Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of *Mycobacterium tuberculosis* thioredoxin C

Gareth Hall,[a]* Manish Shah,[a] Paul A. McEwan,[a] Charles Laughton,[a] Malcolm Stevens,[a] Andrew Westwell[b] and Jonas Emsley[a]*

[a]Centre for Biomolecular Sciences, School of Pharmacy, University of Nottingham, Nottingham NG7 2RD, England, and [b]Welsh School of Pharmacy, Cardiff University, Redwood Building, King Edward VII Avenue, Cardiff CF10 3XF, Wales

Correspondence e-mail: paxgh@nottingham.ac.uk, jonas.emsley@nottingham.ac.uk

Received 30 May 2006
Accepted 19 September 2006

PDB Reference: thioredoxin C, 2i1u, r2i1usf.

Mycobacterium tuberculosis is a facultative intracellular parasite of alveolar macrophages. *M. tuberculosis* is able to propagate in harsh environments within cells such as phagocytes, despite being exposed to reactive oxygen and nitrogen intermediates. The thioredoxin redox system is conserved across the phyla and has a well characterized role in resisting oxidative stress and influencing gene expression within prokaryotic and eukaryotic cells. *M. tuberculosis* thioredoxin (*Mtb*Trx) has similar functions in redox homeostasis and it has recently been shown that alkyl hydroperoxidase C is efficiently reduced to its active form by *Mtb*TrxC, supporting this notion. To address whether the *Mtb*Trx has similar features to other thioredoxin structures and to examine the opportunities for designing drugs against this target, *Mtb*TrxC has been crystallized and its structure determined to 1.3 Å resolution. Unexpectedly, the structure demonstrates an interesting crystal packing in which five C-terminal residues from the *Mtb*TrxC fold insert into a groove adjacent to the active site. A very similar interaction is observed in structures of human thioredoxins bound to peptides from the target proteins NF-κB and Ref-1.

1. Introduction

Mycobacterium tuberculosis is one of the most devastating pathogens and is thought to annually infect an estimated eight million people, resulting in 2–3 million deaths (Dye *et al.*, 1999). The pathogenesis of *M. tuberculosis* still remains an expanding global health problem that compels new therapeutic and preventative measures, with the emergence of multi-drug-resistant strains creating a worldwide emergency (Glickman & Jacobs, 2001). The resistance of the organism towards the host immune response and its ability to survive and reactivate at a later stage is poorly understood (Akif *et al.*, 2005). An obligate aerobe, *M. tuberculosis* resides within alveolar mononuclear phagocytes and is thought to use a variety of redox systems to protect itself against the oxidative intermediates generated by the macrophages (Shinnick *et al.*, 1995). Fernando *et al.* (1992) have shown that the thioredoxin system can also aid in the regeneration of inactivated proteins caused by the oxidative stress.

The thioredoxin system, consisting of thioredoxin (Trx), thioredoxin reductase (TrxR) and NADPH, is a ubiquitous redox pathway found in a wide variety of phyla (Holmgren, 1985, 1989). TrxR catalyses the reduction of Trx, which when present in its dithiol form is the main disulfide reductase in

cells (Williams *et al.*, 2000). This redox system has a wide variety of biological functions, including maintaining an intracellular reduced state in the face of an oxidizing extracellular environment (Powis *et al.*, 1995). Trx and TrxR within *M. tuberculosis* are likely to play a similar key role resisting oxidative stress as observed in other prokaryotes, thus making thioredoxin a target for anti-*M. tuberculosis* drugs (Zhang *et al.*, 1999). The *M. tuberculosis* genome encodes three thioredoxin proteins (TrxA, TrxB and TrxC), with the first biochemical inhibitors of *M. tuberculosis* thioredoxin C (*Mtb*TrxC) signalling recently being reported (Shah *et al.*, 2006). Since mycobacteria are devoid of the glutathione-dependent detoxification system, the thioredoxin system should significantly contribute towards the pathogen's defence against oxidative intermediates (Cole *et al.*, 1998) and it has been shown that *Mtb*TrxC dominates this detoxification pathway (Jaeger *et al.*, 2004). In addition, alkyl hydroperoxidase (*ahpC*), which is another system known to combat oxidative stress in *M. tuberculosis*, is shown to require *Mtb*TrxC to reduce it before it becomes catalytically active (Manca *et al.*, 1999).

