electronic reprint

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of superoxide dismutase from *Pyrobaculum aerophilum* presents a challenging case in molecular replacement with multiple molecules, pseudo-symmetry and twinning

Sangho Lee, Michael R. Sawaya and David Eisenberg

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of superoxide dismutase from *Pyrobaculum aerophilum* presents a challenging case in molecular replacement with multiple molecules, pseudo-symmetry and twinning

Sangho Lee, Michael R. Sawaya and David Eisenberg*

UCLA-DOE Institute for Genomics and Proteomics, Molecular Biology Institute, Howard Hughes Medical Institute, and Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095-1570, USA

Correspondence e-mail: david@mbi.ucla.edu

Received 4 June 2003
Accepted 9 September 2003

PDB Reference: *P. aerophilum* superoxide dismutase, 1p7g, r1p7gsf.

The crystal structure of superoxide dismutase from the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum* was determined by molecular replacement at 1.8 Å resolution. The structure determination was made especially challenging by the large number of molecules (24) in the asymmetric unit, the presence of a pseudo-crystallographic twofold operator close to a twinning operator and the inability to detect twinning by conventional means. Molecular replacement proceeded at low resolution in pseudo (apparent) space group $P3_212$ and was facilitated by examination of the self-rotation function and native Patterson map. Refinement, however, stalled at an R factor of 40% when high-resolution data were included. Expanding to the lower symmetry space group $P3_2$ decreased R (to 22%) and $R_{\rm free}$ (to 26%), but not by as much as expected for the quality of data. Finally, despite the apparent lack of evidence from conventional twinning tests [*i.e.* plots of the second moment of I and $N(Z)$ distributions], a twinning operator was included in the refinement, lowering R and $R_{\rm free}$ to 16.2 and 21.7%, respectively. The early detection of twinning appears to have been masked by a deviation in the expected intensity distribution caused by the presence of non-crystallographic translational symmetry. These findings suggest the importance of testing twinning operators in cases where pseudo-translational symmetry can explain negative results from conventional twinning tests. The structure reveals a tetrameric assembly with 222 symmetry, similar to superoxide dismutase structures from other organisms. The current structural model represents the metal-free state of the enzyme.

1. Introduction

Molecular replacement becomes non-trivial for a crystal containing multiple molecules in the asymmetric unit. Locating a large number of molecules in the asymmetric unit can be difficult even for search models with high sequence homology, for example higher than 50%, to the query sequence. Several challenging cases have been reported of molecular replacement with multiple molecules in the asymmetric unit (Oh, 1995; Chantalat *et al.*, 1996; Bernstein & Hol, 1997).

An extra level of difficulty is incurred in the molecular-replacement procedure when twinning is present. Twinning refers to abnormal crystal growth in which separate crystal domains, twinning domains, are oriented by a symmetry operation, the twinning operation (Yeates, 1997). Merohedral twinning, in which the twinning domains are superimposable in three dimensions with twinning fraction α, defined as the fractional volume of domains in the second orientation, cannot be detected by examining diffraction patterns, but can be detected by examining the intensity statistics using the

Merohedral Crystal Twinning Server ('Twinning Server'; Yeates, 1997). Perfect merohedral twinning occurs when $\alpha \simeq 0.5$ and partial merohedral twinning occurs when $0 < \alpha < 0.5$. In most cases, molecular replacement can be performed successfully with only slight complications (Yeates, 1997). One simply needs to detect that twinning is present and use the apparent space-group symmetry in the rotation function and the true space-group symmetry in the translation function.

When twinning exists yet all conventional means of detecting twinning falsely yield negative results, structure determination can be especially confusing. In this particular case, molecular replacement initially appeared to be successful. When a non-crystallographic symmetry (NCS) operation nearly coincides with a twinning operation (as is the case here), the effects of twinning can be ignored at low resolution and all steps of molecular replacement can be performed in a pseudo space group which is related to the real space group by a twinning operation (Rees & Lipscomb, 1980), as performed here. However, crystallographic refinement at high resolution will stall with a high R factor. When conventional means of detecting twinning give no indication that twinning is present, it becomes difficult to explain why the R factor is so high or to know how to proceed in the refinement. As we have found, it becomes critical to include a twinning operator in the crystallographic refinement in cases where failure to detect twinning by conventional means can be explained by deviations in the intensity distribution caused by non-crystallographic translational symmetry.

Superoxide dismutases (SODs) catalyze the disproportionation of the biologically toxic superoxide radical to oxygen and hydrogen peroxide in a metal-dependent mechanism (Fridovich, 1995). Structurally, SODs can be classified into three superfamilies according to the metal cofactors: Cu/ZnSODs, Fe or MnSODs and NiSODs. Fe or MnSOD structures from several organisms are known. Two FeSOD structures from hyperthermophiles, those from *Aquifex pyrophilus* (Lim *et al.*, 1997) and *Sulfolobus solfataricus* (Ursby *et al.*, 1999), have been reported.

Pyrobaculum aerophilum is a hyperthermophilic crenarchaeon whose optimal growth temperature is 373 K by aerobic respiration as well as by dissimilatory nitrate reduction (Volkl *et al.*, 1993). The genome sequence of *P. aerophilum* has been determined (Fitz-Gibbon *et al.*, 2002). SOD from *P. aerophilum* (*Pa*SOD) was identified in the course of genome sequencing (Volkl *et al.*, 1996). It shows affinities for both Mn and Fe, with the higher affinity being for Mn (Whittaker & Whittaker, 2000). Therefore, *Pa*SOD belongs to the Fe or MnSOD superfamily.

Here, we report the crystal structure of superoxide dismutase from *P. aerophilum*, determined by molecular replacement. The primary sequence of *Pa*SOD shows a high level of identity to the sequences of other SODs of known structure. However, structure determination by molecular replacement was hampered by the large number of molecules in the asymmetric unit, pseudo-symmetry and partial merohedral twinning. The location of 12 molecules in the asymmetric unit was guided by examination of the self-rotation function and the native Patterson map in pseudo space group $P3_212$. Pseudo-symmetry, in which a local twofold axis happened to be very close to a crystallographic twofold axis, was revealed during refinement, leading to an expansion from space group $P3_212$ to $P3_2$. Partial merohedral twinning, which was masked by NCS, was detected at later stages of refinement, resulting in a significant drop in the R factor. In this report, the procedures used to overcome the challenges in the structure determination of *Pa*SOD are discussed.

2. Materials and methods

2.1. Expression and purification of *Pa*SOD

The cDNA encoding SOD was amplified by polymerase chain reactions from the genomic DNA of *P. aerophilum*. The SOD clone was inserted into a pQE30 vector (Qiagen) to encode an N-terminal histidine-tagged fusion protein. The plasmid pQE30-SOD was expressed in *Escherichia coli*. At an optical density of 0.5 at 600 nm, the cells were induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 m*M*. The cells were harvested 2 h after induction.

