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

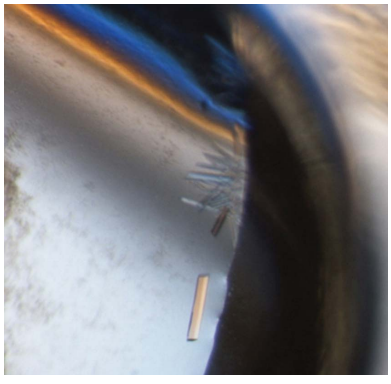
Expression, purification and preliminary crystallographic analysis of *O*-acetylhomoserine sulfhydrylase from *Mycobacterium tuberculosis*

Jiang Yin,[a]* Craig R. Garen,[a] Katherine Bateman,[a] Minmin Yu,[b] Emily Z. Alipio Lyon,[c] Jeff Habel,[b] Heungbok Kim,[c] Li-wei Hung,[b] Chang-Yub Kim[c] and Michael N. G. James[a]*

[a]Protein Structure and Function Group, Department of Biochemistry, School of Molecular and Systems Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada, [b]E. O. Lawrence Berkeley National Laboratory, University of California at Berkeley, California 94720, USA, and [c]Bioscience Division, MS M888, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Correspondence e-mail: jyin@ualberta.ca, michael.james@ualberta.ca

Received 31 March 2011
Accepted 10 May 2011



The gene product of the open reading frame Rv3340 from *Mycobacterium tuberculosis* is annotated as encoding a probable *O*-acetylhomoserine (OAH) sulfhydrylase (MetC), an enzyme that catalyzes the last step in the biosynthesis of methionine, which is an essential amino acid in bacteria and plants. Following overexpression in *Escherichia coli*, the *M. tuberculosis* MetC enzyme was purified and crystallized using the hanging-drop vapor-diffusion method. Native diffraction data were collected from crystals belonging to space group $P2_1$ and were processed to a resolution of 2.1 Å.

1. Introduction

Infection by *Mycobacterium tuberculosis* (*Mtb*) presently kills more humans than any other infectious disease. One third of the world's population is infected with *Mtb*, with millions of new cases accumulating every year. In 2006, 1.7 million people died from tuberculosis alone or in combination with AIDS, according to a report by the World Health Organization (WHO; http://www.who.int/tb).

Mtb has the ability to persist in the macrophages of an infected person for a long period of time. Such a capacity not only ensures the survival of *Mtb via* evasion of the host's humoral immune system, but also allows the pathogen to reactivate its life cycle when the host's immune system has weakened, *e.g.* in the event of a secondary infection such as HIV (Nunn *et al.*, 2005). This creates one of the key obstacles in treating tuberculosis (TB) patients: the treatment regimen is lengthy (6–9 months in order to prevent relapses) and patient noncompliance is quite common. Thus, it is no surprise that there has been a rise in cases of TB infection in recent years, especially in less developed countries. Owing to the rapid appearance of drug-resistant *Mtb* strains, existing anti-TB therapy has become ineffective in containing the disease in certain populations. Novel anti-TB drugs are badly needed and eagerly anticipated.

The TB Structural Genomics Consortium (http://www.webtb.org) was formed in 2000 with the goal of providing a structural basis for the development of novel effective therapeutics against tuberculosis (Murillo *et al.*, 2007; Terwilliger *et al.*, 2003; Smith & Sacchettini, 2003; Goulding *et al.*, 2002). Such a systematic approach is possible as the complete genome of the best-characterized virulent strain of *Mtb*, H37Rv, has been sequenced and many of its open reading frames have been annotated (Cole *et al.*, 1998; Camus *et al.*, 2002).

