Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

addenda and errata

Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of DapA (Rv2753c) from *Mycobacterium tuberculosis*. Corrigendum

Georgia Kefala‡ and Manfred S. Weiss*

EMBL Hamburg Outstation, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany

‡ Present address: Salk Institute, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA. Correspondence e-mail: msweiss@embl-hamburg.de

A correction is made to the *Experimental methods* section of the article by Kefala & Weiss [(2006), *Acta Cryst.* F**62**, 1116–1119].

The first sentence of the *Experimental methods* section of the article by Kefala & Weiss (2006) should read as follows: Each gram of cell pellet was dissolved in 10 ml buffer *A* [20 m*M* Tris pH 8.0, 250 m*M* NaCl, 10 m*M* imidazole, 5%(*v*/*v*) glycerol, 2 m*M* β-mercaptoethanol] plus 10%(*v*/*v*) 10× BugBuster (Novagen) and one Complete Mini EDTA-free protease-inhibitor cocktail tablet (Roche) per 20 ml, and then lysed by sonication four times for 5 min at a time using 0.25 s pulses at 277 K.

The authors would like to thank Dr Simone Weyand, Imperial College London, UK, for pointing out an ommission in the original article.

References

Kefala, G. & Weiss, M. S. (2006). *Acta Cryst.* F**62**, 1116–1119.

  doi:10.1107/S1744309107065566

Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of DapA (Rv2753c) from *Mycobacterium tuberculosis*

Georgia Kefala‡ and Manfred S. Weiss*

EMBL Hamburg Outstation, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany

‡ Present address: Salk Institute, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA.

Correspondence e-mail: msweiss@embl-hamburg.de

Received 5 August 2006
Accepted 28 September 2006

Dihydrodipicolinate synthase from *Mycobacterium tuberculosis* (DHDPS, DapA, Rv2753c) has been cloned, heterologously expressed in *Escherichia coli*, purified using standard chromatographic techniques and crystallized in a monoclinic crystal form. Preliminary diffraction data analysis suggests the presence of two independent tetramers in the asymmetric unit in almost the same relative orientation.

1. Introduction

Mycobacterium tuberculosis (Mtb) infection remains the leading cause of death from a single infectious disease (World Health Organization, 2001), with two million deaths and over eight million new tuberculosis cases annually (Dye *et al.*, 1999). The current WHO-recommended strategy for tuberculosis control, known as directly observed treatment, short-course (DOTS), requires patients to adhere to a drug regimen comprising isoniazid, rifampin, pyrazinamide and ethambutol for a minimum of six months. This prolonged and complicated therapy frequently causes severe side effects, which in turn often results in patient noncompliance. As a consequence, many patients relapse and multi-drug-resistant Mtb strains emerge. The only available human vaccine, *Mycobacterium bovis*-derived bacillus Calmette–Guérin (BCG), provides inconsistent efficacy, with reported variation between 0 and 80% (Fine, 1995). The increase in tuberculosis cases worldwide, particularly amongst immunocompromized individuals, combined with the increase of multi-drug-resistant strains highlights the need for new anti-tuberculosis drugs.

Potential new and effective drugs should inhibit proteins that are essential for bacterial viability but are not present in humans. The lysine-biosynthesis pathway meets both criteria, providing multiple targets for drug design (Hutton *et al.*, 2003). Both the final product of the pathway and the lysine precursor diaminopimelic acid (DAP) are essential for bacterial survival and growth. Whereas lysine is needed for protein biosynthesis, DAP is an important component for the cross-linking of the peptidoglycan layer of mycobacterial cell walls (Cummins & Harris, 1956). Lysine is produced in a nine-step pathway starting from L-aspartate. The first two steps, common also to the glycine, serine and threonine pathways, result in the formation of L-aspartate-β-semialdehyde (L-ASA). DHDPS (dihydrodipicolinate synthase, DapA; EC 4.2.1.52) then catalyses the first unique reaction of L-lysine biosynthesis: an aldol condensation between L-ASA and pyruvate. The product of the reaction is the unstable heterocycle (4*S*)-4-hydroxy-2,3,4,5-tetrahydro-(2*S*)-dipicolinate. The proposed ping-pong mechanism involves Schiff-base formation between pyruvate and an active-site lysine residue, followed by the release of water and enamine formation, which has been suggested to be irreversible. L-ASA then binds to the stable substituted-enzyme form, followed by release of the product and regeneration of the enzyme (Blickling, Renner *et al.*, 1997).

