Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structural and functional analysis of Rv0554 from *Mycobacterium tuberculosis*: testing a putative role in menaquinone biosynthesis

Jodie M. Johnston,[a]* Ming Jiang,[b] Zhihong Guo[b] and Edward N. Baker[a]

[a]School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand, and [b]Department of Chemistry, Center for Cancer Research, The Hong Kong University of Science and Technology (HKUST), Kowloon, Hong Kong SAR, People's Republic of China

Correspondence e-mail: jm.johnston@auckland.ac.nz

Received 1 June 2010
Accepted 30 June 2010

PDB References: Rv0554, 3e3a; 3hss.

Mycobacterium tuberculosis, the cause of tuberculosis, is a devastating human pathogen against which new drugs are urgently needed. Enzymes from the biosynthetic pathway for menaquinone are considered to be valid drug targets. The protein encoded by the open reading frame *Rv0554* has been expressed, purified and subjected to structural and functional analysis to test for a putative role in menaquinone biosynthesis. The crystal structure of Rv0554 has been solved and refined in two different space groups at 2.35 and 1.9 Å resolution. The protein is dimeric, with an α/β-hydrolase monomer fold. In each monomer, a large cavity adjacent to the catalytic triad is enclosed by a helical lid. Dimerization is mediated by the lid regions. Small-molecule additives used in crystallization bind in the active site, but no binding of ligands related to menaquinone biosynthesis could be detected and functional assays failed to support possible roles in menaquinone biosynthesis.

1. Introduction

Mycobacterium tuberculosis (*Mtb*), the causative agent of tuberculosis (TB), is a devastating human pathogen that is responsible for nearly two million deaths annually. Two features of the physiology of *Mtb* that make it particularly difficult to treat are the rise in drug resistance and the ability of the organism to remain in a nonreplicating persistent (NRP) state after phagocytosis by macrophages. As a result, about one-third of the world's population is estimated to have latent or persistent TB infection, with ~5–10% of people with latent infection developing active TB in their lifetime. TB thus remains a significant global health problem.

Menaquinone is an important component of the mycobacterial electron-transport chain and is believed to play a role in both aerobic and anaerobic respiration by mycobacteria. The menaquinone-biosynthesis pathway has thus been suggested as a good target for the development of new antimycobacterial drugs (Kurosu & Begari, 2010; Dhiman *et al.*, 2009; Kurosu *et al.*, 2007; Mdluli & Spigelman, 2006; Anishetty *et al.*, 2005; Lu *et al.*, 2008). A number of the components of the pathway have been found to be essential genes (Sassetti *et al.*, 2003) and recent work with inhibitors of the enzyme MenA has shown inhibition of aerobic growth of the bacteria as well as activity against *Mtb* in its NRP state (Dhiman *et al.*, 2009). Isolation of a novel sulfated derivative of the primary menaquinone from *Mtb* that appears to be associated with virulence (Holsclaw *et al.*, 2008) further supports the significance of studying this biosynthetic pathway.

In many organisms, such as *Escherichia coli* and *Bacillus subtilis*, the menaquinone-biosynthesis genes are co-localized in a similar order and are likely to be co-transcribed (Glasner *et al.*, 2006). Similar clusters are found for Actinobacteria such as *Mtb*, although other genes are interspersed (Glasner *et al.*, 2006). In the *Mtb* genome most of the recognized enzymes of the menaquinone-biosynthetic pathway (MenA, MenB, MenC, MenD, MenE, MenF and UbiE) lie between *Rv0534c* and *Rv0558*. However, recent studies in *E. coli* have implicated a further gene *yfbB* in menaquinone biosynthesis and re-annotated it as *menH*. Conflicting views exist as to the reaction catalysed by the MenH enzyme. MenH has been proposed to act either as a thioesterase that cleaves the CoA moiety from 1,4-dihydroxynaphthoate (DHNA) following the action of MenB (Meganathan, 2001) or to act between MenD and MenC as a (1*R*,6*R*)-2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate (SHCHC) synthase (Jiang *et al.*, 2008; Fig. 1). Structurally, YfbB/MenH has been predicted to be an α/β-hydrolase family member on the basis of its 40% sequence identity to an enzyme from *Vibrio cholerae* (PDB code 1r3d; J. Gorman & L. Shapiro, unpublished work) and genes for similar putative MenH α/β-hydrolases have been found to be clustered in the same location as other menaquinone-biosynthesis genes (Glasner *et al.*, 2006).

