*, 2004). These structures confirm OxdC as a member of the cupin superfamily (Khuri *et al.*, 2001); moreover, the presence of two cupin  $\beta$ -barrel domains per subunit places it in the bicupin subfamily. Each of these cupin domains contains a tightly bound manganese ion liganded by the side chains of His and Glu residues that reside within conserved sequence motifs characteristic of the cupins. The two crystal structures are isomorphous and belong to space group *R*32. There is a single 43.6 kDa subunit per asymmetric unit; the biologically relevant hexamer is generated in the crystal by crystallographic twofold and threefold axes of symmetry.

Single-wavelength anomalous diffraction (SAD) has become a successful method for phasing macromolecular structures; for general reviews on the potential of the SAD method, see Dauter *et al.* (2002) and Dodson (2003). To date, there is only one account in the literature of the anomalous signal of manganese being used for SAD phasing. In this case, there was

one Mn per 44 kDa subunit of glucose isomerase (Ramagopal *et al.*, 2003) and synchrotron X-ray data were collected to relatively high resolution (1.5 Å) at wavelengths ranging from 0.98 to 1.54 Å. In this paper, we describe how X-ray data collected from a single crystal using Cu *K* $\alpha$  radiation ( $\lambda = 1.54$  Å) to medium (2 Å) resolution were sufficient to solve the structure of oxalate decarboxylase by manganese SAD.

## 2. Experimental

### 2.1. Diffraction data

Native *B. subtilis* OxdC was overproduced, purified, crystallized and cryoprotected as published elsewhere (Tanner *et al.*, 2001; Just *et al.*, 2004). Moreover, the collection of the data set used for this study (PDB code 1uw8) has previously been described (Just *et al.*, 2004). However, the salient points and additional pertinent details are given here for clarity. The crystals belong to the rhombohedral space group *R*32, with unit-cell parameters  $a = b = 154.7$ ,  $c = 122.8$  Å,  $\alpha = \beta = 90$ ,  $\gamma = 120^\circ$  (hexagonal setting). They contain a single 43.6 kDa monomer per asymmetric unit, giving an estimated solvent content of 62% (Matthews, 1968). X-ray data were recorded at 100 K using a MAR 345 image-plate detector (X-ray Research) mounted on a Rigaku RU-H3RHB rotating-anode X-ray generator (operated at 50 kV and 100 mA) fitted with Osmic confocal optics and a copper target (Cu *K* $\alpha$ ;  $\lambda = 1.54$  Å). Data were initially

**Table 1**  
Summary of X-ray data for OxdC.

Values in parentheses indicate values for the outer resolution shell.

Subset name	Full (M360 + L360)	M90	M180	M270	M360	L360
Resolution range (Å)	39–2.0	26–2.0	26–2.0	26–2.0	26–2.0	39–3.5
Rotation range (°)	360 × 2	90	180	270	360	360
Unique reflections	38069	37966	38030	38049	38055	7257
Redundancy	23.6	5.1	9.7	14.3	19.5	21.4
Completeness	100.0 (99.8)	99.7 (95.3)	99.9 (98.4)	99.9 (99.6)	100.0 (99.8)	100.0 (99.9)
Wilson <i>B</i> value (Å <sup>2</sup> )	17.5	17.8	17.0	17.1	17.6	18.1
<i>R</i> <sub>merge</sub>	0.084 (0.328)	0.077 (0.306)	0.078 (0.314)	0.078 (0.320)	0.080 (0.325)	0.088 (0.116)
<i>I</i> / <i>σ</i> ( <i>I</i> )	34.4 (6.7)	16.4 (3.6)	23.0 (4.8)	27.8 (5.6)	32.8 (6.5)	37.7 (23.7)

**Table 2**  
Summary of SAD phasing for OxdC.

nd, not determined.