The cell pellets were resuspended in sonication buffer (20 m*M* HEPES pH 7.8, 300 m*M* NaCl, 20 m*M* imidazole). Lysozyme was added to a final concentration of 50 μg ml^{-1}, DNase I to a final concentration of 50 μg ml^{-1} and phenylmethylsulfonyl fluoride to a final concentration of 1 m*M*. The resuspended cells underwent one freeze/thaw cycle followed by sonication to complete cell lysis. The lysate was centrifuged at 35 000*g* for 30 min to remove cell debris. The soluble fraction was heated at 338 K for 20 min and cooled to room temperature. The supernatant after centrifugation at 35 000*g* for 30 min was filtered and applied to a HiTrap chelating column (Pharmacia) charged with 50 m*M* nickel sulfate. *Pa*SOD was eluted with a gradient of 50–500 m*M* imidazole in 20 m*M* HEPES pH 7.8 containing 300 m*M* NaCl. Fractions containing *Pa*SOD were pooled and dialyzed in 20 m*M* HEPES pH 7.8, 300 m*M* NaCl, 5 m*M* EDTA and then dialyzed in 20 m*M* HEPES pH 7.8, 300 m*M* NaCl for crystallization trials.

Selenomethionyl-substituted *Pa*SOD was expressed and purified as described above with the following modifications. Cells bearing the plasmid pQE30-SOD were grown in M9 minimal medium with ampicillin, selenomethionine and glycerol at 310 K. After induction with IPTG at a final concentration of 1 m*M*, the cells were grown for an additional 8 h. At each step of purification, β-mercaptoethanol was added to a final concentration of 5 m*M* in order to prevent the oxidation of selenomethionine.

2.2. Crystallization

The native and selenomethionyl-substituted *Pa*SODs were concentrated using a Centricon (Amicon) with a molecular-weight cutoff of 10 000 Da to 26 and 22 mg ml^{-1}, respectively, for crystallization trials. The proteins were crystallized using

Table 1
Data collection and processing statistics.

Values in parentheses are for the highest resolution shell.

Space group	$P3_2$	$P3_212$
Unit-cell parameters (Å)		
a	163.43	
b	163.43	
c	172.17	
No. molecules in the AU	24	12
Beam source	NSLS†	
Resolution range (Å)	36.4–1.8	
Total reflections	2013670	2034250
Unique reflections	465126	238024
Completeness (%)	97.6 (82.6)	98.4 (88.3)
$R_{\rm merge}$‡ (%)	7.6 (30.2)	9.1 (33.5)
$\langle I/\sigma(I)\rangle$§	20.1 (3.3)	26.4 (4.6)

† National Synchrotron Light Source beamline X8C at Brookhaven National Laboratory. ‡ $R_{\rm merge} = \sum_{hkl}\sum_i |I(hkl)_i - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i \langle I(hkl)\rangle$. § $\langle I/\sigma(I)\rangle$

the hanging-drop vapor-diffusion method at 295 K. Crystallization conditions were found using Wizard screen kits (Emerald BioStructures). The best crystals were grown in reservoir solutions containing 10–15%(*v*/*v*) polyethylene glycol 3000, 0.1 *M* HEPES pH 7.5 and 0.15 *M* calcium acetate with Al's oil (Hampton Research) on top of the reservoir solutions. Trigonal crystals appeared within a week with dimensions of 0.3 × 0.2 × 0.2 mm. Crystals were flash-frozen in liquid nitrogen at 93 K.

2.3. Data collection and processing

Crystals of native and selenomethionyl-substituted *Pa*SOD diffracted X-rays to resolution limits of 2.5 and 1.8 Å, respectively. Data were collected at synchrotron beamline X8C at the National Synchrotron Light Source at Brookhaven National Laboratory using a Quantum 4 CCD area detector. A standard four-wavelength multi-wavelength anomalous dispersion (MAD) experiment was conducted on the selenomethionyl-substituted *Pa*SOD crystals. Both the selenomethionyl and the native data sets were processed using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). Both were processed in the point group *P*312, but the *c* unit-cell parameter was approximately doubled in the selenomethionyl derivative (native, $a = b = 160.54$, $c = 83.84$ Å; selenomethionyl derivative, $a = b = 163.43$, $c = 171.17$ Å). Ultimately, the structure was solved by molecular replacement using the peak-wavelength (0.9785 Å) data set from the selenomethionyl derivative. The data-collection statistics for this 'peak' data set are shown in Table 1. To aid in detection of pseudo-symmetry, statistics were calculated in both point groups *P*3 and *P*312 (see §3.4).

2.4. Phasing and refinement

Initial molecular-replacement trials were attempted using *AMoRe* (Navaza, 1994), *EPMR* (Kissinger *et al.*, 1999) and *CNS* (Brünger *et al.*, 1998), without success. The correct molecular-replacement solution was obtained using *REPLACE* (Tong & Rossmann, 1990). The programs *SOLVE* (Terwilliger & Berendzen, 1999) and *Shake-and-Bake* (Weeks & Miller, 1999) were used for initial efforts in phasing by multi-wavelength anomalous dispersion (MAD) experiments.

All stages of refinement were performed using *CNS*. In the early stages of refinement, strict NCS constraints were applied. NCS map averaging was performed using *RAVE* (Kleywegt & Jones, 1994) and *X-PLOR* (Brünger, 1992). The 24-fold averaged electron-density map was used to build side chains in *O* (Jones, 1978). In the later stages of refinement, NCS constraints were released and individual chains were refined without any restraints. Rigid-body refinement was performed after finding the initial solution. At early stages of refinement, the solution found in space group $P3_212$ did not allow NCS averaging to work. Once the pseudo-symmetry was detected, the space group was changed to $P3_2$. Partial merohedral twinning was implemented in the final refinement.

2.5. Detection of twinning

Detection and analysis of twinning were performed using the Merohedral Crystal Twinning Server (Yeates, 1997) and the program *TRUNCATE* from *CCP*4 (Collaborative Computational Project, Number 4, 1994). Refinement of the twinning fraction was performed using *SHELXL* (Sheldrick & Schneider, 1997).

3. Results and discussion

3.1. Crystallization

Initial conditions produced small crystals in less than 1 d. The small crystals diffracted poorly, probably because of rapid nucleation. Use of oil to control the vapour-diffusion rate has been reported previously (Chayen, 1997). The problem of rapid nucleation was resolved by adding Al's oil (Hampton Research) to the top of the reservoir solutions. The use of oil reduced the nucleation rate by more than twofold and improved the crystal quality, *i.e.* larger sizes and higher resolution diffraction limits were obtained.

