Putative thioredoxin Trx1 from *Thermosipho africanus* strain TCF52B: expression, purification and structural determination using S-SAD

Naheda Sahtout,^{a,‡} Jijin R. A. Kuttiyatveetil,^{a,‡} Michel Fodje^b and David A. R. Sanders^{a,*}

Received 24 March 2016

Accepted 28 April 2016

Edited by J. Newman, Bio21 Collaborative Crystallisation Centre, Australia

‡ These authors contributed equally to the work presented in this publication.

Keywords: thioredoxin; *Thermosipho africanus*; thermophile; sulfur; SAD.

^aDepartment of Chemistry, University of Saskatchewan, 110 Science Place, Saskatoon, SK S7N 5C9, Canada, and

^bCanadian Macromolecular Crystallography Facility, Canadian Light Source Inc., University of Saskatchewan, 44 Innovation Boulevard, Saskatoon, SK S7N 2V3, Canada. *Correspondence e-mail: david.sanders@usask.ca

Thioredoxin is a small ubiquitous protein that plays a role in many biological processes. A putative thioredoxin, Trx1, from *Thermosipho africanus* strain TCF52B, which has low sequence identity to its closest homologues, was successfully cloned, overexpressed and purified. The protein was crystallized using the microbatch-under-oil technique at 289 K in a variety of conditions; crystals grown in 0.2 M MgCl₂, 0.1 M bis-tris pH 6.5, 25% (w/v) PEG 3350, which grew as irregular trapezoids to maximum dimensions of 1.2 × 1.5 × 0.80 mm, were used for sulfur single-wavelength anomalous dispersion analysis. The anomalous sulfur signal could be detected to 2.83 Å resolution using synchrotron radiation on the 08B1-1 beamline at the Canadian Light Source. The crystals belonged to space group *P*2₁2₁2₁, with unit-cell parameters *a* = 40.6, *b* = 41.5, *c* = 56.4 Å, $\alpha = \beta = \gamma = 90.0^\circ$.

1. Introduction

The thioredoxin system is a ubiquitous oxidoreductase system that consists of the enzyme thioredoxin reductase (TrxR), the corresponding thioredoxin (Trx) substrate and the cofactor nicotinamide adenine dinucleotide phosphate (NADPH). Trx is a small ubiquitous protein which plays a role in protein disulfide reduction (Holmgren *et al.*, 1975; Jeng *et al.*, 1994). Trx belongs to the thioredoxin-fold superfamily and contains the common thioredoxin fold: five β -sheets surrounded by four flanking α -helices (Martin, 1995). Trx contains a highly conserved redox-active site consisting of a CXXC motif; in *Escherichia coli* Trx1 (EcTrx1) this active site involves the residues Cys32–Cys35 (Chivers & Raines, 1997). Although extensive knowledge is available about EcTrx1, little is known about the differences that arise between the thioredoxin systems of thermophilic bacteria and that of *E. coli*.

Thermosipho africanus (*T. africanus*) is a thermophilic eubacterium, belonging to the order Thermotogales, that was first isolated from a marine hydrothermal area in Djibouti, Africa (Huber *et al.*, 1989). Strain TCF52B was isolated from a high-temperature oil reservoir in the North Sea and was identified as being a strain of *T. africanus* (Nesbø *et al.*, 2009). Analysis of the complete genome sequence of *T. africanus* strain TCF52B, NCBI reference sequence NC_011653.1, suggested the presence of two putative Trxs, NCBI reference sequences WP_004103357.1 (Trx1) and WP_012580195.1 (Trx2), and a TrxR, NCBI reference sequence WP_004101768.1 (TrxR), as components of its thioredoxin system.

© 2016 International Union of Crystallography

The putative *T. africanus* Trx1 (TaTrx1) is 19 amino acids shorter than EcTrx1 and comparison of the protein sequences indicates only 17% sequence identity. Furthermore, a search of current crystal structures indicated that TaTrx1 shares only 26 and 24% sequence identity with the thioredoxin-like protein BCE_0499 from *Bacillus cereus* (PDB entry 4euy; Midwest Center for Structural Genomics, unpublished work) and *Archaeoglobus fulgidus* Trx3 (PDB entry 4xhm; Bewley *et al.*, 2015), respectively. In spite of the low sequence identity to its closest homologues, TaTrx1 still contains the active-site cysteines typical of Trxs (Fig. 1).