Subset name	Max. resolution (Å)	<i>SHELXD</i>		<i>SHELXE</i>			<i>DM</i>		<i>ARP/wARP</i> Residues fitted‡
		Resolution used (Å)	Sites in top 14 peaks	Contrast (original/inverted)	FOM	mapCC†	FOM	mapCC†	
Full	2.0	3.5	2Mn, 8S	0.09/0.82	0.637	0.669	nd	nd	369
M90	2.0	3.5	2Mn, 1S	0.11/0.09	0.170	0.237	0.700	0.359	0
M180	2.0	3.5	2Mn, 4S	0.41/0.09	0.375	0.481	0.696	0.549	367
M270	2.0	3.5	2Mn, 7S	0.80/0.10	0.621	0.628	nd	nd	368
M360	2.0	3.5	2Mn, 8S	0.10/0.81	0.636	0.660	nd	nd	369
L360	3.5	4.0	2Mn	0.11/0.09	0.181	0.003	0.597	0.003	0

† Versus final map, calculated using *OVERLAPMAP* (see main text for further explanation). ‡ In the cases where *DM* was used, this column shows the result obtained from the *DM*-phased map.

collected to a maximum resolution of 2 Å: a total of 360 × 1° images were recorded in a single sweep at the rate of 10 min per image. None of the reflections were flagged as overloads. Subsequently, a 3.5 Å resolution pass was recorded as a single sweep of 180 × 2° images at 150 s per image. The X-ray data were then processed and merged using the *HKL* software package (version 1.97; Otwinowski & Minor, 1997). The resultant data set was of high quality, being virtually complete to the resolution limit of 2 Å, highly redundant (23.6 overall) and relatively strong in the outer resolution shell [*I*/*σ*(*I*) = 6.7 for data in the range 2.03–2.00 Å]. Details of the data-collection and processing statistics are summarized in Table 1.

### 3. SAD phasing

There are two Mn atoms and 10 S atoms per 43.6 kDa monomer of OxdC, although only eight of the S atoms were ordered in the structure that was derived previously from these data (PDB code 1uw8; Just *et al.*, 2004). The *K* X-ray absorption edge of manganese lies at approximately 1.9 Å and at the wavelength used in this experiment (1.54 Å) the imaginary component of the anomalous scattering (*f*'') for manganese is 2.8 electrons. The theoretical anomalous signal arising from the Mn alone can be

estimated from the Bijvoet ratio ( $\langle \Delta F \rangle / \langle F \rangle$ ) to be 1.5% based on two manganese ions and 43.6 kDa of protein per asymmetric unit (Hendrickson & Ogata, 1997). The contribution from sulfur is smaller, having an *f*' value of only 0.56 electrons at this wavelength and an estimated Bijvoet ratio of 0.6% based on eight ordered S atoms per asymmetric unit. The overall Bijvoet ratio for the manganese and sulfur together was subsequently estimated to be 1.6% using the formula for mixed anomalous substructure atoms derived by Olczak *et al.* (2003). This is significantly higher than the minimum value of 0.6% required for successful phasing suggested by Wang assuming error-free data (Wang, 1985).

The graphical user interface *HKL2MAP* (T. Schneider; <http://shelx.uni-ac.gwdg.de/~trs/mad/hkl2map-0.1-tut.pdf>) was used to run several programs from the *SHELX* suite, enabling a smooth transition from the preparation and analysis of scaled diffraction data using *SHELXC*, through substructure solution by *SHELXD* (Schneider & Sheldrick, 2002), to phasing with *SHELXE* (Sheldrick, 2002). In *SHELXC*, the anomalous signal was judged to be significant (*i.e.*  $\langle \text{DANO} \rangle / \langle \sigma I \rangle$  value greater than 1.5) to 3.5 Å resolution. Thus, only the data to 3.5 Å were used to locate the anomalous scatterers in *SHELXD*. The number of phase trials was limited to 100

and the correct solution was found in 95 of these. In the best solution, the top two sites corresponded to the two Mn ions expected in the asymmetric unit and the next eight sites corresponded to the eight ordered S atoms. These sites were then passed to *SHELXE* (without editing or refinement) for phasing, phase extension to 2 Å and density modification (solvent content set to 62%). A clear difference in contrast was seen between the original and inverted-hand enantiomorph, with the latter being the correct one. When used for phasing, this gave a pseudo-free correlation coefficient of 69.5% (defined in [http://shelx.uni-ac.gwdg.de/shelx/shelx\\_de.pdf](http://shelx.uni-ac.gwdg.de/shelx/shelx_de.pdf)).