Thioredoxin proteins are highly conserved across prokaryotes and eukaryotes, having sequence identities of between 27 and 69% (Eklund *et al.*, 1991). *Mtb*TrxC has 50 and 29% sequence identity to *E. coli* and human thioredoxins, respectively. We report the crystal structure of *Mtb*TrxC determined to 1.3 Å resolution, showing a novel crystal packing with five C-terminal residues overlaying the hydrophobic groove of the redox-active site.

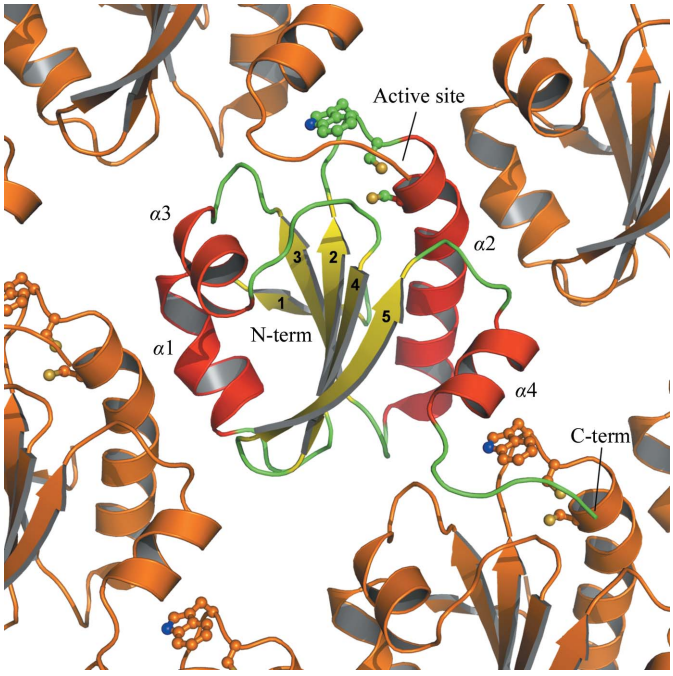

Figure 1
Ribbon diagram portraying the secondary structure and fold of *Mtb*TrxC. The crystal packing present in the primitive triclinic unit cell of *Mtb*TrxC is also shown, portraying the C-terminal tail packed into the active-site groove of a symmetry mate. Figures were prepared using *PyMOL* (DeLano, 2002).

Table 1
Crystallographic statistics.

Values in parentheses are for the highest resolution shell.

Space group	*P*1
Unit-cell parameters (Å, °)	a = 26.48, b = 29.23, c = 30.95, α = 88.37, β = 88.15, γ = 66.67
Resolution limits (Å)	18.90–1.30 (1.37–1.30)
No. of observed/unique reflections	22036/11739
Completeness (%)	2.1 (2.0)
Data redundancy	94.7 (92.3)
R_{merge} (%)	17.8 (22.6)
Mean $I/\sigma(I)$	8.7 (4.2)
R_{work}/R_{free} (%)	21.2 (28.3)/24.0 (34.1)
R.m.s. bond lengths (Å)	0.009
R.m.s. bond angles (°)	1.3
No. of protein residues	108
No. of water molecules	159
Average *B* factor, protein (Å^2)	8.5
R.m.s. *B*, protein (Å^2)	0.7
Average *B* factor, solvent (Å^2)	20.9
Residues in Ramachandran plot regions (%)	
Most favoured	95.8
Additional allowed	4.2

