Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Purification, crystallization and preliminary X-ray analysis of an unusual thioredoxin from the gastric pathogen *Helicobacter pylori*

Heather Filson,[a]† Áine Fox,[b]† Dermot Kelleher,[b] Henry J. Windle[b] and David A. R. Sanders[a]*

[a]Department of Chemistry, University of Saskatchewan, Saskatoon, Saskatchewan S7N 5C9, Canada, and [b]Department of Clinical Medicine, Trinity College, Dublin and Dublin Molecular Medicine Centre, Dublin 8, Ireland

† These authors made an equal contribution to this work.

Correspondence e-mail: david.sanders@usask.ca

Received 3 April 2003
Accepted 1 May 2003

Thioredoxin-2 (HP1458) from *Helicobacter pylori* is a member of the thioredoxin family, but possesses the unusual active-site motif CPDC (compared with CGPC in other thioredoxins). *H. pylori* is deficient in the glutaredoxin system, making the thioredoxin system the sole reduction system in the bacterium and critical for its ability to survive oxidative stress. The recombinant protein has been overexpressed, purified and crystallized. This is the first reported crystallization of a thioredoxin possessing this unusual active site. Single crystals have been obtained using the sitting-drop technique. Crystals diffract to 2.4 Å resolution and belong to space group $P4_1$, with unit-cell parameters $a = b = 40.21$, $c = 64.65$ Å. Molecular replacement using *AMoRe* proved unsuccessful; however, implementation of the program *BEAST* gave successful molecular-replacement solutions.

1. Introduction

Helicobacter pylori is a bacterium that colonizes the human gastric epithelium, is a major cause of chronic gastritis and is associated with peptic ulcer disease in addition to being an independent risk factor for gastric cancer (Marshall & Warren, 1984; Blaser, 1987; Forman *et al.*, 1991). As an inhabitant of the gastric tissue, *H. pylori* is exposed to an environment rich in inflammatory infiltrates and reactive oxygen species (ROS) as a consequence of the host immune response to infection (*e.g.* Ernst, 1999). These host responses and pathogen-induced alterations to the gastric environment encourage the persistence of damaging oxidative and carcinogenic agents (Drake *et al.*, 1996; Obst *et al.*, 2000; Shirin *et al.*, 2001; Bhattacharjee *et al.*, 2002). In addition, *H. pylori* secretes O_2^-, thus providing a direct role for the bacterium in modulating the gastric redox environment (Nagata *et al.*, 1998).

H. pylori is significantly different from other prokaryotes in terms of the deficiency in its complement of thiol-dependent redox-active proteins, being equipped with only one of the two major, almost ubiquitous, intracellular disulfide-reducing systems, the thioredoxin (Trx) system, and lacking the glutaredoxin/glutathione (Grx) system. Generally, these systems have a major role in protecting cells from oxidative damage. As few cell types are deficient in the glutathione/glutaredoxin system (Kanzok *et al.*, 2001), the regulation of thiol-mediated redox homeostasis is of interest in cells/organisms where significant deficiencies in redox biochemistry are noted. As a consequence of this deficiency, it is likely that the redox metabolism of *H. pylori* is significantly different from that of other prokaryotes. In the context of pathogens and parasites, this field of research is now receiving attention as novel mechanisms of maintaining redox balance have implications for the development of effective therapeutics (Gommel *et al.*, 1997; Kanzok *et al.*, 2000; Alger & Williams, 2002; Lopez *et al.*, 2002).