3.2. Testing twinning in the pseudo space group $P3_212$

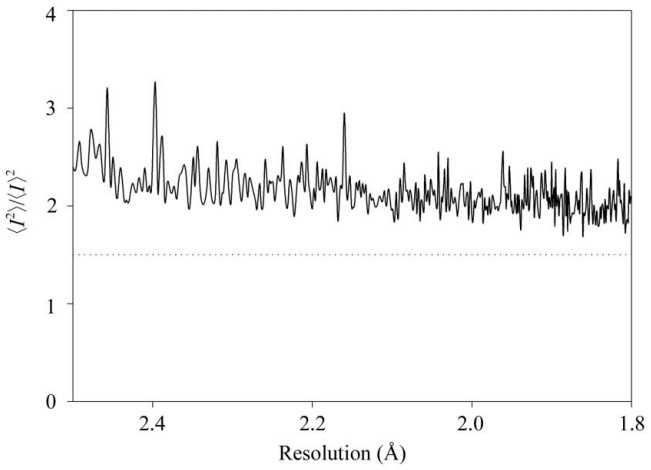
Since trigonal space groups are prone to twinning, both native and selenomethionyl data sets (in the pseudo space group $P3_212$) were tested for twinning using the Twinning Server (http://www.doe-mbi.ucla.edu/Services/Twinning; Yeates, 1997). No strong evidence for twinning was found by the Twinning Server. Perfect twinning-test results are shown for the selenomethionyl data set (Fig. 1*a*). Perfect twinning is indicated when the value of $\langle I^2\rangle/\langle I\rangle^2$ for acentric reflections is 1.5. As shown in Fig. 1(*a*), the calculated values of $\langle I^2\rangle/\langle I\rangle^2$ are closer to 2, the expected value for untwinned data. The $N(Z)$ plot given in the output of *TRUNCATE* (Collaborative Computational Project, Number 4, 1994) showed no evidence of twinning (Fig. 1*b*), consistent with the results from the Twinning Server. Furthermore, the crystal morphology did not appear to be twinned (Fig. 1*c*), nor did the Matthews coefficient suggest an impossibly low $V_{\rm M}$ as had been the case in other twinned crystals. Therefore, all phasing efforts, including molecular replacement and multi-wavelength anomalous

dispersion, were performed in pseudo space groups $P3_112$ or $P3_212$ (rather than $P3_1$ or $P3_2$ as would have been indicated if twinning were detected).

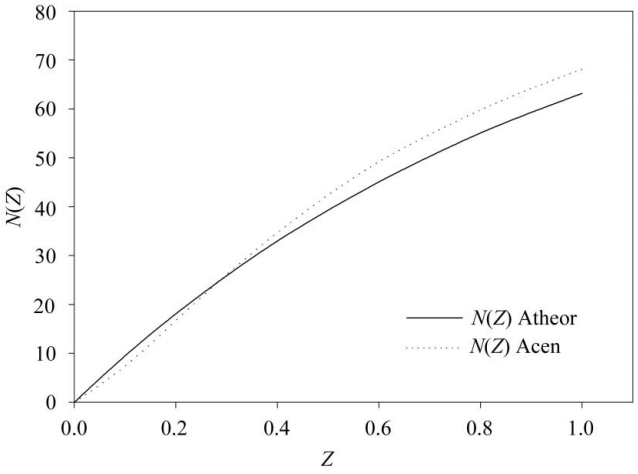
3.3. Finding a correct molecular-replacement solution

Initial molecular-replacement trials using the native data set were unsuccessful. Several molecular-replacement programs were tried, including *AMoRe* (Navaza, 1994), *EPMR* (Kissinger *et al.*, 1999) and *CNS* (Brünger *et al.*, 1998). The crystal structures of three SOD homologs were used as search models: MnSOD from *Thermus thermophilus* (Ludwig *et al.*, 1991), Y34F mutant human MnSOD (Guan *et al.*, 1998) and FeSOD from *S. sulfactaricus* (Ursby *et al.*, 1999). The sequence identities of these search models to *Pa*SOD are high: 41, 44 and 55%, respectively, giving a favourable outlook for finding a solution. However, exhaustive combinatorial searches using all search models in a variety of oligomeric states, including dimers and tetramers as well as monomers, failed to produce a convincing solution. Limited success was achieved with the program *REPLACE* (Tong & Rossmann, 1990) using a dimer as a search model. Comparison of translation-function peak heights in space groups $P3_112$ and $P3_212$ indicated the former as the correct choice of enantiomorph. All six expected molecules could be located in the asymmetric unit; however, the model could not be refined below an *R* factor of 48% using 2.5 Å resolution data. In retrospect, this obstacle could probably have been overcome had we been able to detect the presence of twinning in the native data set and treated the data accordingly.

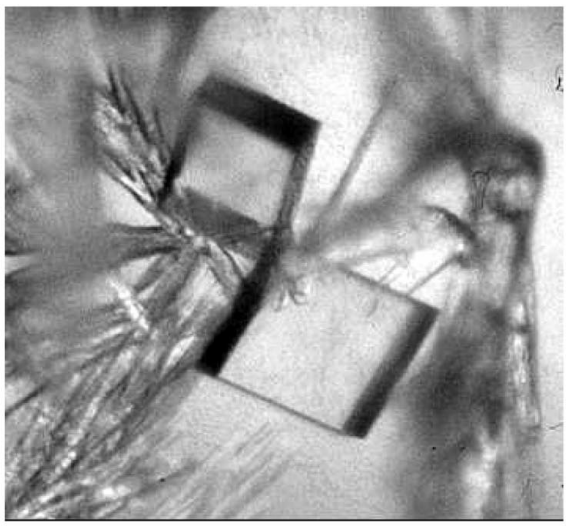
Reasoning that difficulties in refinement were caused by an insurmountable degree of model bias, we pursued the possibility of solving the SOD structure using the selenomethionyl derivative and MAD techniques, but again our attempts met with failure. Compared with the native crystal, the selenomethionyl crystal contained a doubled *c* unit-cell dimension and twice as many molecules in the asymmetric unit. A change in handedness of the threefold screw (from $P3_112$ to $P3_212$) was postulated and later confirmed to accompany the doubled unit-cell dimension, reasoning that the additional molecules in the selenomethionyl crystal could approximate the 3_1 screw present in the native crystal. 36 selenium sites were expected given space group $P3_212$ with 12 molecules in the asymmetric unit (three methionine residues per molecule excluding the N-terminal residue). Despite excellent quality data (1.8 Å resolution), the programs *SOLVE* (Terwilliger & Berendzen, 1999) and *Shake-and-Bake* (Weeks & Miller, 1999) failed to find a sufficient number of selenium sites to produce an



(*a*)



(*b*)



(*c*)