Examination of the *Mtb* genome suggested that the open reading frame *Rv0554* could encode an enzyme with a function analogous to YfbB/MenH. Rv0554 from *Mtb* has been shown to cause slow growth of the bacterium in transposon mutagenesis studies (Sassetti *et al.*, 2003), it is located within the *men* gene cluster between *menC* and *menD*, and it is predicted to be in an operon with Rv0552 and MenC (Rv0553) regulated by the protein Rv1719 (Krawczyk *et al.*, 2009). It is also relevant that the degenerate *M. leprae* genome, with its many pseudogenes, still has an intact Rv0554 homologue that is located similarly between *menC* and *menD*. Expression of this homologue has been experimentally validated, with the protein being detected in the cell wall and the cytosol of *M. leprae* (Marques *et al.*, 2008).

Rv0554 has only ∼20% sequence identity to *E. coli* MenH, but sequence comparisons show that MenH sequences are highly variable, with only 15 residues being fully conserved across 47 species (Jiang *et al.*, 2009). Importantly, *Mtb* Rv0554 is predicted to have the same α/β-hydrolase fold as *E. coli* MenH and appears to be the only α/β-hydrolase located in close proximity to the *Mtb* menaquinone-biosynthesis genes. Although it is annotated as a possible cofactor-free bromoperoxidase BpoC, this reflects a slightly higher (∼30%) sequence identity with several haloperoxidases. Given the wide functional diversity of α/β-hydrolase enzymes, an alternative function as MenH is also possible.

In addition to its putative role in menaquinone biosynthesis, Rv0554 has been identified as a favoured *Mtb* drug target in a recent *in silico* target-identification analysis (Raman *et al.*, 2008) and has been shown by microarray data to be upregulated in the hypoxic and mildly acidic conditions (Kim *et al.*, 2008) thought to be linked to persistence. Here, we present the crystal structure of Rv0554, with the aim of characterizing it as a potential drug target, together with functional studies intended to establish whether this protein is (or is not) an authentic MenH enzyme.

2. Methods

2.1. Cloning, expression and purification

Rv0554 was amplified by PCR, cloned into a pProEX Hta vector containing an N-terminal cleavable His_6 tag and transformed into rare-codon *E. coli* BL21 pRIL cells. Overnight cultures were grown in MDG media (Studier, 2005) or LB (Luria broth) with chloramphenicol/ampicillin selection (34 and 100 µg ml^{-1}, respectively) and used to inoculate 1 l ZYM-5052 autoinduction media (Studier, 2005) or LB. The cultures were grown at 310 K for several hours and then transferred to 297 K for overnight expression using either