H. pylori contains two genes encoding thioredoxins [(Trx1 (HP824) and Trx2 (HP1458), The Institute for Genomic Research annotation; http://www.tigr.org]. The *H. pylori* thioredoxins show relatively low primary sequence homology to one another (35% identity). Trx1 has the highly conserved motif CGPC found in all thioredoxins studied to date with few exceptions. The amino-acid residues flanked by the vicinal cysteine residues are known to influence the redox potential of the molecule (Chivers *et al.*, 1997). Trx2 has the unusual motif CPDC not found in other Trxs. Further, Trx2 has an unusually high pI (8.8) compared with the acidic Trx1 (pI 5.3). The sequence differences within the redox active-site motif (Trx1, CGPC; Trx2, CPDC) suggest that the Trxs are functionally dissimilar; this is perhaps an essential divergence in roles given the absence of the Grx system. Generally, Trxs are retained within the cytoplasm or periplasm of bacteria; however, recent proteomic analyses of the *H. pylori* secretome have demonstrated constitutive secretion of Trx (Bumann *et al.*, 2002; Kim *et al.*, 2002). The only other published report of a secreted thioredoxin occurs in humans, where it has been linked to an increase in breast cancer (Powis & Montfort, 2001). This raises some intriguing possibilities in relation to the effects

they may have on host gastric cells. For example, *H. pylori* Trx is an efficient reductant of human immunoglobulin and mucin *in vitro* (Windle *et al.*, 2000) and thus it is conceivable that these molecules may have a role in the colonization process (Windle *et al.*, 2000; Bumann *et al.*, 2002).

We report here the overexpression, purification, crystallization and initial data collection for Trx2 from *H. pylori*. This paper is the first reported crystallization of a thioredoxin with the unusual active-site motif CPDC. The structural studies will provide insight into the redox role that Trx2 performs in *H. pylori*.

2. Protein overexpression and purification

2.1. Cloning and overexpression

The pPROK1 vector (Clontech) containing the *H. pylori trx2* gene was provided by Dr Leslie Poole (Wake Forest University). The *trx2* gene encodes a protein with 104 amino acids and a predicted molecular weight of 11.7 kDa (Baker *et al.*, 2001). The plasmid was used to transform *Escherichia coli* BL21(DE3)*pLysS* using the heat-shock method of transformation and growth took place on LB plates with ampicillin (50 µg ml^{-1}) and chloramphenicol (50 µg ml^{-1}).

Single colonies were selected and grown overnight in broth for a starter culture. This was used to inoculate 800 ml of LB medium containing ampicillin (50 µg ml^{-1}) and chloramphenicol (50 µg ml^{-1}). The cells were grown at 310 K until the OD_{600} reached 0.6. Overexpression was induced by the addition of 0.4 m*M* isopropyl-β-D-thiogalactopyranoside (IPTG) and the culture was incubated at 310 K for a further 3 h. The cells were harvested by centrifugation (5 min, 5000*g*, 277 K) and the pellets were stored at 253 K until use.

2.2. Purification

Trx2 was purified essentially as described by Baker *et al.* (2001), with minor modifications. The thawed and lysed pellets of cells were resuspended in buffer (5 m*M* potassium phosphate pH 7.0) and sonicated (5 × 30 s pulses with cooling on ice in a Branson Sonifier Model 450). Sonicated material was centrifuged (13 000*g*, 277 K) for 25 min and the supernatant was filtered (0.45 µm, Pall Acrodisc low-protein-binding filter) and dialysed overnight against 5 m*M* potassium phosphate buffer pH 7.0 at 277 K. The dialysate was applied to an anion-exchange column (24 × 2.5 cm; Whatman DE52) pre-equilibrated with 5 m*M* potassium phosphate pH 7.0 buffer (flow rate 0.6 ml min^{-1}) and the non-binding fraction was collected and subsequently applied to a carboxymethylcellulose column (24 × 2.5 cm; Whatman CM52) pre-equilibrated with 5 m*M* potassium phosphate buffer pH 7.0 and eluted with a linear gradient of potassium phosphate (5–40 m*M*). Fractions containing pure Trx2 (identified by SDS–PAGE with silver staining) were pooled and dialyzed against 30 m*M* potassium phosphate pH 7.0. The protein was then concentrated to 12 mg ml^{-1} and stored at 253 K with 2 m*M* DTT. Typical yields were 13–15 mg of protein from 1 l of culture.