Figure 1
Tests for perfect twinning in the pseudo space group $P3_212$. Perfect twinning was tested on the data set processed in the pseudo space group $P3_212$ using (*a*) the Twinning Server (Yeates, 1997) and (*b*) the $N(Z)$ plot given in the output from *TRUNCATE* (Collaborative Computational Project, Number 4, 1994). In (*a*), the values of $\langle I^2\rangle/\langle I\rangle^2$ for acentric reflections were calculated in thin resolution shells. Each thin resolution shell contains approximately 400 reflections. The expected values $\langle I^2\rangle/\langle I\rangle^2$ are 2.0 for untwinned data and 1.5 for perfectly twinned data. The values of $\langle I^2\rangle/\langle I\rangle^2$ in the observed data set are closer to 2.0, indicating no apparent sign of twinning. In (*b*), $N(Z)$ refers to the cumulative distribution function and Z, defined as $I/\langle I\rangle$, refers to the intensity relative to the mean intensity. $N(Z)$ values for acentric data are plotted for theoretical data (solid line) and for real data (dotted line). The $N(Z)$ values for twinned data are expected to be significantly lower than those for the theoretical data. No evidence for twinning is found from this $N(Z)$ plot. (*c*) Morphology of *Pa*SOD crystals. No morphological features implying twinning are observed.

interpretable electron-density map. Again, in retrospect, the presence of undetected twinning (see §3.4 and §3.5) is believed to be the culprit. We later found that the anomalous difference Fourier map calculated with phases from the final refined model was able to reveal the correct positions of the selenium sites.

In a final attempt at structure determination, we decided to return to molecular replacement, this time using the selenomethionyl data set in the hope that the superior quality of this data set over the native data set would lead to success. A useful molecular-replacement solution was indeed found using conventional rotation and translation functions as implemented in the program *REPLACE* (Tong & Rossmann, 1990). The search model consisted of a dimer of FeSOD from *S. solfataricus* (Ursby *et al.*, 1999) in which the side chains of non-identical residues were replaced by alanine and regions of insertions or deletions were removed. The presence of six dimers in the asymmetric unit was suggested by a Matthews coefficient (V_M) of 2.14 Å^3 Da^{-1} (Matthews, 1968) and an estimated solvent content of 46% assuming a $P3_212$ cell. The six dimers were located one by one in a sequential manner.

Table 2
Initial molecular-replacement solution.

θ_1 (°)	θ_2 (°)	θ_3 (°)	x†	y†	z†	CC‡	R§
91	4	266	0.2111	0.0926	0.2396	23.02	50.58
			0.5458	0.7583	0.2400	22.50	50.88
			0.5458	0.7708	0.7400	29.70	47.14
			0.8792	0.4375	0.7400	36.25	47.80
34	3	324	0.2125	0.1042	0.7400	39.54	47.80
			0.8792	0.4292	0.2400	41.45	49.45

† All translations are in fractional coordinates. ‡ CC, correlation coefficient. CC = $\sum_h (F_h^{obs} - \langle F_h^{obs}\rangle)(F_h^{calc} - \langle F_c^{calc}\rangle)/[\sum_h (F_h^{obs} - \langle F_h^{obs}\rangle)^2 \sum_h (F_h^{obs} - \langle F_h^{calc}\rangle)^2]^{1/2}$. § R factor = $100\sum_h |F_h^{obs} - kF_h^{calc}|/\sum_h F_h^{obs}$, where $k = \sum_h F_h^{obs}/\sum_h F_h^{calc}$.

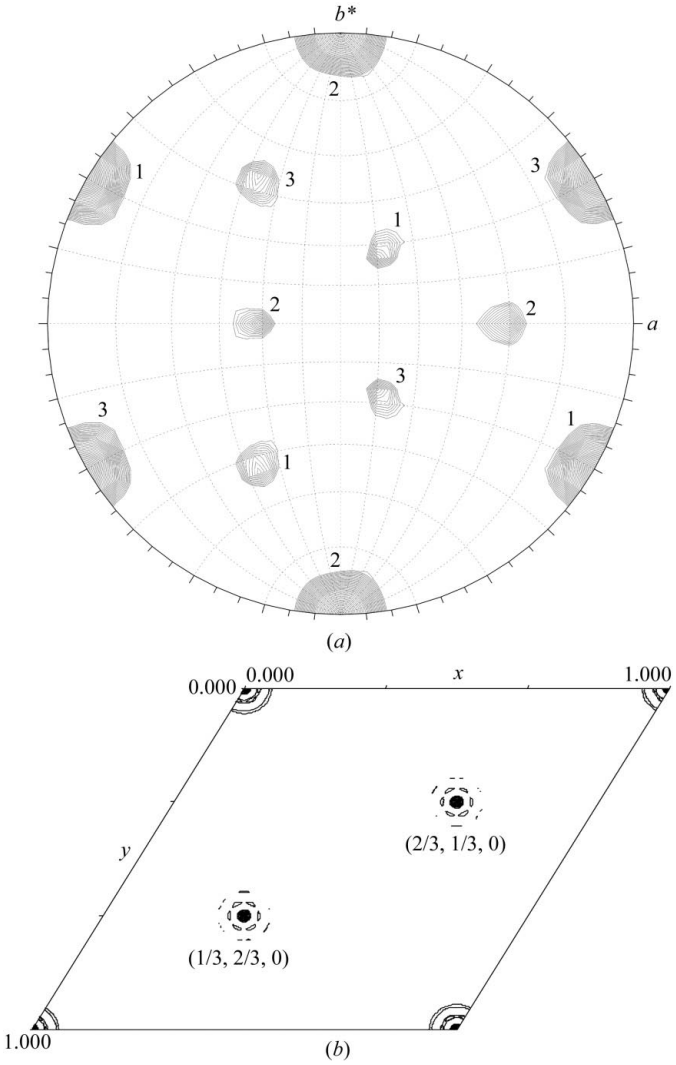


Figure 2
Self-rotation function and native Patterson map in the pseudo space group $P3_212$. Shown are the maps for (*a*) the self-rotation function and (*b*) the native Patterson map of the selenomethionyl data set. Both functions were calculated in the pseudo space group $P3_212$. In (*a*), the self-rotation function was calculated with data in the resolution range 9–4.5 Å using *GLRF* (Tong & Rossmann, 1990) and a Patterson vector radius of 32 Å. Three sets of orthogonal twofold rotation axes ($\kappa = 180°$) are labeled with different numbers. In (*b*), the native Patterson map was calculated with data in the resolution range 20–4 Å using *XFFT* from the *XtalView* suite (McRee, 1999) and contoured at 5σ. Peaks off the origin are labeled. The labeled peaks have a height of 126.7σ.