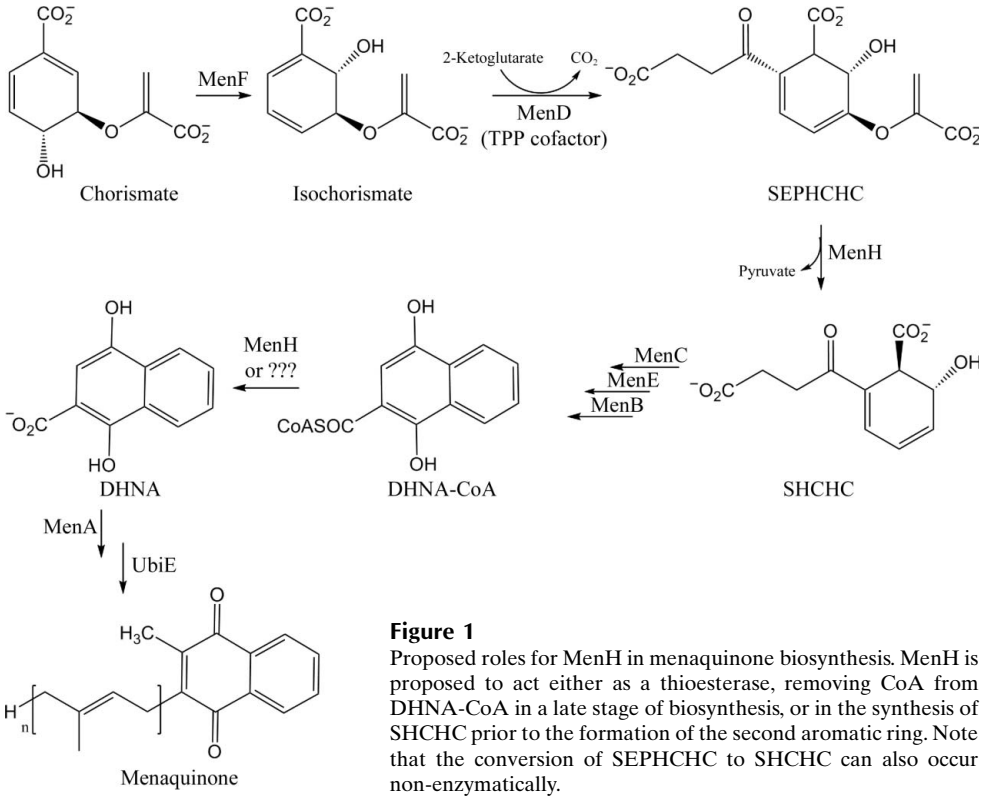

Figure 1
Proposed roles for MenH in menaquinone biosynthesis. MenH is proposed to act either as a thioesterase, removing CoA from DHNA-CoA in a late stage of biosynthesis, or in the synthesis of SHCHC prior to the formation of the second aromatic ring. Note that the conversion of SEPHCHC to SHCHC can also occur non-enzymatically.

Table 1
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

	Orthorhombic form	Tetragonal form
Space group	$P2_12_12_1$	$P4_12_12$
Unit-cell parameters (Å, °)	a = 48.70, b = 97.89, c = 146.17, $\alpha = \beta = \gamma = 90$	$a = b$ = 100.45, c = 135.16, $\alpha = \beta = \gamma = 90$
Molecules per asymmetric unit	2	2
Solvent content (%)	54.3	53.4
Resolution range (Å)	50.0–2.35 (2.43–2.35)	50.0–1.90 (1.97–1.90)
$R_{\rm merge}$ (%)	11.4 (74.8)	8.7 (67.8)
No. of reflections	201036	1593886
No. unique reflections	29742	54772
Mean $I/\sigma(I)$	19.7 (2.0)	53.0 (6.1)
Multiplicity	6.8	29.1
Completeness (%)	99.5 (96.9)	99.0 (98.2)

IPTG induction (0.8 m*M* final concentration when OD_{600} reached 0.6–1.0) or autoinduction (Studier, 2005).

Bacterial cell pellets containing overexpressed His_6-tagged Rv0554 were resuspended in 20–30 ml lysis buffer (50 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl, 20 m*M* imidazole) with a complete EDTA-free protease-inhibitor tablet, lysozyme, DNase and RNase added and then lysed by sonication or cell disruption (Constant Cell Disruption Systems), followed by pelleting of the insoluble material by centrifugation (20 000*g* for 20 min). The supernatant containing soluble His_6-tagged Rv0554 was then loaded onto a 5 ml Ni^{2+} HiTrap affinity column pre-equilibrated in lysis buffer, washed with 6–10 column volumes of lysis buffer and eluted with a gradient to 100% elution buffer (50 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl, 500 m*M* imidazole). The protein eluted between ∼150 and 300 m*M* imidazole. His_6-tagged Rv0554 was then further purified by size-exclusion chromatography using a Superdex 200 10/30 or 16/60 column in 50 m*M* Tris–HCl pH 8, 150 m*M* NaCl. The protein eluted at a size consistent with a dimer. Protein samples were analysed by dynamic light scattering (DynaPro, Protein Solutions), SDS–PAGE and native PAGE.