2. Materials and methods

2.1. Cloning, expression and purification

The open reading frame for *M. tuberculosis* TrxC (Rv3914) was amplified from genomic DNA, kindly supplied by T. H. M. Ottenhoff (Leiden University Medical Centre, Leiden, The Netherlands), using standard PCR protocols and the resultant product was cloned into the expression vector pTrcHisA (Invitrogen). This plasmid was transformed into *Escherichia coli* BL21 (DE3) cells for expression. Cultures were grown at 310 K until an OD_{595} of 0.6 was reached, prior to being induced with isopropyl β-D-thiogalactopyranoside (IPTG) and left to incubate overnight at 295 K. The cells were harvested by centrifugation at 4500*g* and resuspended in 20 m*M* Tris–HCl pH 7.9, 5 m*M* imidazole and 0.5 *M* NaCl prior to lysis by sonication (Branson). His-tagged *Mtb*TrxC was purified from the supernatant by Ni^{2+}-affinity chromatography using a pre-charged nickel-chelate affinity resin (Novagen). The protein was eluted using 20 m*M* Tris–HCl pH 7.9, 1 *M* imidazole and 0.5 *M* NaCl, with the *Mtb*TrxC fractions pooled together and dialysed overnight in 1 l 20 m*M* Tris–HCl pH 7.9 and 0.5 *M* NaCl to remove the imidazole. The His tag was cleaved using enterokinase (Sigma) at an enzyme:substrate ratio of 1:100(*w*:*w*), resulting in an extra five N-terminal residues Asp, Arg, Trp, Gly and Ser prior to the start of the native protein sequence. After digestion, the protein was concentrated to a volume of 5 ml and loaded onto a HiLoad Superdex 75 gel-filtration column (Amersham Biosciences) pre-equilibrated with 20 m*M* Tris–HCl pH 7.9 and 0.5 *M* NaCl. Fractions of the protein were collected and analysed by SDS–PAGE before being pooled and concentrated (Vivaspin column) to 3 mg ml^{-1}.

2.2. Crystallization and data collection

Crystals grew in 10% 2-propanol, 0.1 *M* Na HEPES pH 7.5 and 20% PEG 4K using sitting-drop vapour diffusion at 293 K with a total drop volume of 2.0 µl. The crystals belonged to

space group $P1$, with unit-cell parameters a = 26.48, b = 29.23, c = 30.95 Å, α = 88.37, β = 88.15, γ = 66.67°. A single crystal was flash-frozen using liquid nitrogen in mother liquor plus 20% glycerol. An initial data set was collected to 2.0 Å using a Rigaku Micromax-007 X-ray generator (λ = 1.5418 Å) and R-AXIS IV^{++} image plate and a subsequent data set was collected to 1.3 Å resolution at the European Synchrotron Research Facility (ESRF) using beamline ID14-2 (λ = 0.9792 Å). As a result of the crystal form, a number of unique reflections are too close to the rotation axis and are subsequently lost. Therefore, both the high- and low-resolution data sets were indexed and integrated separately using *MOSFLM* (Leslie, 1999) before being merged and then scaled using *SCALA* (Evans, 1993) from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994), giving an overall $R_{\rm merge}$ of 0.178 and a completeness of 94.7% (Table 1).