In this study, we report the molecular cloning, protein purification and preliminary X-ray crystallographic analysis of the Rv3340 gene product, which is annotated as a probable *O*-acetylhomoserine (OAH) sulfhydrylase (MetC). *Mtb*MetC is an enzyme that catalyzes the final step in the biosynthesis of methionine. Structurally related enzymes include OAH sulfhydrylases and cystathionine α-, β- and γ-lyases from bacteria, plants, yeast and animals, all of which require a bound pyridoxal 5′-phosphate (PLP) molecule as a cofactor. By sequence alignment, the closest structural neighbor of *Mtb*MetC in the Protein Data Bank (PDB) is OAH2 from *Thermus thermophilus*

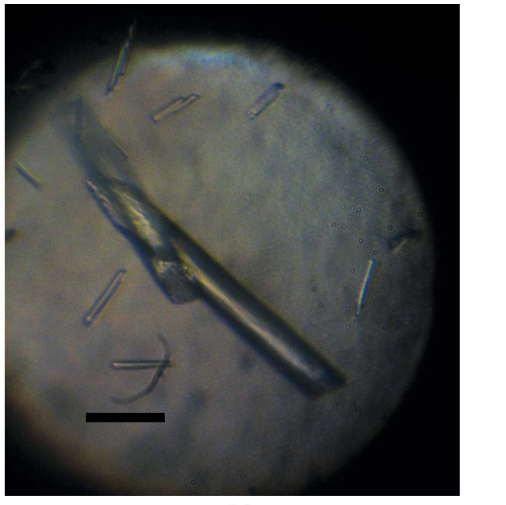
(PDB entry 2ctz; T. Imagawa, Y. Kousumi, H. Tsuge, H. Utsunomiya, A. Ebihara, N. Nakagawa, S. Yokoyama & S. Kuramitsu, unpublished work), which has ∼53% sequence identity to *Mtb*MetC. The crystal structure of *T. thermophilus* OAH2 shows an overall fold composed of three domains: N-terminal, central and C-terminal domains. The active site, where PLP is often covalently attached to a conserved lysine residue, is formed by residues from all three domains plus residues from a neighboring OAH2 molecule in the crystal structure. The functional assembly of these enzymes is thought to be a homotetramer.

2. Experimental methods

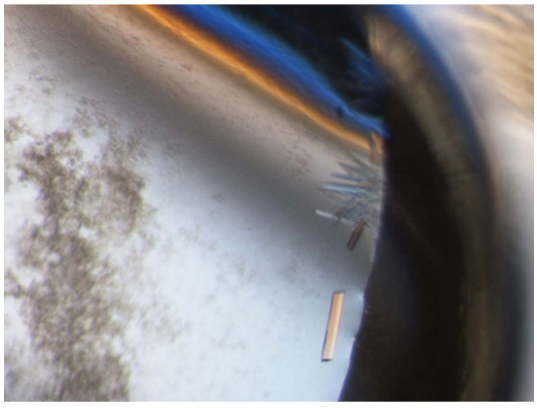
2.1. Cloning, expression and purification

The template DNA for Rv3340 amplification by the polymerase chain reaction (PCR) was from a bacterial artificial chromosome (BAC) genomic library of the *Mtb* H37Rv strain from L'Institut Pasteur (Brosch *et al.*, 1998; Gordon *et al.*, 1999). The PCR primer sequences (Integrated DNA Technologies) for directional cloning of the amplified product into the Gateway cloning system (Invitrogen) were Rv3340F, 5′-gggacaagtttgtacaaaaaagcaggctcc**gaaaacctgtattttca-gggt**ATGAGCGCCGACAGCAATAGCACCG-3′, and Rv3340R, 5′-gggaccactttgtacaagaaagctgggtcTCAGAACGCCGCCACGGACTG-CGG-3′. The bold type in the forward primer encodes the cleavage site (ENLYFQ^G) for tobacco etch virus protease (TEV^{pro}). Complete proteolysis by TEV^{pro} leaves a single glycine residue at the amino-terminus as indicated. The PCR product containing the full open reading frame encoding residues 1–449 of Rv3340 was gel purified and inserted into the pDONR221 plasmid (Invitrogen) to generate an entry clone plasmid, pDONR3340. Subsequently, the coding region of Rv1018c in pDONR1018c was cloned into an expression plasmid pVP16 (Novagen) with an amino-terminal maltose-binding protein (MBP) fusion partner. The coding sequence of Rv3340 in this expression plasmid (pMBP-3340) was confirmed by DNA-sequencing analysis (DNA Core Facility, Department of Biochemistry, University of Alberta, Canada).