The multi-step molecular-replacement procedure was guided by NCS information obtained from the self-rotation function and the native Patterson map. Specifically, the self-rotation function revealed the presence of three sets of three mutually orthogonal twofold rotation axes ($\kappa = 180°$ section), suggesting the conservation of a tetrameric SOD assembly with 222 symmetry (Fig. 2*a*) (it was later shown that the crystallographic twofold does indeed generate the tetramer from the dimer used as the search model). Furthermore, the native Patterson map revealed strong translational symmetry peaks at (1/3, 2/3, 0), (2/3, 1/3, 0), (1/3, 2/3, 1/2), (2/3, 1/3, 1/2) and (0, 0, 1/2) (Fig. 2*b*), suggesting that the first dimer positioned in the asymmetric unit could be related to the remaining five dimers by these translation vectors.

Both the rotation and translation functions (performed in $P3_212$) could be readily interpreted. The cross-rotation function was calculated using data in the resolution range 8–3 Å and a Patterson vector radius of 32 Å. The top peak in the rotation function was 6.9σ above the mean and was well separated from other peaks. The translation function was calculated using data in the resolution range 20–3.5 Å and yielded a top peak of 8σ above the mean. Fixing the first dimer in the unit cell yielded a correlation coefficient of 23.0% and an R factor of 50.6% (Table 2). As predicted, the second, third and fourth dimers were found to be related to the first dimer by translational symmetry corresponding to the peaks found in the native Patterson map. The fifth and sixth dimers could not be confidently located in the translation function, suggesting that the remaining two dimers had a slightly different orientation to the first four. The orientations of these last two dimers were determined by performing a fine-grid search of the cross-rotation function with a grid size of 0.5°. When all six dimers, corresponding to 12 molecules in the space group $P3_212$, were located, the correlation coefficient was 41.5%, with an R factor of 49.5% (Table 2). As more

dimers were found, the correlation coefficients increased. The *R* factors, however, fluctuated as more dimers were located (Table 2).

3.4. Detection of pseudo-symmetry

After several rounds of initial model building and refinement, the *R* and R_{free} values were stalled at 40 and 46%, respectively, at 2.3 Å resolution. Calculations revealed that the r.m.s. distances among C^{α} atoms of 12 molecules in the asymmetric unit ranged from 0.5 to 1.6 Å. These large r.m.s. differences severely limited the usefulness of NCS-averaged maps and suggested that the twofold rotational axis in the space group $P3_212$ might instead be a NCS twofold axis in the space group $P3_2$ in nearly the same orientation. Although little support for this hypothesis can be gained from comparing R_{merge} values in space group $P3_2$ (7.6%) with $P3_212$ (9.1%) (Table 1), expansion of the model to space group $P3_2$ brought a significant drop in *R* and R_{free} to 22 and 26%, respectively. Furthermore, expansion of the model to lower symmetry followed by rigid-body refinement enabled us to apply NCS averaging successfully in map calculations. Calculations subsequently revealed that the NCS twofold axis is misoriented from the crystallographic twofold axis by a rotation of 2°. Such a subtle misorientation of NCS twofold axis would not have been detected at the stage of finding the molecular-replacement solution.

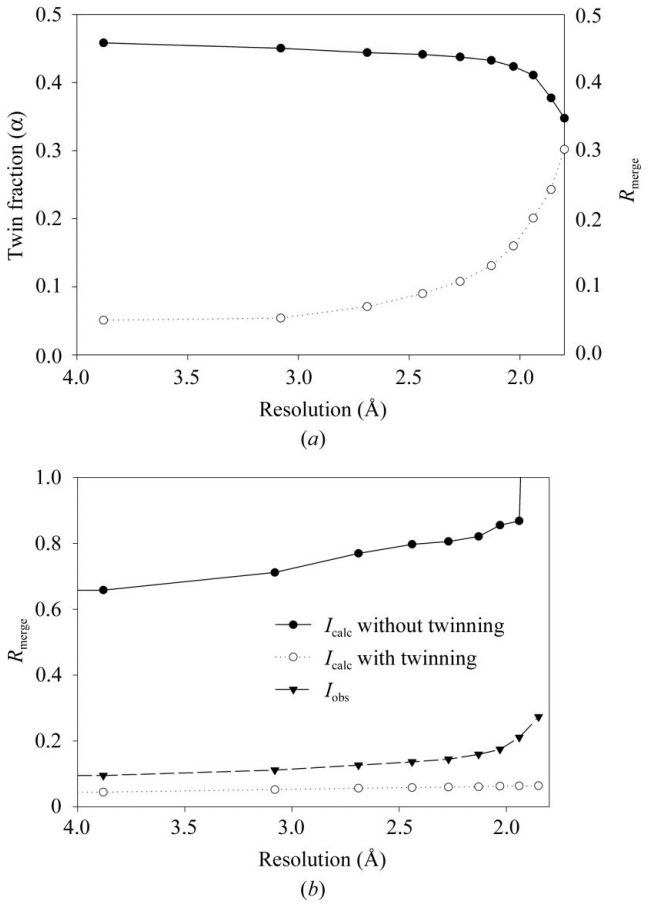


Figure 3
Data suggesting partial twinning. (*a*) Estimated twinning fraction α *versus* R_{merge} in resolution shells in space group $P3_2$ (filled circles) calculated using the Twinning Server (Yeates, 1997). We assumed partial twinning with the twinning operation being twofold rotation along *a** and *b** axes in space group $P3_2$. R_{merge} values (open circles) in the same resolution shells as those used to calculate the twinning fractions were calculated using *SCALEPACK* (Otwinowski & Minor, 1997). (*b*) R_{merge} values *versus* resolution shells for twinning-related reflections in space group $P3_2$, shown with R_{merge} values on I_{calc} without implementing twinning (filled circles), on I_{calc} with implementing twinning (open circles) and I_{obs} (filled reverse triangles).

3.5. Detection of partial twinning in space group $P3_2$ in the presence of NCS

After several further rounds of refinement, *R* values were once again stalled, leading us to re-examine the possibility of twinning in space group $P3_2$. We sought a twinning operator that might be employed to lower the *R* factor and perhaps explain the similarity in R_{merge} values in space groups $P3_2$ and $P3_212$. The partial twinning test (Yeates, 1997) yielded an estimated twinning fraction of 0.45 and a twinning operator corresponding to the crystallographic twofold symmetry axis of $P3_212$. The partial twinning test is not by itself a proof of twinning. It cannot distinguish between NCS and twinning operations if the NCS axis happens to be close to a potential twinning operator (Yeates & Fam, 1999), as it is in this case.