An experimentally phased map at 2.0 Å resolution showed clear contrast between solvent and protein regions and good connectivity in the latter. Automatic model building was then carried out using *ARP/wARP* (Perrakis *et al.*, 1999) and this was highly successful in placing 369 of the possible 385 residues in four chains. It is notable that the final model of OxdC derived previously from this data set (PDB code 1uw8) contains only a further eight residues. A superposition of these two models gave a root-mean-square deviation of 0.12 Å over all common main-chain atoms. Re-running *SHELXE* with only the two Mn sites was also successful, indicating that the eight S atoms contributed relatively little to the overall phasing. This is not surprising given that the sulfurs added very little to the overall Bijvoet ratio estimated above. The phasing statistics are summarized in Table 2.

### 4. The importance of redundancy and resolution

In order to establish the minimum quantity of raw data required for a successful structure solution by SAD, the structure solution was repeated using subsets of the original data set. By running the *STRATEGY* option in *MOSFLM* (Leslie, 1992), it was predicted that a 90° wedge of data corresponding to images 141–230 of the 2 Å pass would give virtually complete anomalous data. This subset was reprocessed and denoted M90 (for 90° medium-resolution subset). Subsequently, images 51–230 were reprocessed as M180, 51–320 as M270 and 1–360 as M360. Finally, the complete 3.5 Å resolution pass was reprocessed as L360, denoting 360° low-resolution subset. The statistics of all subsets are summarized in Table 1 alongside those for the full data set. These subsets were analysed by the *SHELX* suite in the same manner as the complete data set, with the

exception of L360, where a 4 Å resolution cutoff was imposed in *SHELXD*. In all cases, *SHELXD* was able to find the two Mn sites, but all eight sulfurs were only found using M360. The figures of merit (FOMs) after *SHELXE* were respectable for only the M270 and M360 subsets, but these could be 'improved' significantly for M90 and M180 by further density modification using the program *DM* (Cowtan, 1994). The *SHELXE* phases were sufficient for automated model building by *ARP/wARP* for only the M270 and M360 subsets, although the *DM*-modified phases also enabled successful chain tracing for M180. The *DM* phases for M90, however, were not good enough for automated model building, suggesting that the relatively high FOM was misleading. With the L360 data, *SHELXD* only found the two Mn sites and the phasing statistics from *SHELXE* were poor. The FOM was significantly increased by running *DM* and this enabled a map to be calculated that showed contrast between protein and solvent regions, but little connectivity. Re-running *DM* on the L360 phases, incorporating phase extension to 2 Å (using structure factors from the complete data set), also failed to yield an interpretable map, confirming that the starting phases were poor. The combination of lower resolution (3.5 versus 2.0 Å) and weaker data (the exposure per degree was eight times shorter for this pass) were undoubtedly the main reasons for the failure of phasing with the L360 subset. Nevertheless, radiation damage could also have been a factor, since the low-resolution pass was collected after some 60 h of exposure. However, an inspection of scale and temperature factors in the *SCALEPACK* output suggested that there was no significant decay.

As an alternative measure of the quality of the SAD phases, the program *OVERLAPMAP* (Brändén & Jones, 1990) was used to compare all the SAD-phased maps to a map produced using phases calculated from the final OxD model (PDB code 1uw8). The resultant correlation coefficients were a good gauge of map quality, with a value of greater than 0.5, indicating a SAD-phased map that could be automatically traced with *ARP/wARP*. All phasing statistics are summarized in Table 2.