The three-dimensional structures of DHDPSs from *Escherichia coli* (Mirwaldt *et al.*, 1995; Blickling, Renner *et al.*, 1997), *Thermotoga maritima* (Joint Center for Structural Genomics, http://www.jcsg.org, unpublished data), *Nicotiana sylvestris* (Blickling, Beisel *et al.*, 1997)

Table 1
Data-collection and processing statistics.

Values in parentheses are for the highest resolution bin (2.32–2.28 Å).

No. of crystals	1
Wavelength (Å)	0.981
Crystal-to-detector distance (mm)	170
Rotation range per image (°)	0.5
Total rotation range (°)	499.5
Resolution range (Å)	99.0–2.28 (2.32–2.28)
Space group	$P2_1$
Unit-cell parameters (Å, °)	a = 94.79, b = 87.37, c = 139.85, β = 107.78
Mosaicity (°)	0.72
Total No. of reflections	1011351
Unique reflections	99300
Redundancy	10.2
$I/\sigma(I)$	28.0 (12.5)
Completeness (%)	99.9 (100.0)
$R_{\rm merge}$ (%)	8.5 (19.2)
$R_{\rm r.i.m.}$† (%)	9.0 (20.3)
$R_{\rm p.i.m.}$† (%)	2.8 (6.4)
Overall B factor from Wilson plot (Å^2)	30.9
Optical resolution (Å)	1.71

† $R_{\rm r.i.m.}$ is the redundancy-independent merging R factor, $R_{\rm r.i.m.} = 100\sum_{hkl}[N/(N-1)]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, and $R_{\rm p.i.m.}$ is the precision-indicating merging R factor, $R_{\rm p.i.m.} = 100\sum_{hkl}[1/(N-1)]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, with N being the redundancy of a given reflection (Weiss, 2001).

and more recently from *Bacillus anthracis* (Blagova *et al.*, 2006) have been determined. In all cases the enzyme is a homotetramer. The monomers show a $(\beta/\alpha)_8$-barrel fold with a C-terminal extension, which seems to have no obvious function with respect to the catalytic mechanism or the regulation of the enzyme.

Here, we report the successful cloning, expression, purification and crystallization of Rv2753c from *M. tuberculosis* and the characterization of the obtained crystals by X-ray crystallographic methods.

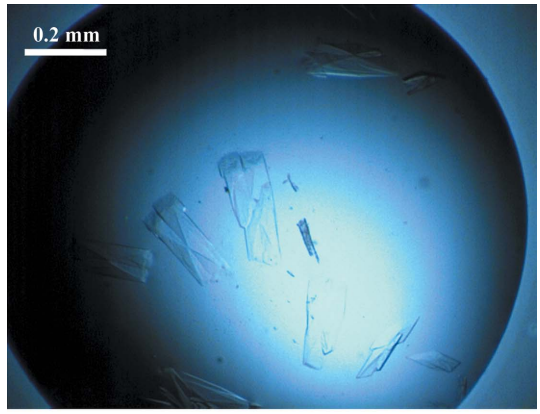
2. Experimental methods

2.1. Cloning and heterologous expression

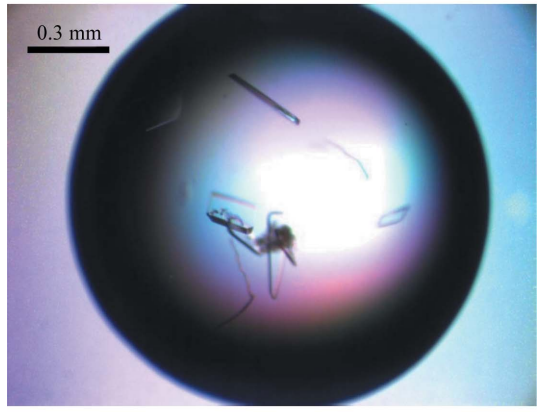
The 903 bp Rv2753c open reading frame was amplified by PCR from Mtb H37Rv genomic DNA. The following oligonucleotides (obtained from MWG Biotech) were used as forward and reverse primers: 5′-AAAA**CCATGG**TGACCACCGTCGGATTCGACG-TCG-3′ (forward) and 5′-AAAA**CTCGAG**TTACCGAAGCACCG-AGGCCGCGCGCA-3′ (reverse). The amplified fragment containing 5′-*Nco*I and 3′-*Xho*I restriction sites (shown in bold in the primer sequence) was digested and ligated to the pETM-11 expression vector (EMBL) digested with the same restriction enzymes. Upstream of the *Nco*I site the vector contains a His_6 tag followed by a tobacco etch virus (TEV) protease-cleavage site. This adds three amino acids (Gly, Ala and Met) to the N-terminus of the protein after TEV cleavage. The construct was sequenced to confirm the cloning of the *dapA* gene sequence in frame. The recombinant plasmid was used to transform *E. coli* BL21-CodonPlus(DE3)-RP cells (Stratagene) for expression of the 31 kDa protein. Cells from an overnight 5 ml preculture were grown in LB broth medium containing kanamycin (30 µg ml^{-1}) and chloramphenicol (10 µg ml^{-1}) at 310 K and 200 rev min^{-1}. The culture was induced with 0.25 m*M* isopropyl β-D-thiogalactopyranoside (IPTG) at an $OD_{600\,\rm nm}$ of approximately 0.6 at 293 K. Following induction, the culture was incubated for about 15 h at 293 K and 220 rev min^{-1} and was then harvested. The cells were stored at 193 K until further processing.