The expression strain *Escherichia coli* BL21 (DE3) (Novagen) was transformed to ampicillin resistance by pMBP-3340. An overnight culture from a single colony was used to inoculate 2 l Terrific Broth (TB) supplemented with 100 µg ml^{-1} ampicillin. Shaking of the culture at 310 K was continued until its $OD_{600\,nm}$ reached 0.8. Subsequently, the temperature of the culture was shifted to 295 K and protein overexpression was induced by adding isopropyl β-D-1-thio-galactopyranoside (IPTG) to a final concentration of 0.5 m*M*. After overnight incubation, the cells were harvested by centrifugation at 7000*g* for 15 min. Bacterial pellets were resuspended in 40 ml buffer *A* (20 m*M* NaH_2PO_4/Na_2HPO_4 pH 7.4, 300 m*M* NaCl, 1 m*M* DTT) supplemented with a tablet of Complete protease-inhibitor cocktail (Roche). For protein purification, the cells were lysed by freeze–thaw and then subjected to ultrasonication in buffer *A*. The lysate was cleared by centrifugation (30 min, 20 000*g*) and the supernatant was loaded onto a 5 ml MBPTrap column (GE Healthcare) pre-equilibrated with buffer *A* (20 m*M* Tris pH 7.5, 300 m*M* NaCl, 1 m*M* DTT and 0.02% NaN_3). Following competitive elution of the MBP-Rv3340 fusion protein from the column with 10 m*M* maltose (Sigma) in buffer *A*, the MBP tag was removed by proteolytic cleavage using recombinant tobacco etch virus protease (AcTEV; Invitrogen). The site recognized by the AcTEV protease is encoded in the forward



(*a*)



(*b*)

Figure 1
Morphology of Rv3340 crystals. The approximate dimensions of the Rv3340 crystals in crystal forms I (*a*) and II (*b*) grown over a period of 2–5 d were 20 × 20 × 100 and 100 × 100 × 50 µm, respectively. The black bar in (*a*) indicated a measurement of 50 µm.

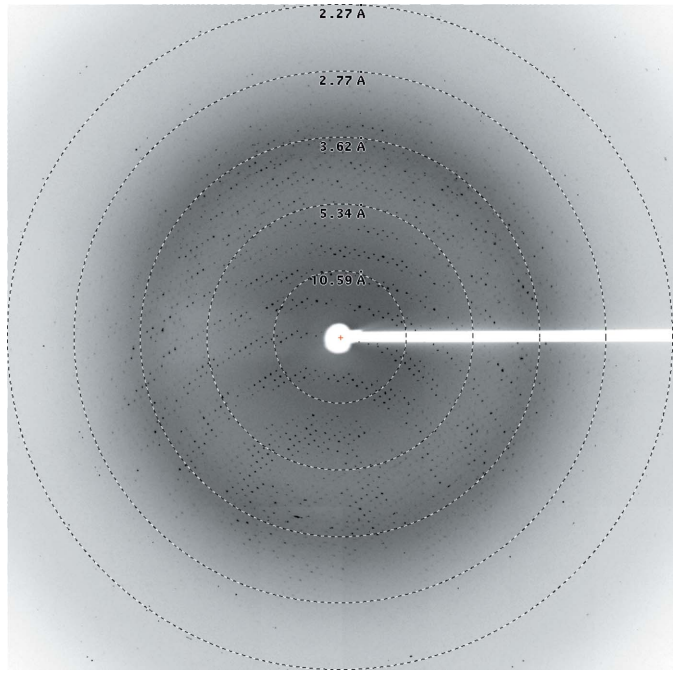


Figure 2
A representative frame showing the diffraction quality of form I Rv3340 crystals. Rings indicate resolution ranges.

primer (shown in bold in the previously mentioned sequence); there is a single glycine residue left in the P1′ position following complete proteolysis. After overnight dialysis against buffer *B* (10 m*M* Tris pH 7.5, 100 m*M* NaCl, 50 m*M* KCl, 0.02% NaN_3), the cleaved protein mixture was loaded onto a HisTrap column (GE Healthcare) and the flowthrough fractions containing liberated Rv3340 were dialyzed against 10 m*M* Tris–HCl pH 7.4 and 100 m*M* NaCl. The resulting solution was concentrated to ∼10 mg ml^{-1} using an Amicon Ultra filtration unit (30 kDa cutoff, Millipore). Most steps of protein purification subsequent to the initial thawing of cells were performed at 277 K and the result of each step was monitored by 15% SDS–PAGE.