2.1.1. Purification for DHNA-CoA thioesterase and SHCHC synthase activities. The plasmid pProEXHta Rv0554 was transformed into BL21 (DE3) cells and the resulting transformants were plated onto an LB–agar plate containing ampicillin at 100 µg ml^{-1}. A single colony from the plate was used to inoculate 10 ml LB medium (100 µg ml^{-1} ampicillin) overnight at 310 K. Subsequently, the overnight LB culture was used to inoculate 1 l LB medium containing ampicillin. The culture was grown for about 3 h at 310 K to an OD_{600} of ∼0.8 and then cooled to 289 K. Finally, IPTG was added to the cooled culture to a final concentration of 0.25 m*M* to express the enzyme at 289 K for 24 h before harvesting.

Cell pellets harvested from 1 l bacterial culture were first suspended in 20–30 ml Ni^{2+}–NTA affinity-column buffer (20 m*M* Tris–HCl buffer pH 7.9, 5 m*M* imidazole, 0.5 *M* NaCl) containing 1 m*M* PMSF and the cells were lysed by sonication. The cell lysate was centrifuged at 30 000*g* for 30 min and the supernatant was loaded onto an Ni^{2+}–NTA column. The His_6-tagged Rv0554 was eluted from the Ni^{2+}–NTA column following the manufacturer's instructions (Novagen) and was further purified using a Sephacryl-200 gel-filtration column (GE Healthcare; column buffer 50 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl, 10% glycerol).

2.2. Crystallization

Initial crystallization conditions were found by sitting-drop vapour diffusion (100 nl protein solution + 100 nl precipitant) using a Cartesian Honeybee dispensing robot with an in-house 480-component screen (Moreland *et al.*, 2005). Subsequent optimization of an initial MPD (2-methyl-2,4-pentanediol) condition also utilized sitting-drop vapour diffusion, with a range of drop sizes from 1 µl upwards. Subsequent screens were also undertaken with a variety of additives, including Tacsimate pH 3.8, 7.0 and 8.6, malonic acid and ethylene glycol.

All crystals were grown from protein at a concentration of 7–8 mg ml^{-1}. Orthorhombic crystals which belonged to space group $P2_12_12_1$ were grown from 0.1 *M* sodium acetate pH 4.9, 7.5% MPD. The best diffracting (<2.6 Å) tetragonal crystals, belonging to space group $P4_12_12$, were grown from 0.1 *M* sodium acetate pH 4.92, 2–5% MPD, 10–15% ethylene glycol with either 5% Tacsimate pH 8.6 or 1.65 *M* unbuffered malonic acid added (Table 1).

Cocrystallization experiments were also carried out using a range of compounds similar to potential substrates or products of reactions thought to be possible for Rv0554. These compounds were mixed with the protein at a range of concentrations prior to crystallization or soaked into crystals prior to flash-freezing. These compounds included DHNA (1,4-dihydroxy-2-naphthoic acid, the product of the MenH thioesterase reaction), 2-acetylbenzoic acid, thymine, PMSF, malonyl-CoA, acetyl-CoA and the potassium salts of phosphoenoyl pyruvate (PEP) and SHCHC (the product of the other MenH-catalysed reaction). The best crystals obtained, which diffracted to 1.9 Å resolution, were obtained after the addition of PEP to the protein prior to crystallization, although no PEP was ever observed in the crystal structure.