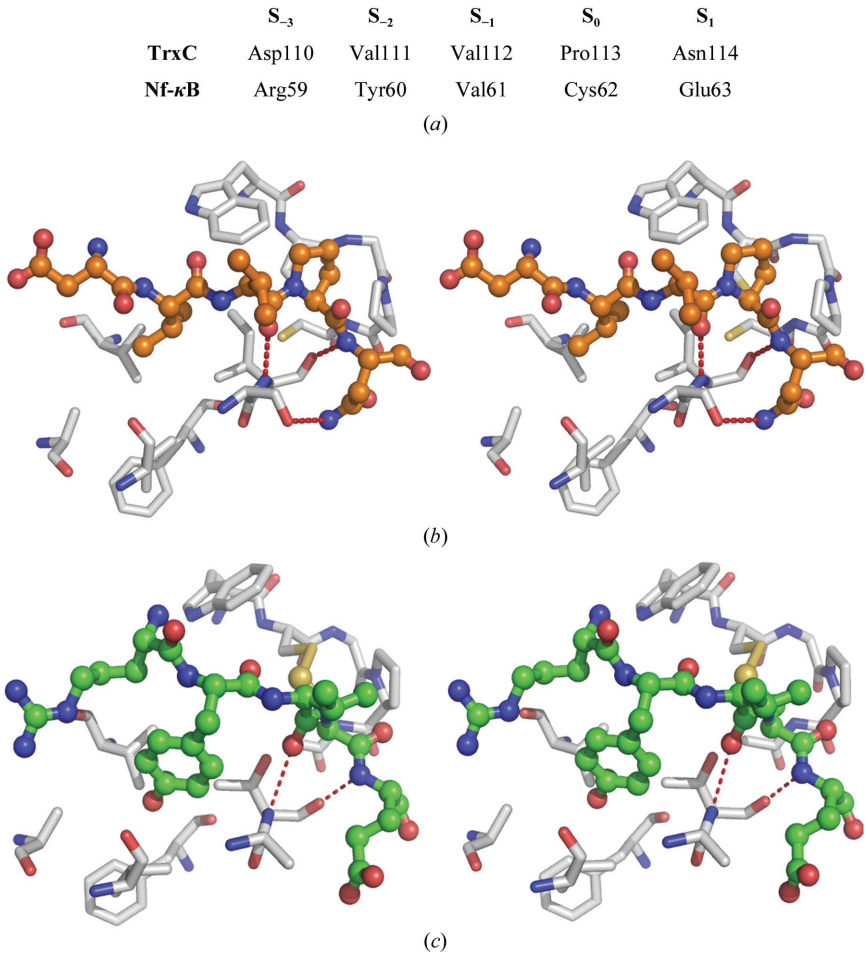

Figure 2
(*a*) Sequence alignment of the polypeptide tail and NF-κB residues interacting with thioredoxin. (*b*) Stereoviews portraying the alignment of the five C-terminal residues of the tail of *Mtb*TrxC and (*c*) five residues of the thioredoxin binding partner NF-κB from the NMR structure 1mdi (Qin *et al.*, 1995), resting in their respective thioredoxin active sites. The residues lining the hydrophobic pockets are also shown.

2.3. Structure determination and refinement

A homology model of *Mtb*TrxC was produced from the *E. coli* thioredoxin crystal structure (Katti *et al.*, 1990; PDB code 2trx) using *SWISSMODEL* (Peitsch & Jongeneel, 1993). The *Mtb*TrxC structure was solved by molecular replacement with *Phaser* (McCoy *et al.*, 2005) using the homology model, locating one molecule in the asymmetric unit. The highest peak in the cross-rotation function gave the correct orientation of the monomer and an R factor of 0.38. The resulting $2F_{\rm o} - F_{\rm c}$ map gave clear electron density, allowing model building of 108 residues. Subsequent refinement took place using *REFMAC*5 (Murshudov *et al.*, 1997), with intermittent cycles of manual model correction using *Coot* (Emsley & Cowtan, 2004). A total of 159 water molecules and seven dual rotamers were modelled using *Coot* to give a final model that has an R factor of 0.212 and $R_{\rm free}$ = 0.240, with 95.8% of residues in the most favoured region and the remainder in the additional allowed area of the Rama-