2.2. Crystallization

The initial screening of commercially available crystallization conditions was performed on native full-length Rv3340 at 295 K using a robotic setup and the sitting-drop vapor-diffusion technique in 96-well Intelli-Plates (Hampton Research). Crystal Screen, Crystal Screen 2 and Index (Hampton Research) were applied using equal

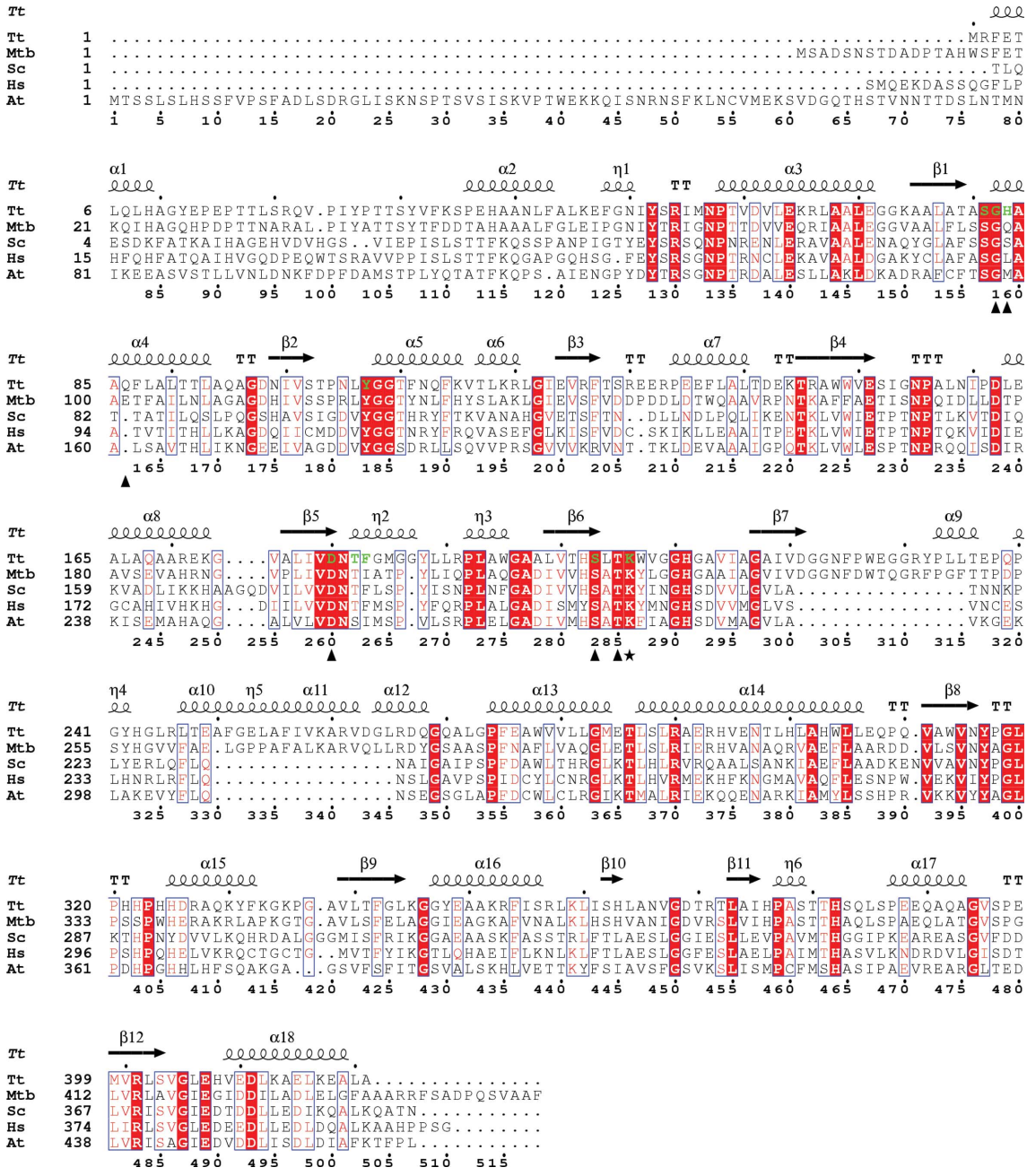


Figure 3
Sequence alignment of *Mtb*MetC (Mtb) with *Thermus thermophilus O*-acetylhomoserine sulfhydrylase (*Tt*OAH2; Tt), *Arabidopsis thaliana* cystathionine *β*-lyase (*At*CBL; At), *Homo sapiens* cystathionine *γ*-lyase (*Hs*CGL; Hs) and *Saccharomyces cerevisiae* cystathionine *γ*-lyase (*Sc*CGL; Sc). The secondary structures corresponding to *Tt*OAH2 (PDB entry 2ctz) are shown at the top of the sequence alignment. Helices, *β*-strands and turns are represented by coils, arrows and 'T's, respectively. Conserved residues are highlighted in red, whereas residues with similar chemical properties are colored red. The identities between *Mtb*MetC and *Tt*OAH2, *Sc*CGL, *Hs*CGL and *At*CBL are 53, 35, 34 and 27%, respectively. Black triangles indicate the residues of *Tt*OAH2 that make hydrogen bonds to a covalently bound cofactor pyridoxal 5′-phosphate (PLP), whereas the key lysine residue that forms a Schiff base with the PLP cofactor is marked by a star. Residues that form van der Waals contacts with PLP are colored green. Sequence alignment was performed using the program *ClustalW* (Thompson *et al.*, 1994) and the figure was generated with the program *ESPript* (Gouet *et al.*, 1999).

Table 1
Crystal parameters and data-collection statistics for native Rv3340 crystals.

Values in parentheses are for the highest resolution shell.