In this paper, we present X-ray crystallography and diffraction data of TaTrx1 obtained using sulfur single-wavelength anomalous diffraction (S-SAD). Determining the crystal structure is the first step in providing an understanding of the distinguishing features of the thioredoxin system from *T. africanus*.

2. Materials and methods

2.1. Cloning

The gene encoding TaTrx1 was amplified using the polymerase chain reaction (PCR) from the *T. africanus* strain TCF52B genomic DNA using 5'-CATGCCATGGCTATGAAGATAGAATACTTTAAG-3' and 5'-CGGGATCCTTATTCTTTATTACTTATTTCTAGG-3' as the forward and reverse primers, respectively (Table 1). The PCR product was digested with the restriction enzymes NcoI and BamHI prior to insertion into the pEHISTEV vector (Liu & Naismith, 2009). The clone was transformed into *E. coli* DH5 α competent cells and the constructs were verified by sequencing at the Applied Genomics Centre, National Research Council, Saskatoon, Saskatchewan, Canada.

2.2. Overexpression and purification

To overexpress the N-terminally hexahistidine-tagged TaTrx1 protein, the construct was transformed into *E. coli*

Table 1

Summary of cloning and the construct produced.

The NcoI and BamHI restriction sites are underlined in the forward and reverse primers, respectively. The additional N-terminal residues of the tag, encoding a hexahistidine tag and a *Tobacco etch virus* (TEV) protease cleavage site, are underlined in the amino-acid sequence of the construct produced.

Source organism	<i>T. africanus</i> strain TCF52B
DNA source	Genomic
Forward primer	5'-CATGCCATGGCTATGAAGATAGAATACTTTAAG-3'
Reverse primer	5'-CGGGATCCTTATTCTTTATTACTTATTCTAAG-3'
Cloning vector	pEHISTEV
Expression vector	pEHISTEV
Expression host	<i>E. coli</i> BL21-CodonPlus-RIL
Complete amino-acid sequence of the construct produced	MSYYHHHHHHYDIPTTENLYFQGAMAMKIEYFK-NDKCSVCKAMPLPKIQTIAKNFDIDIEVIDVIE-NPSYPAQKLVTFTPTVITLDKEFEIKRFARNF-SISEVINTIERYLEISNKE

BL21-CodonPlus-RIL competent cells (Table 1). A single colony was used to culture Luria-Bertani (LB) medium supplemented with 50 $\mu\text{g ml}^{-1}$ kanamycin. The culture was grown at 310 K and 250 rev min $^{-1}$ to an OD₆₀₀ of 0.6. Expression of the protein was induced with the addition of isopropyl β -D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 M, and the induced cells were cultured for an additional 24 h at 303 K and 250 rev min $^{-1}$. The cells were harvested by centrifugation at 3500 rev min $^{-1}$ for 30 min at 277 K.

The cell pellets were resuspended in 50 mM potassium phosphate pH 8, 150 mM NaCl, 1 $\mu\text{g ml}^{-1}$ 4-(2-aminoethyl)-benzenesulfonyl fluoride hydrochloride (AEBSF), 0.4 $\mu\text{g ml}^{-1}$ lysozyme, 0.2 $\mu\text{g ml}^{-1}$ DNase and left to stir for 30 min at 277 K. The lysate was sonicated and then centrifuged at 15 000 rev min $^{-1}$ for 30 min at 277 K. The supernatant was passed through a 0.45 μm filter (Sarstedt) and loaded onto a Ni-Sepharose column (GE Healthcare) pre-equilibrated with 50 mM potassium phosphate pH 8, 150 mM NaCl at room temperature. The column was washed with four column

Table 2

Summary of the optimum crystallization conditions.