3. Crystallization, data collection and structure solution

3.1. Crystallization

Crystal screening was carried out using the sitting-drop method of vapor diffusion. Initial trials were carried out using Crystal Screens 1 and 2 (Hampton Research) using 4 µl drops consisting of 2 µl protein (12 mg ml^{-1} in 30 m*M* potassium phosphate pH 7.0) and 2 µl well buffer in Crystal Clear strips (Douglas Instruments) at 293 K. Addition of 10 m*M* DTT was shown to give a significant improvement in crystal size and morphology. The final crystal conditions comprised sitting drops consisting of 5 µl of 3 mg ml^{-1} protein (30 m*M* potassium phosphate pH 7.0, 10 m*M* DTT) plus 5 µl of well solution [30% PEG 6000, 0.1 *M* $(NH_4)_2SO_4$, 10 m*M* DTT], yielding rod-shaped crystals of dimensions 0.1 × 0.1 × 1–2 mm (Fig. 1). Crystals took 5–7 d to grow.

3.2. X-ray data collection

Crystals of Trx2 were frozen by moving the crystal to a solution containing 35% PEG 6000, 0.1 *M* $(NH_4)_2SO_4$, 10 m*M* DTT and 5% glycerol. The crystal was then moved into solutions with increasing amounts of glycerol (10 and 15%) and finally frozen by submerging in liquid nitrogen.

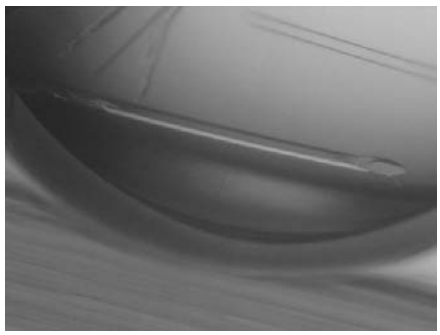

Figure 1
Typical crystals of Trx2 from *H. pylori*.

Table 1
Data-collection statistics.

Values in parentheses refer to the highest resolution shell (2.53–2.40 Å).

Wavelength (Å)	1.00
Resolution (Å)	40.16–2.40
Space group	$P4_1$
Unit-cell parameters (Å, °)	$a = b = 40.21$, $c = 64.65$, $\alpha = \beta = \gamma = 90$
Total measurements	55405
Unique reflections	4079
V_M† (Å^3 Da^{-1})	2.1
Solvent (%)	43
$I/\sigma(I)$	6.6 (2.3)
Average redundancy	13.6 (13.8)
Data completeness (%)	100 (100)
R_{merge}‡ (%)	7.3 (47.1)

† Assuming a monomer in the asymmetric unit. ‡ $R_{merge} = \sum I(h)_i - I(h)/\sum\sum I(h)_i$, where $I(h)$ is the measured diffraction intensity and the summation includes all observations.

Initial data were collected using the Mail-in data service of the NSLS. Nine crystals were sent and two data sets were finally collected, the best to 2.4 Å, with 360° of data collected. The data were collected on beamline X12C. The data were initially processed on-site and were later reprocessed using *MOSFLM* (Leslie, 1992).

The space group was initially indexed in a tetragonal space group (unit-cell parameters a = 39.74, b = 39.74, c = 64.95 Å, $\alpha = \beta = \gamma = 90°$) and the data were integrated and merged in Laue group $P4$ using *MOSFLM* (Leslie, 1992) and *SCALA* (Collaborative Computational Project, Number 4, 1994); however, further examination of the systematic absences showed that the space group was either $P4_1$ or $P4_3$. Calculation of the Matthews volume (Matthews, 1968) gave a V_M of 2.1 Å^3 Da^{-1} and a solvent content of 43%, assuming one molecule per asymmetric unit and a molecular weight of 11.7 kDa. Table 1 summarizes the data set.