2.2. Protein purification

Each gram of cell pellet was dissolved in 10 ml buffer *A* [20 m*M* Tris pH 8.0, 250 m*M* NaCl, 10 m*M* imidazole, 5%(*v*/*v*) glycerol, 2 m*M* β-mercaptoethanol] plus 10% BugBuster (Novagen) and one Complete Mini EDTA-free protease-inhibitor cocktail tablet (Roche) per 20 ml and then lysed by sonication four times for 5 min at a time using 0.25 s pulses at 277 K. After each 5 min sonication run, the ice around the vial was replenished. The cell debris was pelleted by centrifugation for 60 min at 277 K and 20 000 rev min^{-1}. The crude lysate was filtered through a 0.22 µm membrane and was loaded onto a 5 ml Hi-Trap Chelating HP column (GE Healthcare) charged and equilibrated with Ni^{2+} and buffer *A*, respectively. In order to remove unbound proteins, the column was first washed with 25 ml of buffer *A*, then with 25 ml buffer *B* [20 m*M* Tris pH 8.0, 1 *M* NaCl, 10 m*M* imidazole, 5%(*v*/*v*) glycerol, 2 m*M* β-ME] and then with 25 ml buffer *C* [20 m*M* Tris pH 8.0, 250 m*M* NaCl, 50 m*M* imidazole, 5%(*v*/*v*) glycerol, 2 m*M* β-ME]. The protein was eluted by running a linear gradient from 50 to 800 m*M* imidazole (in buffer *C*). The peak fractions were pooled and dialyzed against buffer *D* [20 m*M* Tris pH 8.0, 250 m*M* NaCl, 5%(*v*/*v*) glycerol, 2 m*M* β-ME, 5 m*M* EDTA]. During overnight dialysis at 277 K, the His_6 tag was cleaved off by adding recombinant TEV protease in a molar ratio of 1:100. The dialysis was then continued for several more hours against buffer *D* without EDTA and the cleaved and dialyzed protein mixture was then passed through an Ni–NTA column pre-equilibrated with buffer *D* without EDTA. The protein in the flow-



(*a*)



(*b*)

Figure 1
Crystals of Rv2753c from *M. tuberculosis*. (*a*) Initial crystals observed under the conditions 30%(*w*/*v*) PEG 4000, 180 m*M* $MgCl_2$, 100 m*M* Tris–HCl pH 8.5; (*b*) crystal obtained under the optimized crystallization conditions 28%(*w*/*v*) PEG 4000, 170 m*M* $MgCl_2$, 100 m*M* Tris–HCl pH 8.5.

through was subsequently purified by gel filtration (Superdex 200, 16/60, GE Healthcare) using buffer *D* containing 5 m*M* DTT instead of 2 m*M* β-ME for both equilibration and elution. The peak fractions analyzed by SDS–PAGE were pooled and concentrated. The purity of the protein was estimated by SDS–PAGE and its oligomeric state was probed by dynamic light scattering (DLS).