	Crystal form I	Crystal form II
Space group	$P2_1$	$P2_1$
Synchrotron beamline	SSRL BL9-2	ALS 5.0.2
Unit-cell parameters		
a (Å)	84.12	78.20
b (Å)	108.31	123.60
c (Å)	98.28	91.00
β (°)	98.89	113.90
No. of molecules in asymmetric unit	4	4
Matthews coefficient (Å^3 Da^{-1})	2.33	2.12
Solvent content (%)	47.2	42.0
Data collection		
Temperature (K)	100	100
Detector type	MAR 325 CCD	ADSC Quantum 315
Crystal-to-detector distance (mm)	350	200
Oscillation angle (°)	0.5	1
Oscillation range (°)	180	180
Wavelength (Å)	0.98	1.0
Resolution (Å)	50.0–2.5 (2.59–2.50)	20–2.1 (2.14–2.10)
Unique reflections	56684	95061
Mosaicity (°)	0.8	0.3
Multiplicity	3.5 (2.6)	3.5 (3.2)
$\langle I/\sigma(I)\rangle$	11.6 (1.2)	14.5 (2.2)
Completeness (%)	94.0 (67.0)	93.8 (91.3)
$R_{\rm merge}$† (%)	13.7 (63.6)	11.5 (69.1)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the weighted average intensity for all observations i of reflection hkl.

volumes (0.2 µl) of protein solution and reservoir solution. Rv3340 readily crystallizes in several screening conditions, with the largest crystals observed in drops containing Index screen condition No. 59. To obtain crystals suitable for X-ray diffraction, equal volumes (1 µl) of protein solution at ∼10 mg ml^{-1} and reservoir solution [0.02 *M* $MgCl_2$, 0.15 *M* HEPES pH 7.5, 20%(*w*/*v*) polyacrylic acid 5.1K] were mixed and equilibrated in hanging drops against 1 ml reservoir solution on a 24-well VDX plate (Hampton Research). These crystals (form I) grew to maximal dimensions of 100 × 20 × 20 µm in 2–5 d (Fig. 1*a*). Crystals of form II were grown at Lawrence Berkeley National Laboratory using a reservoir solution consisting of 0.1 *M* Tris pH 8.0, 25%(*w*/*v*) PEG 3K, 0.3 *M* calcium acetate in the sitting-drop format (Fig. 1*b*).