Method	Microbatch under oil
Temperature (K)	289
Protein concentration (mg ml ⁻¹)	10
Buffer composition of protein solution	50 mM Tris pH 8, 150 mM NaCl, 1 mM DTT
Composition of reservoir solution	0.1 M bis-tris pH 6.5, 0.2 M MgCl ₂ , 25%(w/v) PEG 3350
Volume and ratio of drop	2.4 µl (1:1)

volumes of 50 mM potassium phosphate pH 8, 150 mM NaCl. The N-terminally hexahistidine-tagged TaTrx1 was eluted with 50 mM potassium phosphate pH 8, 150 mM NaCl, 150 mM imidazole. The purity of the fractions was assessed by SDS-PAGE. Fractions containing the N-terminally hexahistidine-tagged TaTrx1 were pooled together and dialyzed against 50 mM Tris pH 8, 150 mM NaCl, 1 mM DTT. The dialyzed N-terminally hexahistidine-tagged TaTrx1 was concentrated to approximately 10 mg ml⁻¹ using a 5 kDa cutoff concentrator (Sartorius), flash-cooled in liquid nitrogen and stored at 193 K.

2.3. Crystallization

Initial broad screening of the purified N-terminally hexahistidine-tagged TaTrx1 using commercial crystallization kits (Qiagen and Microlytic) was completed using the microbatch-under-oil technique at 289 K. Crystallization drops consisted of equal volumes (1.2 µl) of protein solution and precipitant solution topped with a layer of paraffin oil (Hampton Research). Positive hits from the initial broad screening were identified in a number of conditions from different commercial screens. The conditions that produced crystals included The Classics Suite condition No. 71 [0.2 M sodium acetate, 0.1 M sodium cacodylate pH 6.5, 30%(w/v) PEG 8000], The Classics

Figure 2

Irregular trapezoid crystals of the N-terminally hexahistidine-tagged *T. africanus* Trx1 protein grown using the microbatch method under paraffin oil at 289 K in 0.2 M MgCl₂, 0.1 M bis-tris pH 6.5, 25%(w/v) PEG 3350 with maximum dimensions of 1.2 × 1.5 × 0.8 mm.

Table 3

Data collection and processing from S-SAD.

Crystal-to-detector distances are given in the order of collection of the data sets. Values in parentheses are for the highest resolution shell.

	Native data set	S-SAD data set
Diffraction source	08B1-1, Canadian Light Source	
Detector	Rayonix MX300HE CCD X-ray detector	
Wavelength (Å)	1.7712	
Temperature (K)	100	
Energy (keV)	7.0	
Beam size (µm)	150	
Exposure time per image (s)	5	
Crystal-to-detector distance (mm)	120, 121, 122, 123, 124, 119, 118, 117, 120, 120	
No. of frames	360 [first 9 data sets], 720 [final data set]	
δ (°)	1.0 [first 9 data sets], 0.5 [final data set]	
κ (°)	0 [first 8 data sets], 30 [final 2 data sets]	
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	40.6, 41.5, 56.4	
α, β, γ (°)	90.0, 90.0, 90.0	
Mosaicity (°)	0.22	
Resolution range (Å)	33.03–1.90 (1.95–1.90)	
Total No. of reflections	949486 (25960)	949447 (25960)
No. of unique reflections	7586 (368)	13894 (702)
Completeness (%)	96.0 (64.6)	95.9 (65.4)
<i>I</i> /σ(<i>I</i>)	72.51 (2.77)	55.28 (2.02)
Multiplicity	125.2 (70.5)	68.3 (37.0)
<i>R</i> _{meas} (%)	11.2 (216.4)	11.1 (216.5)
<i>CC</i> _{1/2} (%)	100.0 (67.6)	100.0 (66.5)