2.3. Crystallization

The initial crystallization screening was performed in 24-well plates using the hanging-drop vapour-diffusion method. 1 µl protein solution (15 mg ml^{-1}) and 1 µl reservoir solution taken from Crystal Screens 1 and 2 (Hampton Research) were equilibrated against 1 ml reservoir solution. The first crystals were observed in the condition 30%(*w*/*v*) PEG 4000, 180 m*M* $MgCl_2$, 100 m*M* Tris–HCl pH 8.5 at 293 K. Subsequently, this condition was optimized to 28%(*w*/*v*) PEG 4000, 170 m*M* $MgCl_2$, 100 m*M* Tris–HCl pH 8.5.

2.4. Diffraction data collection and processing

Crystals of the native protein were treated with 20%(*v*/*v*) MPD in reservoir solution as a cryoprotectant and flash-cooled in a nitrogen stream at 100 K. Diffraction data were then collected on the XRD beamline at the ELETTRA synchrotron (Trieste, Italy) using a MAR CCD (165 mm) detector. The data were indexed and integrated using *DENZO* (Otwinowski & Minor, 1997) and scaled using *SCALEPACK* (Otwinowski & Minor, 1997). The redundancy-independent merging *R* factor $R_{\rm r.i.m.}$ as well as the precision-indicating merging *R* factor $R_{\rm p.i.m.}$ (Weiss, 2001) were calculated using the program *RMERGE* (available from http://www.embl-hamburg.de/~msweiss/projects/msw_qual.html or from MSW upon request). The relevant data-collection and processing parameters are given in Table 1. Intensities were converted to structure-factor amplitudes using the program *TRUNCATE* (French & Wilson, 1978; Collaborative Computational Project, Number 4, 1994). The optical resolution of the data set was calculated using the program *SFCHECK* (Vaguine *et al.*, 1999). Calculation of the self-rotation function was carried out using the program *MOLREP* (Collaborative Computational Project, Number 4, 1994) and structure-factor amplitudes to a maximum resolution of 4.0 Å. The native Patterson synthesis was calculated using the program *FFT* (Collaborative Computational Project, Number 4, 1994).

3. Results and discussion

Expression of Rv2753c in *E. coli* BL21-CodonPlus(DE3)-RP cells resulted in approximately 40% of the protein in the soluble fraction and 60% in inclusion bodies. After the two-step chromatographic procedure, the final yield of pure protein was approximately 1 mg from a 1 l culture. From the gel-filtration column, the protein eluted with an apparent molecular weight of approximately 120 kDa, which is consistent with a tetrameric state. This result was confirmed by DLS.

The purity of the sample prepared for crystallization was at least 95% as estimated by SDS–PAGE. Crystal growth was observed about two weeks after the experiments had been set up. Initially crystals grew with various defects (Fig. 1*a*), but after optimization of the crystallization conditions single crystals suitable for diffraction experiments were obtained. The crystals typically grew to dimensions of about 300 × 100 × 50 µm, although they sometimes grew as large as 1000 × 200 × 50 µm. They diffract X-rays to about 2.0 Å resolution.

From one crystal of Rv2753c, an X-ray diffraction data set with a redundancy of about 10 was collected at the X-ray diffraction beamline of the ELETTRA synchrotron (Trieste, Italy) to a resolution of 2.28 Å. The data-processing and scaling statistics (Table 1) indicate that the data set is of good quality. The $I/\sigma(I)$ value of 12.5 in the outer resolution bin suggests that the crystals diffract significantly further than 2.28 Å, which is corroborated by the value of 1.71 Å for the optical resolution (Table 1). The space-group assignment turned out to be difficult. The diffraction pattern could be indexed in a primitive monoclinic setting with unit-cell parameters *a* = 94.79,