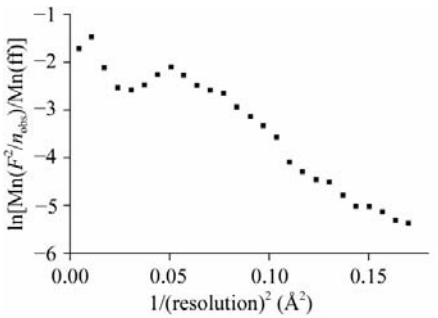

Figure 2
Wilson plot of the data for Trx2 to 2.4 Å. The *B* factor is calculated as 60 Å^2 and the scale factor is 1.68. Data taken from *TRUNCATE* (Collaborative Computational Project, Number 4, 1994).

At the highest resolution, the $R_{\rm merge}$ becomes quite high (47%) owing to a slight splitting of the spots at high resolution. We believe this is because of the tendency of the long thin crystals to bend with the contour of the frozen drops. Fig. 2 shows the Wilson plot for this data. The B factor is calculated as 60 Å^2 and the scale factor is 1.68. These values are reasonable given the limited resolution of the data. The trace in Fig. 2 remains linear for the higher resolution data, indicating that the data remain consistent, even with the higher $R_{\rm merge}$.

3.3. Structure solution

All structure-solution trials were run using the *CCP*4 package (Collaborative Computational Project, Number 4, 1994). Molecular replacement was initially performed using *AMoRe* (Navaza, 2001). The sequence homology between Trx2 from *H. pylori* and known thioredoxin structures is quite low, the highest being 27% identity to the *E. coli* (Katti *et al.*, 1990), human (Weichsel *et al.*, 1996) and *Spinacea oleracea* (Trx-M, oxidized; Capitani *et al.*, 2000) structures. These structures all contain the conserved active-site CGPC, unlike Trx2. These structures were used for molecular-replacement searches. As it is not possible to distinguish between space groups $P4_1$ and $P4_3$ based on systematic absences, all of the molecular-replacement searches were performed in both $P4_1$ and $P4_3$ space groups.

Models for *AMoRe* were derived from the structure of the known thioredoxins, either using the entire known structure, a structure using a polyalanine backbone or a model created by predicting the Trx2 structure using *MODELLER* (Šali & Blundell, 1993) and the known structures. An entire rotation search was performed for each model and the top 20 solutions were used for translation and fitting. None of these searches resulted in identifiable solutions.

Further molecular-replacement trials were run using *BEAST* (Read, 2001). The entire known thioredoxin structures from *E. coli*, human and Trx-M were used. These searches were performed using the structures either singly or as a combined group. An entire rotation search was performed for each trial. The top 20 solutions were then used in a translation search of the entire Cheshire cell (x = 0–1.0, y = 0–0.5, z = 0) with the rotation fixed. The Trx-M model structure gave the best results, determined as the difference in the log likelihood gain (LLG) of the best solutions compared with the background LLG values (solution LLG = 42.7, background = 19.7; signal-to-noise = 8.5). The solution for $P4_1$ showed a significant improvement compared with that for $P4_3$ (LLGs of 15.0 and 12.4; signal-to-noise = 4.3). We expect that this molecular-replacement solution will be sufficient to allow us to determine the structure of Trx2. We will then progress to structures of Trx2 in complex with substrate proteins in order to thoroughly understand the role that Trx2 plays in both maintaining the redox environment of *H. pylori* and in the infective process.

We thank the NIH Institute, National Center for Research Resources for supporting the Brookhaven National Laboratory Biology/PX Mail-in program and the BNL/Biology beamline X12C at the National Synchrotron Light Source (NSLS). We thank Howard Robinson for his great assistance in this effort, for collecting the data and carrying out the initial processing. We thank Drs Leslie Poole and Laura Baker (Wake Forest University School of Medicine, Winston-Salem, USA) for the Trx2 construct (pPROK1). We thank Dr Randy Read for his invaluable assistance with running and interpreting his program *BEAST*. Funding for this project was provided in part by NSERC (RGPIN250238-02) and BioResearch Ireland.