2.3. Data collection

Crystals for synchrotron data collection were first rinsed in cryoprotectant (30% glycerol in reservoir solution) and then flash-cooled in liquid nitrogen. Native data sets were collected on beamline BL9-2 at the Stanford Synchrotron Radiation Laboratory (SSRL) and on beamline 5.0.2 at the Advanced Light Source (ALS). Raw data were indexed, integrated and scaled using the *HKL*-2000 program suite (Otwinowski & Minor, 1997).

3. Results

Native data sets (Fig. 2) were collected from form I and II Rv3340 crystals on beamline BL9-2 of the Stanford Synchrotron Radiation Laboratory and beamline 5.0.2 of the Advanced Light Source, respectively. Crystal parameters and crystallographic statistics of the two data sets are summarized in Table 1. Notably, the crystals of forms I and II diffracted to resolutions of 2.5 and 2.1 Å, respectively. According to the calculated mass of Rv3340 including the N-terminal glycine resulting from AcTEV cleavage, ∼47 427 Da, the Matthews coefficient (Matthews, 1968) is calculated to be 2.33 Å^3 Da^{-1} with a solvent content of 47.2% for form I and 2.12 Å^3 Da^{-1} with a solvent content of 42.0% for form II. These results correspond to the presence of four *Mtb*MetC molecules per asymmetric unit. Furthermore, a self-rotation function in crystal form I showed five prominent non-origin peaks at (ω = 37°, φ = ±100°), (ω = 53°, φ = ± 93°) and (ω = 82°, φ = 0°) with χ = 180°. No significant peaks were seen in sections with χ = 120, 90 or 60°. This is in line with a dimer-of-dimers organization in the asymmetric unit. *Mtb*MetC is predicted to be structurally related to the cystathionine lyases. The biological assembly of these thiol lyases from other bacterial species, plants, yeast and animals is believed to be that of a homotetramer. Such a notion has been strongly supported by the X-ray crystallographic findings that the crystals of many of these enzymes often contain one or two tetramers in the asymmetric unit (PDB entries 1pg8, 1gc0, 1gc2, 1ukj, 2o7c, 2nmp, 3cog, 3elp, 1qgn, 1i41, 1i43, 1i48, 1cs1 and 1n8p; Inoue *et al.*, 1995; Motoshima *et al.*, 2000; Kudou *et al.*, 2007; Sun *et al.*, 2009; Steegborn *et al.*, 1999; Wahl *et al.*, 1997; Messerschmidt *et al.*, 2003). While these MetC-related thiol lyases also crystallized in crystal forms with either one (PDB codes 1y4i, 2rfv, 2cb1, 2fq6, 2qgn and 1cl1; Mamaeva *et al.*, 2005; Ejim *et al.*, 2007; Nikulin *et al.*, 2008; Clausen *et al.*, 1996) or two (PDB codes 2ctz, 1e5e, 1e5f, 1ibj, 1pff and 1cl2; Breitinger *et al.*, 2001; Clausen *et al.*, 1996) molecules per asymmetric unit, the crystallographic symmetry always allowed the reconstruction of a tetrameric tertiary assembly *via* the symmetry operations. Based on the significant sequence conservation between *Mtb*MetC and other thiol lyase enzymes (Fig. 3), we posit that *Mtb*MetC also assembles into a crystallographic homotetramer. Structural determination of *Mtb*MetC is presently under way using the molecular-replacement technique as the coordinates of several structural homologues with high sequence identity to *Mtb*MetC have been deposited in the PDB.

X-ray diffraction data for form II Rv3340 crystals were collected on beamline BL9-2 at the Stanford Synchrotron Radiation Laboratory. We would also like to thank the staff at beamline 5.0.2 of the Advanced Light Source at Lawrence Berkeley National Laboratory for help with data collection on form I Rv3340 crystals. Research in the laboratory of MNGJ is funded by the Alberta Heritage Foundation for Medical Research (AHFMR) and the Canadian Institute of Health Research (CIHR). MNGJ held a Canada Research Chair in Protein Structure and Function. JY is grateful for the support of a postdoctoral fellowship from AHFMR.