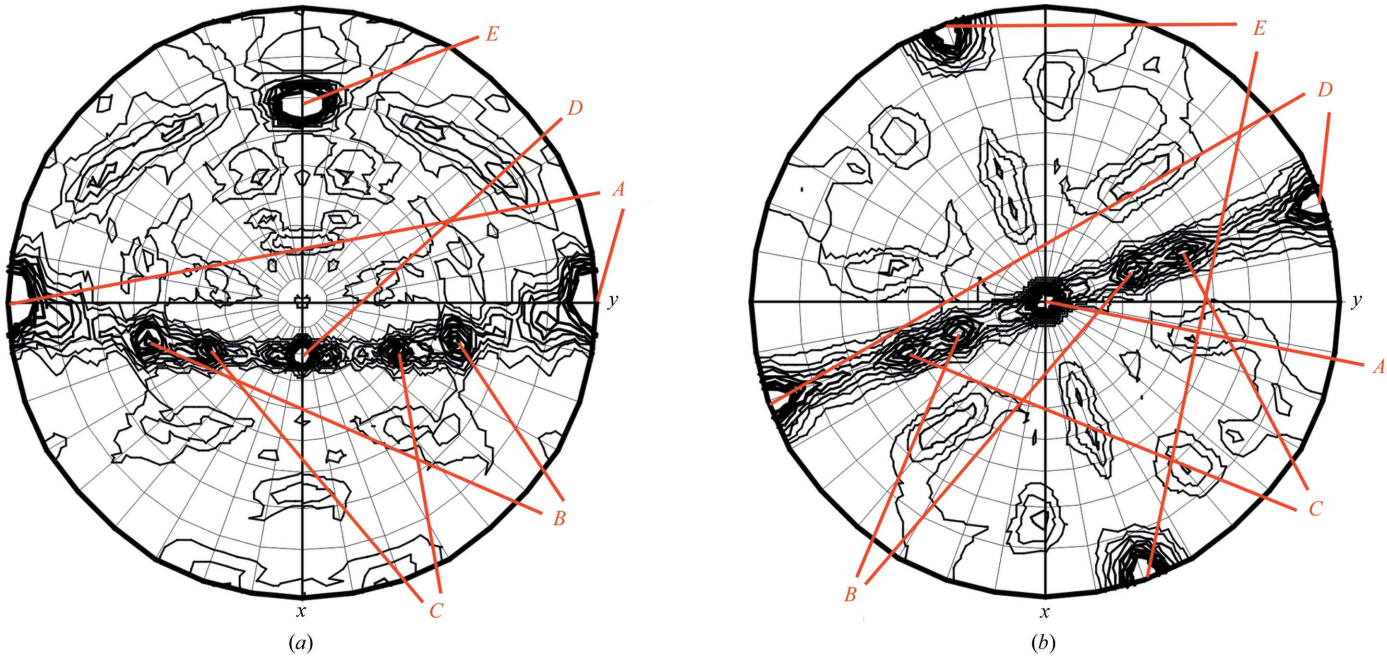


Figure 2
Self-rotation function, $\kappa = 180^\circ$ section, based on data collected from a crystal of Rv2753c. (*a*) Monoclinic setting and (*b*) pseudo-orthorhombic setting of the crystal axes. This figure was produced using the program *MOLREP* (Collaborative Computational Project, Number 4, 1994).

b = 87.37, c = 139.85 Å, β = 107.78°, but because of the approximate relationship $a \simeq -2c\cos\beta$, indexing was also possible in a *C*-centred orthorhombic setting with unit-cell parameters a = 94.79, b = 263.15, c = 87.37 Å, especially when the indexing was based on only one diffraction image. However, data reduction was only possible assuming monoclinic symmetry. An explanation for this phenomenon is provided by examining the self-rotation function (Fig. 2). The stereographic projection of the peaks in the self-rotation function in the monoclinic setting (Fig. 2*a*) shows the crystallographic twofold along the *y* direction (indicated by the two peaks *A* on the equator) and six further peaks indicating the presence of a tetramer with D_2 symmetry. Two of the six peaks occur in the *xz* plane (peaks *D* and *E* lying on the meridian) and are therefore perpendicular to the crystallographic twofold along the *y* direction. If the data are indexed in the *C*-centred orthorhombic setting and reduced without any assumption of symmetry, the stereographic projection of the self-rotation function looks as in Fig. 2(*b*). The *y* axis (*A*) of the monoclinic system becomes the *z* axis in the orthorhombic systems and the two noncrystallographic axes which occur in the *xz* plane in the monoclinic system (*D* and *E*) now occur in the *xy* plane in the orthorhombic system (peaks *D* and *E* on the equator in Fig. 2*b*). However, since these two axes are not parallel to the unit-cell axes, the data cannot be reduced assuming orthorhombic symmetry. Thus primitive monoclinic has to be the correct crystal system.