To determine whether the data set in the space group $P3_2$ was really twinned, we calculated estimated twinning fractions α as a function of resolution shells using the Twinning Server and compared the estimated twinning fractions with the R_{merge} values in the corresponding resolution shells (Fig. 3*a*). If two reflections are related by a twin law, the extent of disagreement in the intensities of these two reflections will stay constant regardless of resolution. On the other hand, the extent of disagreement will become larger in higher resolution shells if the same two reflections are related by NCS. The twinning fraction in the highest resolution shell is 24% lower than the twinning fraction in the lowest resolution shell. Taken alone, this figure may imply that the data are not twinned. However, the modest decrease observed in the twinning fraction values at high resolution may arise from errors in intensity measurements (larger R_{merge} values in the higher resolution shells) rather than from NCS. Thus, this twinning test also appears to be ambiguous.

A more convincing line of evidence in support of the presence of partial twinning comes from a comparison of R_{merge} values between twin-related reflections for both I_{calc} and I_{obs} (Fig. 3*b*). As we have seen, R_{merge} on I_{obs} is approximately 10–25% across the resolution range. However, R_{merge} on I_{calc} prior to implementation of the twinning operator is approximately 65–85%. Clearly, NCS is insufficient to explain the low R_{merge} observed in the twin-related I_{obs}. Implementation of the twinning operator brings R_{merge} on I_{calc} down to approximately 5%. The improved ability to model the observed data with the implementation of the twinning operator provides clear support for the presence of twinning.

With caution, we monitored the changes in *R* and R_{free} after introducing partial twinning into the refinement processes.

Introduction of partial twinning into the refinement in *CNS* reduced the *R* values significantly, reinforcing the argument above regarding the partial twinning of the data set in space group $P3_2$. Refinement in *SHELXL* (Sheldrick & Schneider, 1997), which employs a different algorithm dealing with twinning (Herbst-Irmer & Sheldrick, 1998), also showed decreased *R* values, confirming partial twinning of the data set. Implementing a twinning fraction of 0.463, refined using *SHELXL*, caused a dramatic decrease in *R* and $R_{\rm free}$ by 6 and 5%, respectively. The drastic decrease in the *R* and $R_{\rm free}$ values merely from introducing twinning without any model rebuilding strongly supported the presence of partial twinning.

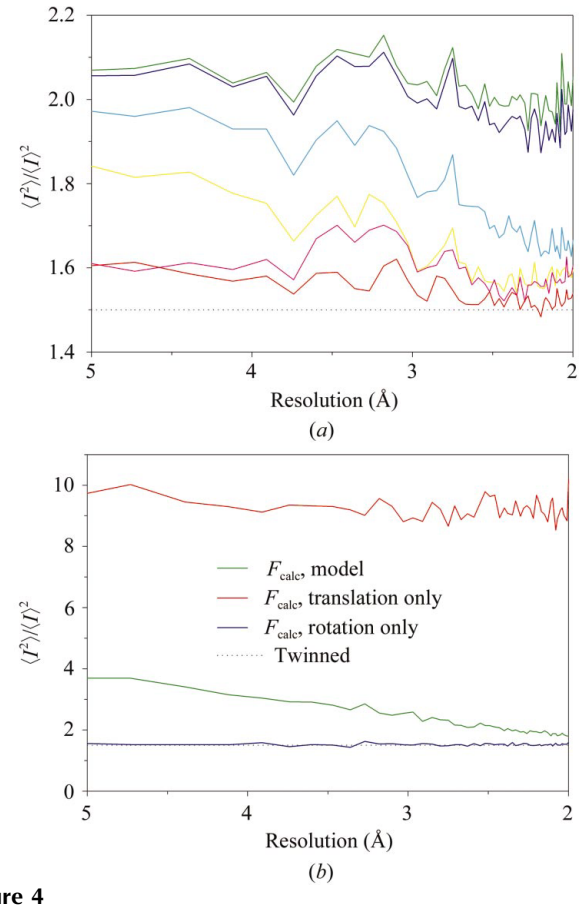


Figure 4
The effect of NCS symmetry elements on masking the detection of twinning. (*a*) Effect of close alignment of the NCS twofold axis with the twinning operator on the detection of twinning using conventional tests. $\langle I^2\rangle/\langle I\rangle^2$ values for acentric reflections are shown in the resolution range 2–5 Å using $F_{\rm calc}$ from models with the NCS twofold axis oriented closely by 0 (green), 0.1 (blue), 0.25 (cyan), 0.5 (yellow), 1 (magenta) and 5° (red), respectively, from the twinning operator (twofold rotation along a^* and b^* axes in the space group $P3_2$). The value of $\langle I^2\rangle/\langle I\rangle^2$ for twinned data, 1.5, is indicated as a dotted line (black). The masking of twinning by the presence of a NCS twofold axis is less significant when the NCS twofold axis is more than 1° away from the twinning operator. (*b*) Effect of pure NCS translational symmetry on masking the detection of twinning. $\langle I^2\rangle/\langle I\rangle^2$ values are shown in the resolution range 2–5 Å using $F_{\rm calc}$ from the finally refined model (green), from a model incorporating NCS translational symmetry only (red) and from a model having NCS rotational symmetry only (blue). The value of $\langle I^2\rangle/\langle I\rangle^2$ for twinned data, 1.5, is indicated as a dotted line (black). The presence of NCS translational symmetry can offset twinning significantly by pseudo-centering (see text for details).

However, the refined twinning fraction of 0.463 is almost the value for perfect twinning. Because the twinning fraction value was very close to 0.5, detwinning was not attempted. Final refinement implementing twinning using *CNS* led to final values of 16.2 and 21.7% for *R* and $R_{\rm free}$, respectively, after releasing all NCS restraints. We believe that the relatively large difference between *R* and $R_{\rm free}$ is acceptable considering that the NCS restraints were released and crystal twinning is present.

Detection of twinning by examination of intensity distribution statistics is generally considered to be robust and efficient, so why did these tests fail in this particular case? Had we been able to detect twinning in the native crystal, its structure might have been successfully refined by implementation of a twinning operator, sparing us the time and expense of producing the selenomethionyl derivative for MAD experiments. There are three possible explanations for failure to detect twinning by analysis of intensity distribution: (i) strong diffraction anisotropy, (ii) close alignment of the NCS twofold with the twinning operator and (iii) the presence of translational NCS elements. No strong anisotropy was detected using the falloff procedure implemented in *TRUNCATE*. To test the effect of close alignment of the NCS rotation axis with the twinning operator on the twinning test, the efficiency of the twinning test was estimated by plotting the second moment of $I_{\rm calc}$ for models oriented with different degrees of alignment of the NCS twofold axis with the twinning operator (Fig. 4*a*). Conventional twinning tests fail to detect twinning only when the NCS twofold axis is less than 1° away from the twinning operator. Our NCS operator was over 2° away from the twinning operator. Therefore, the close alignment of the NCS twofold axis with the twinning operator should not be the major culprit for the failure to detect twinning in this case.

