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
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 \( P3_12 \) 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 \( R_{\text{free}} \) (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 \( R_{\text{free}} \) 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 \( \alpha \), 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
Superoxide dismutases (SODs) catalyze the disproportionalization 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 *P*3212. 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 *P*3212 to *P*3. 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 mM. The cells were harvested 2 h after induction.

The cell pellets were suspended in sonication buffer (20 mM HEPES pH 7.8, 300 mM NaCl, 20 mM 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 mM. The resuspended cells underwent one freeze/thaw cycle followed by sonication to complete cell lysis. The lysate was centrifuged at 35 000g 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 000g for 30 min was filtered and applied to a HiTrap chelating column (Pharmacia) charged with 50 mM nickel sulfate. *Pa*SOD was eluted with a gradient of 50–500 mM imidazole in 20 mM HEPES pH 7.8 containing 300 mM NaCl. Fractions containing *Pa*SOD were pooled and dialyzed in 20 mM HEPES pH 7.8, 300 mM NaCl, 5 mM EDTA and then dialyzed in 20 mM HEPES pH 7.8, 300 mM 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 mM, the cells were grown for an additional 8 h. At each step of purification, β-mercaptoethanol was added to a final concentration of 5 mM 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.

<table>
<thead>
<tr>
<th>Space group</th>
<th>$P3_2$</th>
<th>$P3_{21}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit-cell parameters (Å)</td>
<td>163.43</td>
<td>163.43</td>
</tr>
<tr>
<td>$a$</td>
<td>163.43</td>
<td>163.43</td>
</tr>
<tr>
<td>$b$</td>
<td>172.17</td>
<td>172.17</td>
</tr>
<tr>
<td>$c$</td>
<td>163.43</td>
<td>163.43</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>$\beta$</td>
<td>160.54</td>
<td>160.54</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>110.72</td>
<td>110.72</td>
</tr>
<tr>
<td>No. molecules in the AU</td>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td>Beam source</td>
<td>NSLS²</td>
<td>NSLS²</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>36.4±1.8</td>
<td>36.4±1.8</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.6 (82.6)</td>
<td>98.4 (88.3)</td>
</tr>
<tr>
<td>$R_{merge}$ (%)</td>
<td>7.6 (30.2)</td>
<td>9.1 (33.5)</td>
</tr>
<tr>
<td>$(I/\sigma(I))$</td>
<td>20.1 (3.3)</td>
<td>26.4 (4.6)</td>
</tr>
</tbody>
</table>

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

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 PaSOD 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 PaSOD 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 $P3_{21}$, 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 $P3$ and $P3_{21}$ (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. CNS 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 $P3_{21}$ 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 CCP4 (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 $P3_{21}$

Since trigonal space groups are prone to twinning, both native and selenomethionyl data sets (in the pseudo space group $P3_{21}$) were tested for twinning using the Twinning Server (http://www.biochem.duke.edu/TwinningServer). Perfect twinning-test results are shown for the selenomethionyl data set (Fig. 1a). Perfect twinning is indicated when the value of $(I^2)/\langle I^2 \rangle$ is 1.5. As shown in Fig. 1(a), the calculated values of $(I^2)/\langle I^2 \rangle$ are closer to 2, the expected value for unwinned data. The $N(Z)$ plot given in the output of TRUNCATE (Collaborative Computational Project, Number 4, 1994) showed no evidence of twinning (Fig. 1b), consistent with the results from the Twinning Server. Furthermore, the crystal morphology did not appear to be twinned (Fig. 1c), nor did the Matthews coefficient suggest an impossibly low $V_M$ as had been the case in other twinned crystals. Therefore, all phasing efforts, including molecular replacement and multi-wavelength anomalous

Lee et al. • Superoxide dismutase 2193

Lee et al. • Superoxide dismutase 2193
dispersion, were performed in pseudo space groups $P3_1$ or $P3_2$ (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 (Brüner 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 $P3_1$ and $P3_2$ 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 $P3_1$ to $P3_2$) was postulated and later confirmed to accompany the doubled unit-cell dimension, reasoning that the additional molecules in the selenomethionyl crystal could approximate the 31 screw present in the native crystal. 36 selenium sites were expected given space group $P3_2$ 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