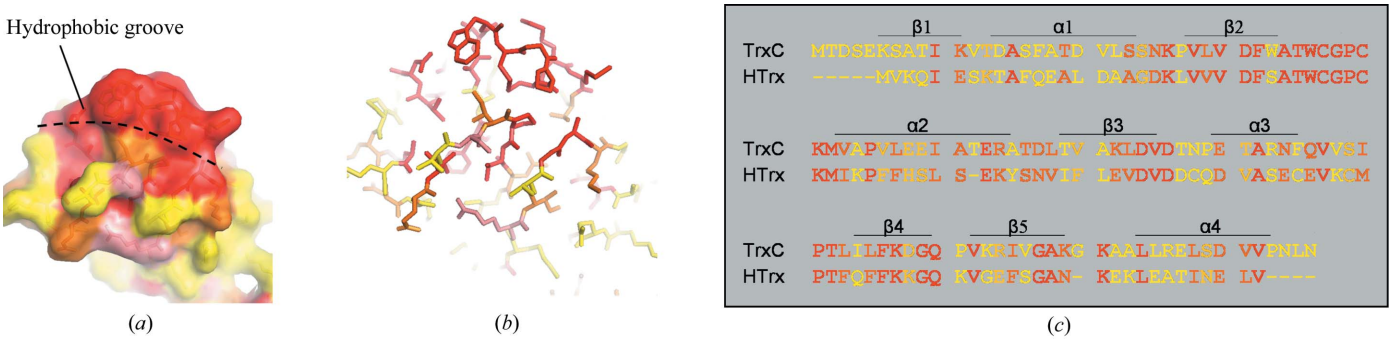

Figure 3
(*a*) Molecular surface and (*b*) ball-and-stick diagrams portraying the sequence conservation between human thioredoxin and *M. tuberculosis* thioredoxin C around the active-site groove, where red shows high conservation and yellow shows low conservation. (*c*) Sequence alignment of *Mtb*TrxC and human thioredoxin (hTrx), with conserved residues coloured in the same scheme and the active site in bold.

chandran plot. Seven N-terminal residues and two C-terminal residues have no interpretable electron density and are assumed to be disordered.

3. Results and discussion

3.1. Molecular structure of *M. tuberculosis* thioredoxin C

M. tuberculosis thioredoxin C has the characteristic thioredoxin fold consisting of a five-stranded β-sheet, three strands of which are parallel and two antiparallel, that forms a hydrophobic core surrounded by four α-helices (Fig. 1). The active site of *Mtb*TrxC is located in the loop region between helix $\alpha2$ and strand $\beta2$ and has a conformation identical to those of other oxidized thioredoxins. The two active-site cysteine residues Cys37 and Cys40 are found at the end of the $\alpha2$ helix, with the Cys40 residue buried more significantly compared with Cys37. The published crystal structure of human thioredoxin (Weichsel *et al.*, 1996; PDB code 1ert) demonstrates dimer formation, which is primarily mediated by a disulfide bridge between Cys73 residues, but also includes numerous hydrophobic contacts and a hydrogen bond between symmetrically related Asp60 residues. Although the dimer has also been observed with a Cys73→Ser mutation, suggesting that the other noncovalent interactions are important, this crystal packing is not seen with other thioredoxin structures. The crystal structure of *Mtb*TrxC does not show any similar dimeric arrangements as observed in the human structures, even though the hydrophobic residues in the region are highly conserved and the relevent aspartic acid residue is still present.

3.2. C-terminal polypeptide tail

A remarkable feature of the *Mtb*TrxC structure is that the five C-terminal residues, consisting of Asp110, Val111, Val112, Pro113 and Asn114, lie in the active-site groove of an adjacent molecule in the crystal (Fig. 1). These residues might have been expected to continue the $\alpha4$ helix, with the three hydrophobic residues Val111, Val112 and Pro113 interacting with the end of the $\beta4$ strand and the start of the $\alpha5$ helix, as seen in other thioredoxin crystal structures. The interaction is comprised of significant hydrophobic and hydrogen-bonding contacts. The hydrophobic packing interactions occur between Pro113 and Trp35 and between Val111 and Val65, Val77 and Ile80, with Val111 sitting in a hydrophobic pocket at the end of the groove. Three hydrogen bonds of less than 3.2 Å feature in the polypeptide interactions: between Ile80 amide and Val108 carbonyl oxygen, Ile80 carbonyl oxygen and Asn110 amide and Ser79 hydroxyl oxygen and Asn110 side-chain amide group. All of these contacts allow the C-terminal polypeptide tail to follow the curvature of the protein and rest between the loop regions of $\beta2$–$\alpha2$ and $\alpha3$–$\beta4$ in a hydrophobic groove that measures 3–4 Å in depth, 3–4 Å in breadth and 11–12 Å in length. Instances of residues from symmetry-related molecules packing into the active site of thioredoxin and thioredoxin-like structures have been reported before in *Anabaena* thioredoxin (Saarinen *et al.*, 1995; PDB code 1thx) and in tryparedoxins (Alphey *et al.*, 2003; PDB code 1o73). However, in both examples the interaction involves soft crystal contacts containing one or two residues, whereas the C-terminal tail found in *Mtb*TrxC is seen to follow the curvature of the structure along the redox-active groove and involves more than five residues.