The presence of NCS translational symmetry appears to be the predominant factor in explaining the lack of detection of twinning by intensity statistics. In a model where the NCS axis was rotated away from the twinning operator but strict translational symmetry was maintained, a plot of the second moment of *I* shows a strong deviation towards the appearance of being 'untwinned' (Fig. 4*b*). We can explain this deviation by the effect of pseudo-centering on the intensity distribution. When translational NCS happens to be close to the rational unit-cell fractions 1/2, 1/3, 1/4, pseudo-centering will cause a large fraction of the reflection intensities to appear as very weak or extinct. The severity of this effect is governed by the precision of the pseudo-centering and by the purity of the translational symmetry (*i.e.* the lack of additional rotation). In this case, pseudo-centering at (2/3, 1/3, *z*) and (1/3, 2/3, *z*) strengthens the class of reflections $h + k = 3n$ while nearly extinguishing other reflections. This deviation effectively counters the influence of twinning, which characteristically decreases the number of very weak and very strong reflections. The overall effect is to mask the detection of twinning by statistical methods. In the observed data from *Pa*SOD, the purity of the NCS translational symmetry was just sufficient to counter the effects of twinning on the intensity distribution, giving the appearance of an untwinned crystal with $\langle I^2\rangle/\langle I\rangle^2$

Table 3
Statistics for atomic refinement of metal-free *Pa*SOD.

Values in parentheses are for the highest resolution shell.

Resolution range (Å)	36.45–1.8
No. of protein atoms	41127
Average *B* factor, protein atoms (Å^2)	21.0
No. of other molecules	
Acetate	19
β-Mercaptoethanol	12
Water	2536
R.m.s.d.	
Bond length (Å)	0.007
Bond angle (°)	1.4
R_{free}	0.217 (0.263)
R	0.162 (0.241)

close to 2. We now know that it is critical to include a twinning operator in the crystallographic refinement in cases where failure to detect twinning by conventional means can be explained by deviations in the intensity distribution caused by non-crystallographic translational symmetry. We therefore recommend calculating a native Patterson map to find whether NCS translational symmetry is present before attempting a routine twinning analysis.

3.6. Structure description

The current model contains 4223 residues in 24 chains. The final refinement statistics are shown in Table 3: 88.9% of residues in the structure fall into the most favoured regions in the Ramachandran plot, 9.1% into additionally allowed regions, 1.8% into generously allowed regions and 0.1% into disallowed regions. Only six residues are in the disallowed regions, but these show reasonable electron densities (data not shown). *Cis*-proline 23 is conserved in 21 of 24 chains in the asymmetric unit.

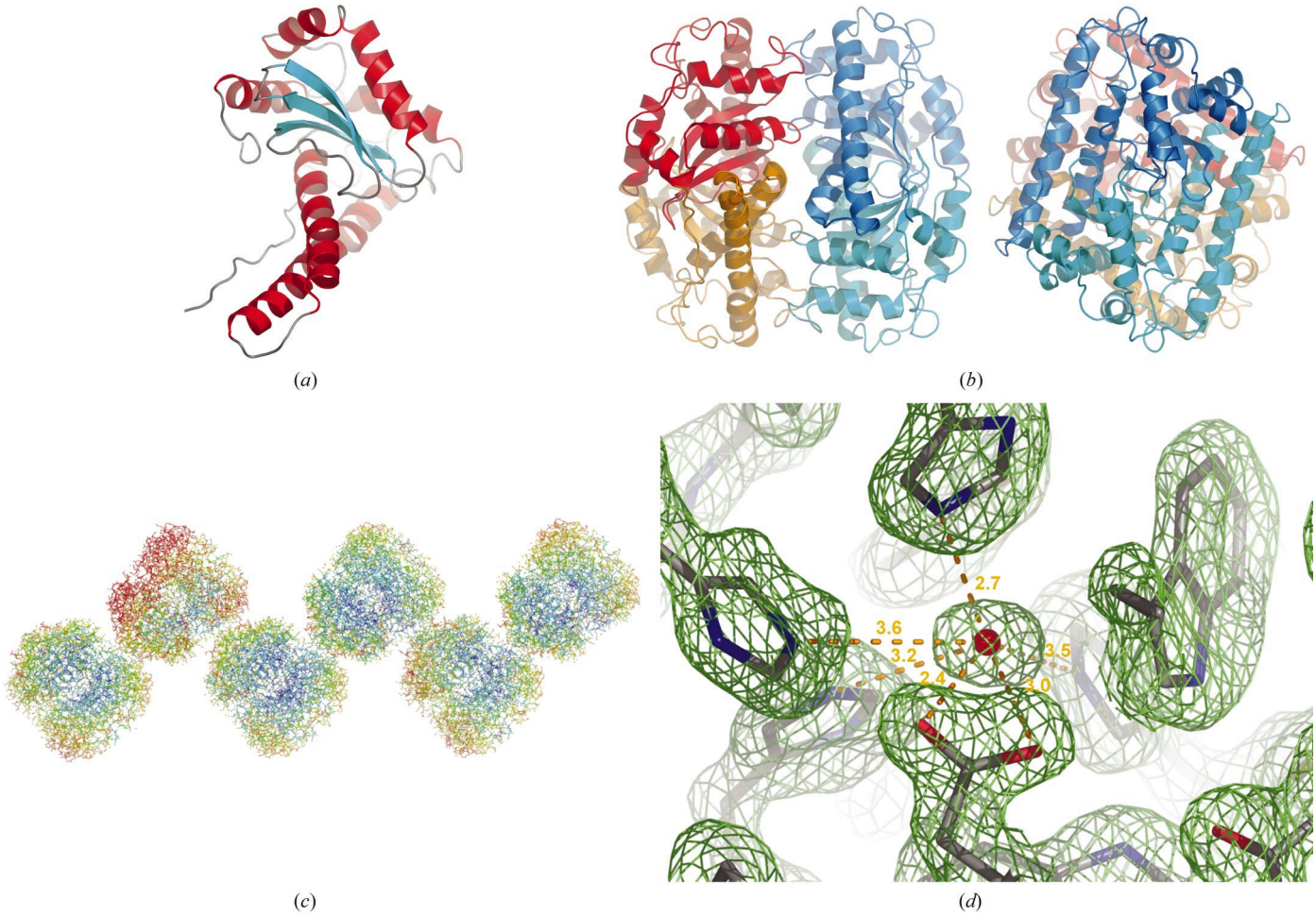


Figure 5
The structure of *Pa*SOD. (*a*) Ribbon representation of the monomeric structure of *Pa*SOD. Helices are colored red and strands cyan. The monomeric structure of *Pa*SOD shows an *α*/*β* fold, typically found in many other Mn or FeSOD structures. (*b*) Ribbon representation of the tetrameric structure of *Pa*SOD. Each chain is colored differently. Two different views are shown. 222 point symmetry is clearly observed for the tetrameric structure of *Pa*SOD. This figure was prepared using *PyMOL* (DeLano Scientific). (*c*) Arrangement of *Pa*SOD subunits in the asymmetric unit. The whole asymmetric unit containing 24 molecules is shown. The asymmetric unit is colored by *B* factor, with red being the highest and blue being the lowest. Notice that two molecules show higher *B* factors than the rest of the asymmetric unit. (*d*) Electron-density map of the active site of *Pa*SOD. The $2F_o - F_c$ map (green) around the active site of *Pa*SOD contoured at 1.5σ is shown. Hydrogen-bonding distances (orange, dotted lines) from the water molecule (red) to the atoms in the conserved active-site residues are shown in Å. Notice that no metal ion is found in the active site of the current structural model. This figure was prepared using *PyMOL* (DeLano Scientific).