---

**Figure 1**
Tests for perfect twinning in the pseudo space group $P3_1$. Perfect twinning was tested on the data set processed in the pseudo space group $P3_1$ 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 methionyl data set in the hope that the superior quality of this return to molecular replacement, this time using the selenomethionyl data set, was able to reveal the correct positions of the selenium sites. Fourier map calculated with phases from the final refined model was able to reveal the correct positions of the selenium sites.

Table 2
Initial molecular-replacement solution.

<table>
<thead>
<tr>
<th>θ1 (°)</th>
<th>θ2 (°)</th>
<th>θ3 (°)</th>
<th>x†</th>
<th>y†</th>
<th>z†</th>
<th>CC†</th>
<th>R§</th>
</tr>
</thead>
<tbody>
<tr>
<td>91</td>
<td>4</td>
<td>266</td>
<td>0.211</td>
<td>0.0926</td>
<td>0.2396</td>
<td>23.02</td>
<td>50.58</td>
</tr>
<tr>
<td>0.5458</td>
<td>0.7583</td>
<td>0.2400</td>
<td>22.50</td>
<td>50.88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5458</td>
<td>0.7708</td>
<td>0.7400</td>
<td>29.70</td>
<td>47.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8792</td>
<td>0.4375</td>
<td>0.7400</td>
<td>36.25</td>
<td>47.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>3</td>
<td>324</td>
<td>0.2125</td>
<td>0.1042</td>
<td>0.7400</td>
<td>39.54</td>
<td>47.80</td>
</tr>
<tr>
<td>0.8792</td>
<td>0.4292</td>
<td>0.2400</td>
<td>41.45</td>
<td>49.45</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† All translations are in fractional coordinates. † CC, correlation coefficient. CC = $\sum (F_{h} - F_{h})/(\sum F_{h} - F_{h})$, where $F_{h}$ is the observed structure factor, $F_{h}$ is the calculated structure factor, and $k = \sum F_{h}$. § R factor = 100$\sum |F_{h} - F_{h}|/\sum F_{h}$.

Figure 2
Self-rotation function and native Patterson map in the pseudo space group P3212. 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 P3212. 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 (κ = 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 (κ = 180° section), suggesting the conservation of a tetrameric SOD assembly with 222 symmetry (Fig. 2a) (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. 2b), 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 P3212) 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 P3212, 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 refine-
ment, the $R$ and $R_{\text{free}}$ values were stalled at 40 and 46%,
respectively, at 2.3 Å resolution. Calculations revealed that
the r.m.s. distances among C$^\alpha$ 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 $P3_12$ 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
$R_{\text{merge}}$ values in space group $P3_2$ (7.6%) with $P3_12$ (9.1%)
(Table 1), expansion of the model to space group $P3_2$ brought
a significant drop in $R$ and $R_{\text{free}}$ 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 mis-
oriented 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.

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 $R_{\text{merge}}$ values in space groups $P3_2$ and
$P3_12$. 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 $P3_12$. 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
$\alpha$ as a function of resolution shells using the Twinning Server
and compared the estimated twinning fractions with the $R_{\text{merge}}$
values in the corresponding resolution shells (Fig. 3a). If two
reflections are related by a twin law, the extent of disagree-
ment 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 $R_{\text{merge}}$ 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
$R_{\text{merge}}$ values between twin-related reflections for both $I_{\text{calc}}$
and $I_{\text{obs}}$ (Fig. 3b). As we have seen, $R_{\text{merge}}$ on $I_{\text{calc}}$ is approxi-
ately 10–25% across the resolution range. However, $R_{\text{merge}}$
on $I_{\text{calc}}$ prior to implementation of the twinning operator is
approximately 65–85%. Clearly, NCS is insufficient to explain
the low $R_{\text{merge}}$ observed in the twin-related $I_{\text{obs}}$. Implementa-
tion of the twinning operator brings $R_{\text{merge}}$ on $I_{\text{calc}}$ 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 $R_{\text{free}}$ after
introducing partial twinning into the refinement processes.