## 5. Discussion

It was first demonstrated in 1981 that the anomalous signal of sulfur was sufficient to solve the structure of a small protein (that of crambin; Hendrickson & Teeter, 1981). However, it was not until comparatively

recently that more challenging macromolecular structures were being routinely solved by the SAD technique using relatively weak anomalous signals (Jaskólski, 1996; Dauter *et al.*, 2000, 2002; Liu *et al.*, 2000; de Graaff *et al.*, 2001; Gordon *et al.*, 2001; Weiss *et al.*, 2001; Lemke *et al.*, 2002; Li *et al.*, 2002; Debreczeni *et al.*, 2003; Dodson, 2003; Ramagopal *et al.*, 2003; Usón *et al.*, 2003; Yang *et al.*, 2003). Clearly, these advances have been driven by a combination of improved instrumentation and the development of powerful software algorithms to process and analyse the X-ray data. Nevertheless, the success of this approach is critically dependent on the availability of good-quality crystals that yield X-ray data sets that are both accurately recorded and highly redundant. In this paper, we describe the structure solution of *B. subtilis* oxalate decarboxylase by SAD using the signal from Mn measured in a data set collected to medium resolution using Cu  $K\alpha$  radiation. Phasing was successful owing to a combination of strong diffraction, high crystal symmetry, relatively high solvent content and well ordered Mn sites. Subsequent analysis showed that very high data redundancy was not essential in this case as the data were of high quality. In general, particularly with less favourable cases (*e.g.* where there is weaker diffraction, lower symmetry or lower solvent content), it would be prudent to collect excess data to ensure successful phasing, assuming radiation damage does not become significant. Nevertheless, given the relatively slow speed of in-house data collection, a structure solution could easily be obtained whilst the experiment is still in progress and thus one could be sure that enough data had been collected before terminating data collection. In this study, we also showed that a weaker data set of lower resolution (the low-resolution pass) was sufficient for locating the Mn sites, but was unable to yield an interpretable map.

To our knowledge, this is the first time that manganese SAD phasing has been used to solve a macromolecular structure using in-house data. In an attempt to predict whether other known structures could theoretically have been solved by manganese SAD using Cu  $K\alpha$  radiation, we searched the PDB for entries containing manganese. This yielded a total of 608 hits that could be reduced to a non-redundant set of 238 by retaining only one representative structure for each protein (*i.e.* excluding variants of the same structure using a 90% sequence-identity cutoff). Next, we reasoned that crystals yielding 1.6 Å

resolution structures would give strong 2 Å resolution data sets using a rotating-anode generator. Thus, we selected a non-redundant set of structures that had been determined to 1.6 Å resolution or better: this applied to a total of 29 entries. For the subsequent analysis, we retained only Mn sites with unit occupancy and temperature factors lower than the overall average for the structure (*i.e.* well ordered sites). This excluded a further six structures that contained only 'unsuitable' Mn sites. Then Bijvoet ratios at Cu  $K\alpha$  wavelength were calculated for the remaining 23 structures, based on only well ordered and fully occupied Mn sites, giving values in the range 0.7–2.0%. According to the 0.6% minimum value suggested by Wang (1985), all of these structures could theoretically be solved with error-free data. However, based on our experience with real data from OxD, if we assume a lower cutoff of 1.5%, this still leaves 14 structures that could potentially be determined by manganese SAD, given highly redundant data collected on a home X-ray source (these were 1d5n, 1e9g, 1fi2, 1i0b, 1k4o, 1mvo, 1nki, 1o6l, 1o9i, 1pl4, 1r2m, 1ro2, 1s95 and 1ues). A more rigorous, less conservative, analysis of the PDB would undoubtedly find significantly more examples. Furthermore, the manganese(II) ion is intermediate between magnesium and calcium in size and can therefore substitute for either in biological systems (Fraústo da Silva & Williams, 1993). Indeed, it has seen extensive use as a spectroscopic probe for both metals (Mildvan & Cohn, 1970; Feig, 2000; Reed & Poyner, 2000). Since magnesium currently occurs in 1786 PDB entries and calcium in 2370, the use of Mn SAD to solve macromolecular crystal structures may potentially have much wider application.