2.3. Data collection and processing

For data collection, crystals were soaked in cryoprotectant and then flash-frozen in liquid N_2. For the orthorhombic ($P2_12_12_1$) crystals, mother liquor supplemented with 20% MPD was used as cryoprotectant, whereas for all other crystals mother liquor supplemented with 20–25% ethylene glycol was used. The $P2_12_12_1$ data set was collected to 2.35 Å resolution on beamline BL9-1 (λ = 0.97907 Å) at the Stanford Synchrotron Radiation Laboratory with an ADSC Quantum 315 CCD detector. For the $P4_12_12$ crystals, a data set was collected to 1.9 Å resolution on beamline 3-BM1 (λ = 0.95665 Å) at the Australian synchrotron equipped with an ADSC Quantum 315R CCD detector. For all data sets the raw data were processed using *DENZO* in *HKL*-2000 and scaled using *SCALEPACK* in *HKL*-2000 (Otwinowski & Minor, 1997). Full details of the best orthorhombic and tetragonal data sets are given in Table 1.

2.4. Structure determination and refinement

The $P2_12_12_1$ crystal structure was solved by molecular replacement using *Phaser* (McCoy *et al.*, 2007) with a search model comprising the monomer of bromoperoxidase A2 (Hecht *et al.*, 1994), which had 29% sequence identity to Rv0554. The Rv0554 sequence was docked into this model with *MODELLER* (Eswar *et al.*, 2006). Iterative model rebuilding was carried out with *Coot* (Emsley & Cowtan, 2004) into σ_A-weighted $2mF_o - DF_c$ and $mF_o - DF_c$ electron-density maps at 2.35 Å resolution, interspersed with cycles of refinement with *CNS* (Brünger *et al.*, 1998). Maps were also generated with *ARP/wARP* (Perrakis *et al.*, 1997) and *Prime&Switch* (Terwilliger, 2004) during the manual building process. The refinement was completed with *REFMAC*5 (Murshudov *et al.*, 1997; Winn *et al.*, 2001) from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) with TLS refinement. The $P4_12_12$ structures were solved by molecular replacement in *MOLREP* (Vagin & Teplyakov, 1997) or *Phaser* using a monomer from the $P2_12_12_1$ Rv0554 structure as a search model. Iterative rounds of model building and refinement were carried out in *Coot* and *REFMAC*5 or *PHENIX* (Adams *et al.*, 2010), respectively. Full refinement and model details are given in Table 2.

2.5. Functional assays

2.5.1. Generic thioesterase activity. Prior to more specific testing, purified His_6-tagged Rv0554 was assayed for generic thioesterase activity using the substrates malonyl-CoA (MCoA) and palmitoyl-CoA (PCoA). A range of 1 ml reactions were set up in 50 m*M* Tris–HCl pH 8, 150 m*M* NaCl buffer containing 0.3 m*M* 5,5′-dithiobis-(2-nitrobenzoic acid) (DTNB), 50 or 22.5 µg His_6-tagged Rv0554 and 0.04 m*M* MCoA or PCoA and BSA to make the final protein concentration in the reaction 0.3 mg ml^{-1}. Reactions were started by addition of enzyme and incubated at 310 K. Samples were taken over the course of 1.5 h and measured for absorbance at 412 nm. Controls with substrates but no enzyme and with enzyme only were also carried out.

2.5.2. SHCHC synthase activity assay. Purified His_6-tagged Rv0554 was tested for SHCHC synthase activity using the method described by Jiang *et al.* (2008). This assay was carried out in 50 m*M* sodium phosphate buffer pH 7.0 containing 0.1% bovine serum albumin (BSA), 15 µ*M* SEPHCHC and the Rv0554 protein at an appropriate concentration. The reaction was monitored in real time by observing the change in absorbance at 290 nm.