References

Breitinger, U., Clausen, T., Ehlert, S., Huber, R., Laber, B., Schmidt, F., Pohl, E. & Messerschmidt, A. (2001). *Plant Physiol.* **126**, 631–642.
Brosch, R., Gordon, S. V., Billault, A., Garnier, T., Eiglmeier, K., Soravito, C., Barrell, B. G. & Cole, S. T. (1998). *Infect. Immun.* **66**, 2221–2229.
Camus, J.-C., Pryor, M. J., Médigue, C. & Cole, S. T. (2002). *Microbiology*, **148**, 2967–2973.
Clausen, T., Huber, R., Laber, B., Pohlenz, H. D. & Messerschmidt, A. (1996). *J. Mol. Biol.* **262**, 202–224.
Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.
Ejim, L. J., Blanchard, J. E., Koteva, K. P., Sumerfield, R., Elowe, N. H., Chechetto, J. D., Brown, E. D., Junop, M. S. & Wright, G. D. (2007). *J. Med. Chem.* **50**, 755–764.
Gordon, S. V., Brosch, R., Billault, A., Garnier, T., Eiglmeier, K. & Cole, S. T. (1999). *Mol. Microbiol.* **32**, 643–655.
Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Goulding, C. W. *et al.* (2002). *Curr. Drug Targets*, **2**, 121–141.
Inoue, H., Inagaki, K., Sugimoto, M., Esaki, N., Soda, K. & Tanaka, H. (1995). *J. Biochem.* **117**, 1120–1125.

Kudou, D., Misaki, S., Yamashita, M., Tamura, T., Takakura, T., Yoshioka, T., Yagi, S., Hoffman, R. M., Takimoto, A., Esaki, N. & Inagaki, K. (2007). *J. Biochem.* **141**, 535–544.
Mamaeva, D. V., Morozova, E. A., Nikulin, A. D., Revtovich, S. V., Nikonov, S. V., Garber, M. B. & Demidkina, T. V. (2005). *Acta Cryst.* F**61**, 546–549.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Messerschmidt, A., Worbs, M., Steegborn, C., Wahl, M. C., Huber, R., Laber, B. & Clausen, T. (2003). *Biol. Chem.* **384**, 373–386.
Motoshima, H., Inagaki, K., Kumasaka, T., Furuichi, M., Inoue, H., Tamura, T., Esaki, N., Soda, K., Tanaka, N., Yamamoto, M. & Tanaka, H. (2000). *J. Biochem.* **128**, 349–354.
Murillo, A. C. *et al.* (2007). *Infect. Disord. Drug Targets*, **7**, 127–139.
Nikulin, A., Revtovich, S., Morozova, E., Nevskaya, N., Nikonov, S., Garber, M. & Demidkina, T. (2008). *Acta Cryst.* D**64**, 211–218.
Nunn, P., Williams, B., Floyd, K., Dye, C., Elzinga, G. & Raviglione, M. (2005). *Nature Rev.* **5**, 819–826.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Smith, C. V. & Sacchettini, J. C. (2003). *Curr. Opin. Struct. Biol.* **13**, 658–664.
Steegborn, C., Messerschmidt, A., Laber, B., Streber, W., Huber, R. & Clausen, T. (1999). *J. Mol. Biol.* **290**, 983–996.
Sun, Q., Collins, R., Huang, S., Holmberg-Schiavone, L., Anand, G. S., Tan, C.-H., van-den-Berg, S., Deng, L.-W., Moore, P. K., Karlberg, T. & Sivaraman, J. (2009). *J. Biol. Chem.* **284**, 3076–3085.
Terwilliger, T. C. *et al.* (2003). *Tuberculosis*, **83**, 223–249.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Wahl, M. C., Huber, R., Prade, L., Marinkovic, S., Messerschmidt, A. & Clausen, T. (1997). *FEBS Lett.* **414**, 492–496.