II Suite condition No. 83 [0.2 M MgCl₂, 0.1 M bis-tris pH 6.5, 25%(w/v) PEG 3350], The PACT Suite condition No. 7 [0.2 M NaCl, 0.1 M sodium acetate pH 5, 20%(w/v) PEG 6000], The PACT Suite condition No. 8 [0.2 M ammonium chloride, 0.1 M sodium acetate pH 5, 20%(w/v) PEG 6000], The PACT Suite condition No. 28 [0.1 M PCB buffer pH 7, 25%(w/v) PEG 1500], The PEGs Suite condition No. 5 [0.1 M sodium acetate pH 4.6, 25%(w/v) PEG 1000], The PEGs Suite condition No. 6 [0.1 M sodium acetate pH 4.6, 25%(w/v) PEG 2000 MME], The PEGs Suite condition No. 64 [0.2 M magnesium nitrate, 20%(w/v) PEG 3350], The PEGs II Suite condition No. 5 [0.1 M MgCl₂, 0.1 M sodium acetate pH 4.6, 25%(w/v) PEG 400], The PEGs II Suite condition No. 6 [0.2 M lithium sulfate, 0.1 M Tris pH 8.5, 25%(w/v) PEG 400] and MCSG-1 condition No. 85 [0.2 M ammonium fluoride, 20%(w/v) PEG 3350]. Ultimately, large irregular-shaped trapezoid crystals from The Classics II Suite condition No. 83, which grew to maximum dimensions of 1.2 × 1.5 × 0.80 mm in 3–4 weeks, were utilized for S-SAD X-ray diffraction analysis (Table 2, Fig. 2). Crystals were washed in a solution consisting of the mother liquor supplemented with 25% PEG 400 as a cryoprotectant and then mounted onto a CryoLoop (Hampton Research) and flash-cooled in liquid nitrogen.

2.4. Data collection and processing

We initially tried to solve the phase problem using molecular replacement, with EcTrx1 (PDB entry 2h6x; R. Godoy-Ruiz, J. A. Gavira, B. Ibarra-Molero & J. M. Sanchez-Ruiz, unpublished work) as the search model. This was not successful, likely owing to the low sequence identity between the two proteins (Fig. 1). A BLAST search for other potential

candidates for molecular replacement was also unsuccessful. We therefore attempted to determine the phase using S-SAD since there was a total of four S atoms in the protein sequence of TaTrx1 and two in the N-terminal hexahistidine tag.

X-ray diffraction data were collected on the 08B1-1 beamline at the Canadian Light Source using a Rayonix MX300HE CCD X-ray detector. Diffraction data were collected at an energy of 7 keV (wavelength of 1.7712 Å). A total of ten data sets were collected with crystal-to-detector distances varying from 117 to 124 mm (Table 3). A total of 360 images were collected for the first nine data sets and 720 images for the final data set, with an exposure of 5 s per image and a crystal oscillation of 1° per frame. The first nine data sets were obtained at a δ angle of 1.0°, while the final data set was obtained at a δ angle of 0.5°. The κ angle was kept at 0° for the first eight data sets and changed to 30° for the final two data sets (Table 3). The ten data sets were processed and merged with *XDS* (Kabsch, 2010) through the *AutoProcess* pipeline (Fodje *et al.*, 2014). Data quality was assessed with the use of *phenix.xtriage* (Terwilliger *et al.*, 2008). Heavy-atom positions were detected, experimental phases were determined and part of the protein model was built using *phenix.autosol* (Terwilliger *et al.*, 2009; Adams *et al.*, 2010). The model was then completed using *phenix.autobuild* (Terwilliger *et al.*, 2008; Adams *et al.*, 2010).

3. Results and conclusions

N-terminally His-tagged TaTrx1 was successfully cloned, overexpressed and purified. The purity of the protein was determined to be greater than 95% as assessed by SDS-PAGE (Fig. 3). Approximately 25 mg of His-tagged protein was obtained from 1 l LB culture after purification. The molecular

Figure 3
SDS-PAGE analysis of the N-terminally hexahistidine-tagged *T. africanus* Trx1 protein purified using Ni-Sepharose affinity chromatography. Lane *M* contains low-range molecular-weight marker (Sigma-Aldrich) and lane 1 contains approximately 15 µg of protein loaded prior to flash-cooling in liquid nitrogen and storage at 193 K.

weight of the His-tagged protein was determined to be around 14 kDa, consistent with the calculated value of 13.8 kDa. Crystals of the His-tagged protein were obtained at 289 K under a variety of conditions using the microbatch-under-oil technique. Crystals that grew in 0.2 M MgCl₂, 0.1 M bis-tris pH 6.5, 25% (w/v) PEG 3350 were ultimately used for diffraction.

Ten data sets were collected from diffraction from a single crystal. When merged, the diffraction data for S-SAD phasing were effective down to 2.83 Å resolution. The collection of further data sets failed to improve the anomalous sulfur signal or the quality of the analysis of the diffraction data. From the data, the positions of all six S atoms could be detected and three additional sites were found corresponding to other nonsulfur heavy elements, such as magnesium from the crystallization solution. The statistics of data collection as well as the parameters obtained are summarized in Table 3. The protein was found to correspond to a Matthews coefficient of 1.8 Å³ Da⁻¹ and a solvent content of 31.8%.