A polypeptide chain interacting with this hydrophobic groove has also been observed in NMR structures of human thioredoxin and residues from its ligand-binding partners NF-κB and Ref-1 (Qin *et al.*, 1995, 1996). Although both of these proteins use cysteine residues to covalently interact with the active site of the thioredoxin, they have local polypeptide sequences that somewhat resemble the *MTb*TrxC tail. Using a nomenclature similar to that described by Qin *et al.* (1995, 1996), where the residue of the binding polypeptide adjacent to active-site cysteine (Cys32 for hTrx, Cys37 for *Mtb*TrxC) is designated S_0, it is possible to describe some important similarities that may give rise to the interaction seen in our structure. For the human thioredoxin-NF-κB complex, S_0 is cysteine, whereas for the polypeptide tail observed in *Mtb*TrxC S_0 is proline (Fig. 2).

Qin *et al.* (1995, 1996) describe an important interaction made by the side chain of a hydrophobic residue at position S_{-2}, which sits in a hydrophobic pocket on the surface of thioredoxin. In the NF-κB and Ref-1 complexes the residue is tyrosine and tryptophan, respectively, and the associated pocket (in human thioredoxin) is formed by residues Phe27, Val59, Ala66, Val71 and Met74. In comparison, the *Mtb*TrxC crystal structure has Val111 in the S_{-2} position and the hydrophobic pocket is formed by residues Phe32, Val65, Ala72, Val77 and Ile80. Thus, we see that the amino acids from the respective thioredoxins involved in these hydrophobic interactions are highly conserved, although owing to its smaller size there are fewer residues found interacting with Val111 of the *Mtb*TrxC polypeptide tail. In fact, most of the residues conserved between all thioredoxins, outside of the active site, are found around this hydrophobic pocket (Fig. 3). The most highly conserved feature across the three complexes is a hydrogen bond that exists between the backbone N atom of the S_1 residue of the binding peptide and the carbonyl O atom of either Thr74 (human) or Ile80 (*Mtb*TrxC) (Fig. 2). This could represent a conserved feature of thioredoxin protein–protein interactions when functioning as a redox protein, signalling modulator or chaperone.

4. Conclusions

The structure of *Mtb*TrxC reveals a five-residue polypeptide tail packing into the shallow groove that contains the active-site cysteine residues of a neighbouring molecule. This interaction undoubtedly arises as a result of crystallization, but it is nevertheless structurally very similar to the protein–protein interactions formed by human thioredoxin with the ligands NF-κB and Ref-1 (Qin *et al.*, 1995, 1996).

As thioredoxin is noted for its solubility, a predominantly hydrophobic outlying region, such as these five C-terminal residues, disrupts the tight and compact nature of the thio-

redoxin fold. The question raised is whether the extended polypeptide tail exists in the solution structure of *Mtb*TrxC and whether it has a discrete function. It seems likely that the polypeptide-tail conformation seen here is a component of the crystallization process and in its natural state the C-terminal residues would continue the fourth α-helix as predicted. If this is the case, then it must be assumed that the energy required to partially unwind the α-helix is compensated by the stable interactions formed between the polypeptide tail and the adjacent protein in the crystal during the crystallization process.

Although arguably artificial, the polypeptide-tail interaction, along with those found between human thioredoxin and its target proteins, allows us to gain a better understanding of the binding interactions that can occur within the active site. It may also be potentially used to allow us to understand the peptide sequence that thioredoxin recognizes within these target proteins. This will enable us to predict the target regions that thioredoxin recognizes in proteins known to interact with this redox protein.