2.5.3. DHNA-CoA thioesterase activity assay. Purified His_6-tagged Rv0554 was assayed for DHNA-COA thioesterase activity using a previously reported method (Jiang *et al.*, 2008). The reactions were carried out in triplicate in 50 m*M* phosphate buffer pH 7.0 containing 10 m*M* $MgCl_2$, 1 m*M* DTT and 15.0 µg ml^{-1} Rv0554, 15.0 µg ml^{-1} BSA (as a negative control) or 15.0 µg ml^{-1} *E. coli* thioesterase EntH (as a positive control). Thioesterase activity towards DHNA-CoA was determined by real-time monitoring of the decrease in absorbance at 392 nm.

Table 2
Refinement and model details.

	Orthorhombic form	Tetragonal form
PDB code	3e3a	3hss
No. of reflections	26997	54732
Resolution range (Å)	50.0–2.35	50.0–1.90
R (%) (No. of reflections)	17.9 (24294)	19.3 (51947)
R_{free} (%) (No. of reflections)	23.8 (2703)	21.9 (2782)
R.m.s.d. from standard values		
Bond lengths (Å)	0.018	0.015
Bond angles (°)	1.7	1.4
Ramachandran plot		
Residues in most favoured regions (%)	97.8	98.9
Outliers	0	0
Model details		
Protein atoms	4213	4130
Water molecules	214	516
Small molecules/ions	2 acetate ions, 1 MPD	3 malonate ions, 6 ethylene glycols, 2 Na^+ ions, 1 Tris, 1 acetate ion
Mean B factors (Å^2)		
Protein (all atoms)	48.4	23.3
Waters	49.9	37.5
Malonic acid (MLA)	—	48.3
Tris (TRS)	—	39.5
Sodium	—	32.15
Ethylene glycol (EDO)	—	51.13
Acetate (ACT)	74	57.2
2-Methyl-2,4-pentanediol (MPD)	65.9	—

3. Results

3.1. Improvement of crystal quality

The initial orthorhombic crystals of Rv0554 were of variable quality and the $P2_12_12_1$ structure was based on the only two crystals that diffracted to relatively high resolution (2.35 Å). Poor-quality tetragonal crystals, with anisotropic diffraction to ∼4 Å resolution, grew alongside the orthorhombic crystals. Additive screening (McPherson & Cudney, 2006) proved crucial in obtaining high-quality reproducible crystals. The use of ethylene glycol as an additive improved the reproducibility of the orthorhombic crystals. Further additive screens gave vast improvements in the diffraction quality and reproducibility of the tetragonal crystal form, firstly when Tacsimate pH 8.6 was included as an additive and then with malonic acid as additive; the crystal structure of the Tacsimate-containing crystals was found to have malonic acid bound in several positions, including the active site, and this was identified as the key additive for ensuring reproducibility and high-quality diffraction.

3.2. Structure of Rv0554

We determined crystal structures of Rv0554 from the orthorhombic crystals and from tetragonal crystals obtained using Tacsimate as an additive and using two different preparations with malonic acid. Apart from some variation in the small molecules bound, one loop rearrangement and differences in the modelling of the His_6 tag, the monomer structures are very similar in all of the crystal structures determined, with root-mean-square (r.m.s.) differences of

0.09– 0.62 Å in C^{α}-atom positions over the entire structure. We present here the highest resolution orthorhombic and tetragonal crystal structures. In each case unbroken density is seen for at least residues 2–261 of the 262-residue protein. In addition, 12–14 residues of the N-terminal His_6 tag could be modelled for both chains of the orthorhombic structure and chain *A* of the tetragonal structure. The tetragonal structure also includes three malonate ions, six ethylene glycol molecules, two sodium ions, one acetate ion and one Tris molecule.