The initial phase of the protein was determined using *phenix.autosol*, which was able to produce an initial model based on 77 residues with a figure of merit (FOM) of 0.43, a score of 43.5 ± 11.2, an *R*_{work} of 0.37 and an *R*_{free} of 0.47. The experimental electron-density map obtained suggested that the data were satisfactory (Fig. 4). From the initial model, the protein was determined to be a monomer with one monomer per asymmetric unit. The initial model could be improved with *phenix.autobuild*. A total of 88 residues were built as one chain, including 80 water molecules, and the *R*_{work} and *R*_{free} values were reduced to 0.21 and 0.24, respectively. The N-terminal His tag was not observed in the structure. A distance alignment matrix search, using the *DALI* server, on the TaTrx1 model from *phenix.autobuild* identified that the three-dimensional structure most closely resembled those of the thioredoxins from *Thermus thermophilus*, *Sulfolobus tokodaii* and *Plasmodium falciparum*, despite their sequence identities being lower than 20% (Holm & Rosenström, 2010). Further refinement of the structure is currently in progress.

Figure 4
Electron-density map around two S atoms which form the disulfide bridge found at the active site of oxidized thioredoxin Trx1 from *T. africanus*. The purple map is a density-modified experimentally phased map from S-SAD. The blue map is a 2*mF*_o - *DF*_c map calculated using phases from the final model built with *phenix.autobuild*. Both maps are contoured at 1.5σ. S atoms are indicated by yellow spheres. This figure was prepared in *PyMOL*.

and details pertaining to the structure and function of the protein will be described in a separate paper.

Acknowledgements

The genomic DNA of *T. africanus* strain TCF52B was provided by Dr Camilla L. Nesbø from the University of Alberta. We thank the staff of the macromolecular beamline 08B1-1 at the Canadian Light Source for their technical assistance in data collection and processing. This work was supported by an NSERC Discovery Grant to DARS. The research described in this paper was performed at the Canadian Light Source, which is supported by the Canada Foundation for Innovation, the Natural Sciences and Engineering Research Council of Canada, the University of Saskatchewan, the Government of Saskatchewan, Western Economic Diversification Canada, the National Research Council Canada and the Canadian Institutes of Health Research.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Bewley, K. D., Dey, M., Bjork, R. E., Mitra, S., Chobot, S. E., Drennan, C. L. & Elliott, S. J. (2015). *PLoS One*, **10**, e0122466.
- Chivers, P. T. & Raines, R. T. (1997). *Biochemistry*, **36**, 15810–15816.
- Fodje, M., Grochulski, P., Janzen, K., Labiuk, S., Gorin, J. & Berg, R. (2014). *J. Synchrotron Rad.* **21**, 633–637.
- Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
- Holmgren, A., Söderberg, B. O., Eklund, H. & Brändén, C.-I. (1975). *Proc. Natl Acad. Sci. USA*, **72**, 2305–2309.
- Huber, R., Woese, C. R., Langworthy, T. A., Fricke, H. & Stetter, K. O. (1989). *Syst. Appl. Microbiol.* **12**, 32–37.
- Jeng, M.-F., Campbell, A. P., Begley, T., Holmgren, A., Case, D. A., Wright, P. E. & Dyson, H. J. (1994). *Structure*, **2**, 853–868.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Liu, H. & Naismith, J. H. (2009). *Protein Expr. Purif.* **63**, 102–111.
- Martin, J. L. (1995). *Structure*, **3**, 245–250.
- Nesbø, C. L., Baptiste, E., Curtis, B., Dahle, H., Lopez, P., Macleod, D., Dlutek, M., Bowman, S., Zhaxybayeva, O., Birkeland, N.-K. & Doolittle, W. F. (2009). *J. Bacteriol.* **191**, 1974–1978.
- Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. (2011). *Mol. Syst. Biol.* **7**, 539.
- Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H. & Hung, L.-W. (2009). *Acta Cryst.* **D65**, 582–601.
- Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst.* **D64**, 61–69.