References

Alger, H. M. & Williams, D. L. (2002). *Mol. Biochem. Parasitol.* **121**, 129–139.
Baker, L. M. S., Raudonikiene, A., Hoffman, P. S. & Poole, L. B. (2001). *J. Bacteriol.* **183**, 1961–1973.
Bhattacharjee, M., Bhattacharjee, S., Gupta, A. & Banerjee, R. K. (2002). *Free Radic. Biol. Med.* **32**, 731–743.
Blaser, M. J. (1987). *Gastroenterology*, **93**, 371–383.
Bumann, D., Aksu, S., Wendland, M., Janek, K., Zimny-Arndt, U., Sabarth, N., Meyer, T. & Jungblut, P. R. (2002). *Infect. Immun.* **70**, 3396–3403.
Capitani, G., Markovic Housley, Z., Delval, G., Morris, M., Jansonius, J. N. & Schurmann, P. (2000). *J. Mol. Biol.* **302**, 135–154.
Chivers, P. T., Prehoda, K. E. & Raines, R. T. (1997). *Biochemistry*, **36**, 4061–4066.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Drake, I. M., Davies, M. J., Mapstone, N. P., Dixon, M. F., Schorah, C. J., White, K. L., Chalmers, D. M. & Axon, A. T. (1996). *Carcinogenesis*, **17**, 559–562.
Ernst, P. (1999). *Aliment. Pharmacol. Ther.* **13**, 13–18.
Forman, D., Newell, D. G., Fullerton, F., Yarnell, J. W., Stacey, A. R., Wald, N. & Sitas, F. (1991). *Br. Med. J.* **302**, 1302–1305.
Gommel, D. U., Nogoceke, E., Morr, M., Kiess, M., Kalisz, H. M. & Flohé, L. (1997). *Eur. J. Biochem.* **248**, 913–918.
Kanzok, S. M., Fechner, A. Bauer, H., Ulschmid, J. K., Muller, H. M., Botella-Munoz, J., Schneuwlys, S., Schirmer, R. & Beckers, K. (2001). *Science*, **291**, 643–646.
Kanzok, S. M., Schirmer, R. H., Türbachova, I., Iozef, R. & Beckers, K. (2000). *J. Biol. Chem.* **275**, 40180–40186.
Katti, S. K., LeMaster, D. M. & Eklund, H. (1990). *J. Mol. Biol.* **167**, 167–184.
Kim, N., Weeks, D. L., Shin, J. M., Scott, D. R., Young, M. K. & Sachs, G. (2002). *J. Bacteriol.* **184**, 6155–6162.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EAMCB Newsl. Protein Crystallogr.* **26**.
Lopez, J. A., Carvalho, T. U., De Souza, W., Flohé, L., Guerrero, S. A., Montemartin, M., Kalisz, H. M., Nogoceke, E., Singh, M., Alves, M. J. M. & Colli, W. (2002). *Free Radic. Biol. Med.* **28**, 767–772.
Marshall, B. J. & Warren, J. R. (1984). *Lancet*, 1(8390), 1311–1315.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Nagata, K., Yu, H., Nishikawa, M., Kashiba, M., Nakamura, A., Sato, E. F., Tamura, T. & Inoue, M. (1998). *J. Biol. Chem.* **273**, 14071–14073.
Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.
Obst, B., Wagner, S., Sewing, K. F. & Beil, W. (2000). *Carcinogenesis*, **21**, 1111–1115.
Powis, G. & Montfort, W. R. (2001). *Annu. Rev. Biophys. Biomol. Struct.* **30**, 421–455.
Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.
Šali, A. & Blundell, T. (1993). *J. Mol. Biol.* **234**, 779–815.
Shirin, H., Pinto, J. T., Liu, L. U., Merzianu, M., Sordillo, E. M. & Moss, S. F. (2001). *Cancer Lett.* **164**, 127–133.
Weichsel, A., Gasdaska, J. R., Powis, G. & Montfort, W. R. (1996). *Structure*, **4**, 735–751.
Windle, H. J., Fox, A., Ni Eidhin, D. & Kelleher, D. (2000). *J. Biol. Chem.* **275**, 5081–5089.