The self-rotation function in the monoclinic setting (Fig. 2*a*) shows four independent peaks (*B*, *C*, *D* and *E*) in addition to the crystallographic twofold along the *y* direction (*A*), although a tetramer exhibiting D_2 symmetry will produce only three independent peaks. A possible explanation of this observation is that peaks *B*, *C* and *E* form the triplet defining the D_2 symmetry of the tetramer and peak *D* is just a consequence of the fact that a noncrystallographic twofold (*E*) is oriented perpendicular to a crystallographic twofold (*A*). However, the solvent content for one tetramer per asymmetric unit would be 72% (based on a molecular weight of 31 086 Da and a unit-cell volume of 1 102 889.5 Å^3) and for two tetramers it would be 45% (Matthews, 1968). Since no further peaks are observed in the self-rotation function, a second tetramer has to be oriented almost exactly like the first tetramer. Support for this comes from the observation that some of the peaks, most notably peaks *B* and *E*, in the self-rotation function are elongated and by the occurrence of a peak in the native Patterson synthesis (calculated to a maximum resolution of 4.0 Å, not shown) at fractional coordinates (x, y, z) = (0.50, 0.46, 0.47).

A second plausible explanation of the self-rotation function would be that peaks *B*, *C* and *E* form a triplet describing the D_2 symmetry of one tetramer and that peaks *A*, *D* and *E* form a triplet describing a second tetramer of D_2 symmetry, in which one of the non-crystallographic axes is oriented parallel to the crystallographic twofold. In this case, the two parallel twofold axes will generate a translation with a component of 0.5 in the *y* direction. Since at lower resolution (6 Å) the peak in the native Patterson synthesis at fractional coordinates (x, y, z) = (0.50, 0.46, 0.47) and its symmetry mate at (0.50, 0.54, 0.47) collapses into one peak at (0.50, 0.50, 0.47), this second scenario remains a possibility, although it does not provide an explanation for the splitting of the peak in the native Patterson at higher resolution.

In summary, the most likely situation is that two DapA tetramers are present in the asymmetric unit of the described monoclinic crystal form. It is currently not clear which one of the two described scenarios is correct, although the first one seems to provide a slightly better explanation of the available data.

A number of DapA structures from other organisms have already been published (Mirwaldt *et al.*, 1995; Blickling, Renner *et al.*, 1997; Joint Center for Structural Genomics, http://www.jcsg.org, unpublished data; Blickling, Beisel *et al.*, 1997; Blagova *et al.*, 2006) and deposited in the Protein Data Bank. These proteins share 25–35% amino-acid sequence identity with DapA from *M. tuberculosis*. Thus, structure solution will be attempted using molecular-replacement methods.

We would like to thank Dr Jeanne Perry (University of California at Los Angeles, USA) for providing genomic Mtb-DNA, Dr Arie Geerlof (EMBL) for help with the solubility screens, Dr Kay Diederichs (University of Konstanz, Germany) for discussions, the TB consortium (http://www.doe-mbi.ucla.edu/TB) for postdoctoral exchange grants to GK and the X-Mtb consortium (http://www.xmtb.org) for funding through BMBF/PTJ grant No. BIO/0312992 A. We would also like to acknowledge the support of this work by the EC Fifth Framework Program 'Transnational Access to Major Research Infrastructures' (Contract No. HPRI-CT-1999-00033) and the staff of the ELETTRA synchrotron (Trieste, Italy) for providing beam time.

References

Blagova, E., Levdikov, V., Milioti, N., Fogg, M. J., Kalliomaa, A. K., Brannigan, J. A., Wilson, K. S. & Wilkinson, A. J. (2006). *Proteins*, **62**, 297–301.
Blickling, S., Beisel, H. G., Bozic, D., Knablein, J., Laber, B. & Huber R. (1997). *J. Mol. Biol.* **274**, 608–621.
Blickling, S., Renner, C., Laber, B., Pohlenz, H. D., Holak, T. A. & Huber, R. (1997). *Biochemistry*, **36**, 24–33.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cummins, C. S. & Harris, H. (1956). *J. Gen. Microbiol.* **14**, 583–600.
Dye, C., Scheele, S., Dolin, P., Pathania, V. & Raviglione, M. C. (1999). *J. Am. Med. Assoc.*, **282**, 677–686.
Fine, P. E. (1995). *Lancet*, **346**, 1339–1345.
French, G. S. & Wilson, K. S. (1978). *Acta Cryst.* A**34**, 517–525.
Hutton, C. A., Southwood, T. J. & Turner, J. J. (2003). *Mini Rev. Med. Chem.* **3**, 115–127.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mirwaldt, C., Korndörfer, I. & Huber, R. (1995). *J. Mol. Biol.* **246**, 227–239.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.
Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.
World Health Organization (2001). *Global Tuberculosis Control.* Geneva, Switzerland: World Health Organization.