This work was funded by the BBSRC. We would like to thank I. Usón for assistance with the structure solution, A. Olczak for help with the calculation of Bijvoet ratios and the referees of this manuscript for constructive comments. CEMS is grateful for support to attend the MAX-INF Workshop on Phasing and Refinement in Barcelona (March 2004).

## References

- Anand, R., Dorrestein, P. C., Kinsland, C., Begley, T. P. & Ealick, S. E. (2002). *Biochemistry*, **41**, 7659–7669.
- Brändén, C. & Jones, T. A. (1990). *Nature (London)*, **343**, 687–689.
- Cowtan, K. (1994). *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **31**, 34–38.

- Dauter, Z., Dauter, M. & Dodson, E. (2002). *Acta Cryst.* **D58**, 494–506.
- Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* **D56**, 232–237.
- Debreczeni, J. E., Bunkoczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). *Acta Cryst.* **D59**, 688–696.
- Dodson, E. (2003). *Acta Cryst.* **D59**, 1958–1965.
- Feig, A. L. (2000). *Manganese and Its Role in Biological Processes*, Vol. 37, edited by A. Sigel & H. Sigel, pp. 157–182. New York: Marcel Dekker.
- Fraústo da Silva, J. J. R. & Williams, R. J. P. (1993). *The Biological Chemistry of the Elements*. Oxford: Clarendon Press.
- Gordon, E. J., Leonard, G. A., McSweeney, S. & Zagalsky, P. F. (2001). *Acta Cryst.* **D57**, 1230–1237.
- Graaff, R. A. de, Hilge, M., van der Plas, J. L. & Abrahams, J. P. (2001). *Acta Cryst.* **D57**, 1857–1862.
- Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
- Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
- Jaskólski, M. (1996). *Acta Cryst.* **D52**, 1075–1081.
- Just, V. J., Stevenson, C. E., Bowater, L., Tanner, A., Lawson, D. M. & Bornemann, S. (2004). *J. Biol. Chem.* **279**, 19867–19874.
- Khuri, S., Bakker, F. T. & Dunwell, J. M. (2001). *Mol. Biol. Evol.* **18**, 593–605.
- Lemke, C. T., Smith, G. D. & Howell, P. L. (2002). *Acta Cryst.* **D58**, 2096–2101.
- Leslie, A. G. W. (1992). *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **26**.
- Li, S., Finley, J., Liu, Z. J., Qiu, S. H., Chen, H., Luan, C. H., Carson, M., Tsao, J., Johnson, D., Lin, G., Zhao, J., Thomas, W., Nagy, L. A., Sha, B., DeLucas, L. J., Wang, B.-C. & Luo, M. (2002). *J. Biol. Chem.* **277**, 48596–48601.
- Liu, Z. J., Vysotski, E. S., Chen, C. J., Rose, J. P., Lee, J. & Wang, B.-C. (2000). *Protein Sci.* **9**, 2085–2093.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Mildvan, A. S. & Cohn, M. (1970). *Adv. Enzymol. Relat. Areas Mol. Biol.* **33**, 1–70.
- Oleczak, A., Cianci, M., Hao, Q., Rizkallah, P. J., Raftery, J. & Helliwell, J. R. (2003). *Acta Cryst.* **A59**, 327–334.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
- Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* **D59**, 868–875.
- Reed, G. H. & Poyner, R. R. (2000). *Manganese and Its Role in Biological Processes*, Vol. 37, edited by A. Sigel & H. Sigel, pp. 183–207. New York: Marcel Dekker.
- Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* **D58**, 1772–1779.
- Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.
- Tanner, A., Bowater, L., Fairhurst, S. A. & Bornemann, S. (2001). *J. Biol. Chem.* **276**, 43627–43634.
- 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.
- Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
- Weiss, M. S., Sicker, T., Djinic Carugo, K. & Hilgenfeld, R. (2001). *Acta Cryst.* **D57**, 689–695.
- Yang, C., Pflugrath, J. W., Courville, D. A., Stence, C. N. & Ferrara, J. D. (2003). *Acta Cryst.* **D59**, 1943–1957.