The *Pa*SOD monomer structure reveals an α/β fold (Fig. 5*a*). The overall structure assumes an L shape. The N-terminus has a long loop parallel to the first long helix. The two long N-terminal helices are connected by a short loop. The C-terminal part of the *Pa*SOD monomer structure consists of short helices and strands. The *Pa*SOD monomer structure is similar to known Fe or MnSOD structures.

The structure of *Pa*SOD tetramer, which is a biologically active unit (Whittaker & Whittaker, 2000), shows 222 point symmetry (Fig. 5*b*). Like *S. sulfataricus* SOD, *Pa*SOD forms a compact tetramer (Fig. 5*b*). *Pa*SOD forms a tetramer *in vitro*, as revealed by size-exclusion chromatography (data not shown).

The asymmetric unit contains six tetramers. Although the r.m.s. deviation of C^{α} atoms among the 24 molecules is relatively small (0.5–0.7 Å), the *B*-factor distribution reveals that only two of 24 chains show higher *B*-factor distributions (Fig. 5*c*). β-Mercaptoethanol molecules, included in the protein solution, are located at local twofold rotational axes throughout the asymmetric unit.

Careful examination of the electron-density map around the active site revealed no evidence of the presence of a metal ion (Fig. 5*d*). A water molecule was found in the active site. Inclusion of $MnCl_2$ in the reservoir solutions failed to produce crystals with a manganese ion bound at the active site. Incorporation of a manganese ion into the active site was achieved by boiling according to a previously published protocol (Whittaker & Whittaker, 2000). The structure of the metal-bound *Pa*SOD was solved by molecular replacement (manuscript in preparation). The current structural model represents the 'metal-free' state of *Pa*SOD.

4. Conclusions

The structure determination of *Pa*SOD presents a challenging case of molecular replacement with the presence of multiple molecules in the asymmetric unit, pseudo-crystallographic symmetry and partial twinning. The most frustrating element in the structure determination was the inability to detect twinning at an early stage owing to the presence of NCS translational symmetry. Although a crude molecular-replacement solution was achievable without considering twinning, refinement of the model stalled when high-resolution data were included. Since conventional means of detecting twinning gave no indication that twinning was present, much time was wasted in the pursuit of structure determination by an alternate means, *i.e.* MAD experiments. As we have found, it is important to test twinning operators in cases where NCS translational symmetry can explain negative results from conventional twinning tests. Simple inspection of the native Patterson map indicates the presence of translational symmetry. A significant drop in R and $R_{\rm free}$ with the implementation of a twinning operator can serve as a justification for its use.

We thank Dr Hanjing Yang for providing us with the plasmid encoding *Pa*SOD, Dr Edward M. Marcotte for contributions to this work at its early stages, Dr. Duilio Cascio for help in data collection and processing, and Drs Todd Yeates and Peter Muller for discussion about twinning. This work was supported by funds from DOE, HHMI and NIH.

References

Bernstein, B. E. & Hol, W. G. J. (1997). *Acta Cryst.* D**53**, 756–764.
Brünger, A. T. (1992). *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR*. New Haven, CT, USA: Yale University Press.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Chantalat, L., Wood, S. D., Rizkallah, P. & Reynolds, C. D. (1996). *Acta Cryst.* D**52**, 1146–1152.
Chayen, N. E. (1997). *J. Appl. Cryst.* **30**, 198–202.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Fitz-Gibbon, S. T., Ladner, H., Kim, U.-J., Stetter, K. O., Simon, M. I. & Miller, J. H. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 984–989.
Fridovich, I. (1995). *Annu. Rev. Biochem.* **64**, 97–112.
Guan, Y., Hickey, M. J., Borgstahl, G. E. O., Hallewell, R. A., Lepock, J. R., O'Connor, D., Hsieh, Y., Nick, H. S., Silverman, D. N. & Tainer, J. A. (1998). *Biochemistry*, **37**, 4722–4730.
Herbst-Irmer, R. & Sheldrick, G. M. (1998). *Acta Cryst.* B**54**, 443–449.
Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.
Kleywegt, G. J. & Jones, T. A. (1994). *Proceedings of the CCP4 Study Weekend. From First Map to Final Model*, edited by S. Bailey, R. Hubbard & D. Waller, pp. 59–66. Warrington: Daresbury Laboratory.
Lim, J.-H., Yu, Y. G., Han, Y. S., Cho, S.-J., Ahn, B.-Y., Kim, S.-H. & Cho, Y. (1997). *J. Mol. Biol.* **270**, 259–274.
Ludwig, M. L., Metzger, A. L., Pattride, K. A. & Stallings, W. C. (1991). *J. Mol. Biol.* **219**, 335–358.
McRee, D. E. (1999). *Practical Protein Crystallography*. San Diego: Academic Press.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Oh, B.-H. (1995). *Acta Cryst.* D**51**, 140–144.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Rees, D. C. & Lipscomb, W. N. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 277–280.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **276**, 319–343.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Tong, L. & Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 783–792.
Ursby, T., Adinolfi, B. S., Al-Karadaghi, S., De Vendittis, E. & Bocchini, V. (1999). *J. Mol. Biol.* **286**, 189–205.
Volkl, P., Huber, R., Drobner, E., Rachel, R., Burggraf, S., Trincone, A. & Stetter, K. (1993). *Appl. Environ. Microbiol.* **59**, 2918–2926.
Volkl, P., Markiewicz, P., Baikalov, C., Fitz-Gibbon, S., Stetter, K. O. & Miller, J. H. (1996). *Nucleic Acids Res.* **24**, 4373–4378.
Weeks, C. M. & Miller, R. (1999). *J. Appl. Cryst.* **32**, 120–124.
Whittaker, M. M. & Whittaker, J. W. (2000). *J. Biol. Inorg. Chem.* **5**, 402–408.
Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.
Yeates, T. O. & Fam, B. C. (1999). *Structure*, **7**, R25–R29.