![Figure 3](image-url)

Data suggesting partial twinning. (a) Estimated twinning fraction $\alpha$ versus
$R_{\text{merge}}$ 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 $P3_2$. $R_{\text{merge}}$ values (open circles) in the same resolution
shells as those used to calculate the twinning fractions were calculated
using SCALEPACK (Otwinowski & Minor, 1997). (b) $R_{\text{merge}}$ values versus
resolution shells for twinning-related reflections in space group
$P3_2$, shown with $R_{\text{merge}}$ values on $I_{\text{calc}}$ without implementing twinning
(filled circles), on $I_{\text{calc}}$ with implementing twinning (open circles) and $I_{\text{obs}}$
(filled reverse triangles).
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 $R_{\text{free}}$ by 6 and 5%, respectively. The drastic decrease in the $R$ and $R_{\text{free}}$ values merely from introducing twinning without any model rebuilding strongly supported the presence of partial twinning.

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 $R_{\text{free}}$, respectively, after releasing all NCS restraints. We believe that the relatively large difference between $R$ and $R_{\text{free}}$ 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 axis 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 $I_{\text{calc}}$ for models oriented with different degrees of alignment of the NCS twofold axis with the twinning operator (Fig. 4a). 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 $I_{\text{calc}}$ shows a strong deviation towards the appearance of being ‘untwinned’ (Fig. 4b). 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 PaSOD, 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 $(F^2)/(I)^2$.
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.

Table 3
Statistics for atomic refinement of metal-free PaSOD.

<table>
<thead>
<tr>
<th></th>
<th>Values in parentheses are for the highest resolution shell.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>36.45–1.8</td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>41127</td>
</tr>
<tr>
<td>Average B factor, protein atoms (Å²)</td>
<td>21.0</td>
</tr>
<tr>
<td>No. of other molecules</td>
<td>Acetate: 19, β-Mercaptoethanol: 12, Water: 2536</td>
</tr>
<tr>
<td>R.m.s.d. Bond length (Å)</td>
<td>0.007</td>
</tr>
<tr>
<td>R.m.s.d. Bond angle (°)</td>
<td>1.4</td>
</tr>
<tr>
<td>RFree</td>
<td>0.217 (0.263)</td>
</tr>
<tr>
<td>R</td>
<td>0.162 (0.241)</td>
</tr>
</tbody>
</table>

Figure 5
The structure of PaSOD. (a) Ribbon representation of the monomeric structure of PaSOD. Helices are colored red and strands cyan. The monomeric structure of PaSOD shows an α/β fold, typically found in many other Mn or FeSOD structures. (b) Ribbon representation of the tetrameric structure of PaSOD. Each chain is colored differently. Two different views are shown. 222 point symmetry is clearly observed for the tetrameric structure of PaSOD. This figure was prepared using PyMOL (DeLano Scientific). (c) Arrangement of PaSOD 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 PaSOD. The 2F_o – F_c map (green) around the active site of PaSOD 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 PaSOD monomer structure reveals an αβ fold (Fig. 5a). 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 PaSOD monomer structure consists of short helices and strands. The PaSOD monomer structure is similar to known Fe or MnSOD structures.

The structure of PaSOD tetramer, which is a biologically active unit (Whittaker & Whittaker, 2000), shows 222 point symmetry (Fig. 5b). Like S. sulfataricus SOD, PaSOD forms a compact tetramer (Fig. 5b). PaSOD 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 Cα 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. 5c). 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. 5d). A water molecule was found in the active site. Inclusion of MnCl2 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 PaSOD was solved by molecular replacement (manuscript in preparation). The current structural model represents the ‘metal-free’ state of PaSOD.

4. Conclusions

The structure determination of PaSOD 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 Rfree 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 PaSOD. 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


Lee et al. • Superoxide dismutase 2199