This conserved interaction motif indicates the likelihood that prokaryotic thioredoxins will also utilize this groove to interact with ligands in a manner reminiscent of human thioredoxin. In *E. coli*, a proteome analysis resulted in the identification of a total of 80 proteins found to be associated with the *E. coli* Trx1, implicating thioredoxin in at least 26 distinct cellular processes, including cell division, transcriptional regulation, energy transduction, protein folding and degradation, and several biosynthetic pathways (Kumar *et al.*, 2004; Zeller & Gabriele, 2006). The exact ligand-binding partners for *Mtb*TrxC are currently unknown, but they are likely to be linked to redox homeostasis and gene regulation and may form conserved interactions utilizing the thioredoxin groove in a manner reminiscent of the crystal structure packing observed here.

References

Akif, M., Suhre, K., Verma, C. & Mande, S. C. (2005). *Acta Cryst.* D**61**, 1603–1611.

Alphey, M. S., Gabrielsen, M., Micossi, E., Leonard, G. A., McSweeney, S. M., Ravelli, R. B., Tetaud, E., Fairlamb, A. H., Bond, C. S. & Hunter, W. N. (2003). *J. Biol. Chem.* **278**, 25919–25925.

Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, CA, USA.

Dye, C., Scheele, S., Dolin, P., Pathania, V. & Raviglione, M. C. (1999). *JAMA*, **282**, 677–686.

Eklund, H., Gleason, F. K. & Holmgren, A. (1991). *Proteins*, **11**, 13–28.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Evans, P. R. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 114–122. Warrington: Daresbury Laboratory.

Fernando, M. R., Nanri, H., Yoshitake, S., Nagata-Kuno, K. & Minakami, S. (1992). *Eur. J. Biochem.* **209**, 917–922.

Glickman, M. S. & Jacobs, W. R. Jr (2001). *Cell*, **104**, 477–485.

Holmgren, A. (1985). *Annu. Rev. Biochem.* **54**, 237–271.

Holmgren, A. (1989). *J. Biol. Chem.* **264**, 13963–13966.

Jaeger, T., Budde, H., Flohe, L., Menge, U., Singh, M., Trujillo, M. & Radi, R. (2004). *Arch. Biochem. Biophys.* **423**, 182–191.

Katti, S. K., LeMaster, D. M. & Eklund, H. (1990). *J. Mol. Biol.* **212**, 167–184.

Kumar, J. K., Tabor, S. & Richardson, C. C. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 3759–3764.

Leslie, A. G. W. (1999). *Acta Cryst.* D**55**, 1696–1702.

McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.

Manca, C., Paul, S., Barry, C. E. III, Freedman, V. H. & Kaplan, G. (1999). *Infect. Immun.* **67**, 74–79.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Peitsch, M. C. & Jongeneel, V. (1993). *Int. Immunol.* **5**, 233–238.

Powis, G., Briehl, M. & Oblong, J. (1995). *Pharmacol. Ther.* **68**, 149–173.

Qin, J., Clore, G. M., Kennedy, W. P., Huth, J. R. & Gronenborn, A. M. (1995). *Structure*, **3**, 289–297.

Qin, J., Clore, G. M., Kennedy, W. P., Kuszewski, J. & Gronenborn, A. M. (1996). *Structure*, **4**, 613–620.

Saarinen, M., Gleason, F. K. & Eklund, H. (1995). *Structure*, **3**, 1097–1108.

Shah, M., Wells, G., Bradshaw, T. D., Laughton, C. A., Stevens, M. F. G. & Westwell, A. D. (2006). *Lett. Drug Des. Discov.* **3**, 419–423.

Shinnick, T. M., King, C. H. & Quinn, F. D. (1995). *Am. J. Med. Sci.* **309**, 92–98.

Weichsel, A., Gasdaska, J. R., Powis, G. & Montfort, W. R. (1996). *Structure*, **4**, 735–751.

Williams, C. H., Arscott, L. D., Muller, S., Lennon, B. W., Ludwig, M. L., Wang, P. F., Veine, D. M., Becker, K. & Schirmer, R. H. (2000). *Eur. J. Biochem.* **267**, 6110–6117.

Zeller, T. & Gabriele, K. (2006). Submitted.

Zhang, Z., Hillas, P. J. & Ortiz de Montellano, P. R. (1999). *Arch. Biochem. Biophys.* **363**, 19–26.