3.2.1. Monomer fold. The Rv0554 monomer (Fig. 2) has the classic α/β-hydrolase fold, comprising a core β-sheet of mostly parallel strands surrounded on both sides by helices (Ollis *et al.*, 1992; Holmquist, 2000; Heikinheimo *et al.*, 1999). The Rv0554 fold represents a minimal version as it contains only seven β-strands compared with the usual eight and lacks the extensive elaboration that is often found. It does, however, have a helical lid domain made up of five α-helices (H5–H9) that constitute the middle third of the protein (residues ∼114–192). This helical lid domain is common in other α/β-hydro-

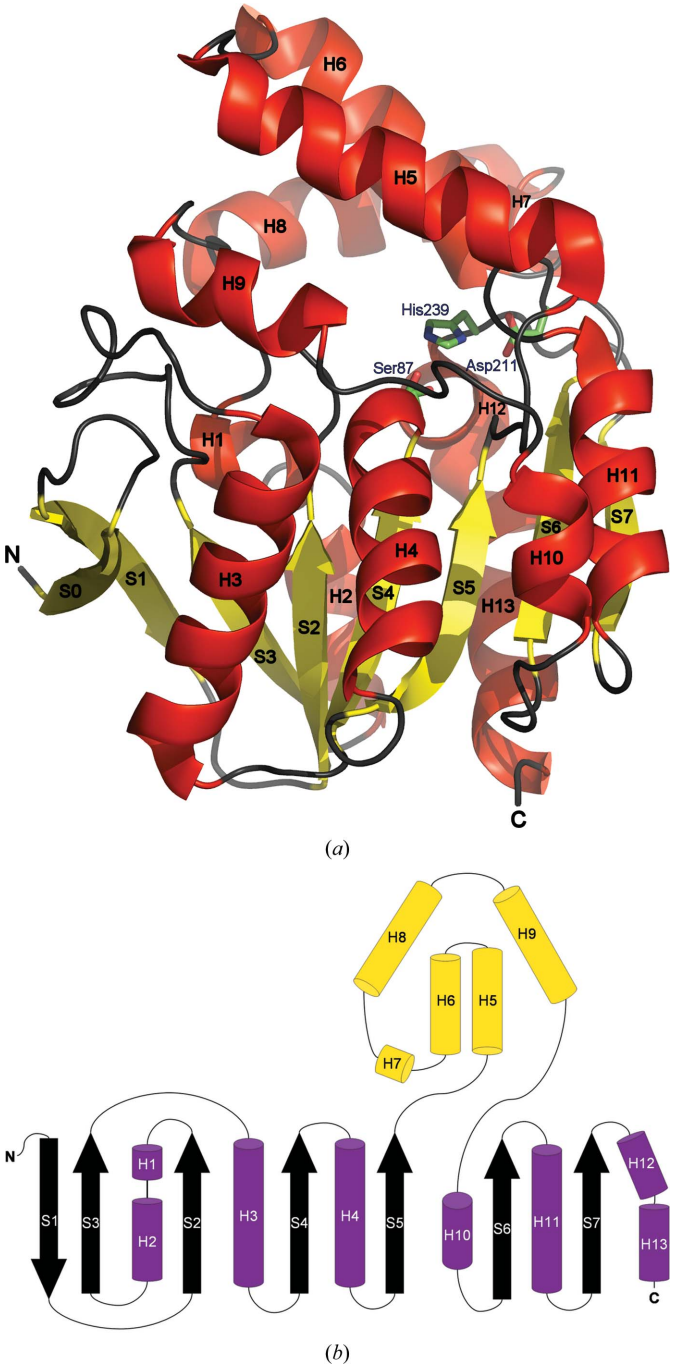

Figure 2
Monomer fold of Rv0554. (*a*) Ribbon diagram showing α-helices (red) and β-strands (yellow). The lid domain is shown at the top, above the residues of the catalytic triad, which are shown in stick mode (green; note that the serine has two alternate conformations). Strand S0 comprises residues from the His_6 tag and would not be present in the native enzyme. (*b*) Topology diagram. The core-domain α-helices are shown in magenta and the β-strands are shown in black. The five α-helices comprising the lid domain are shown in yellow.

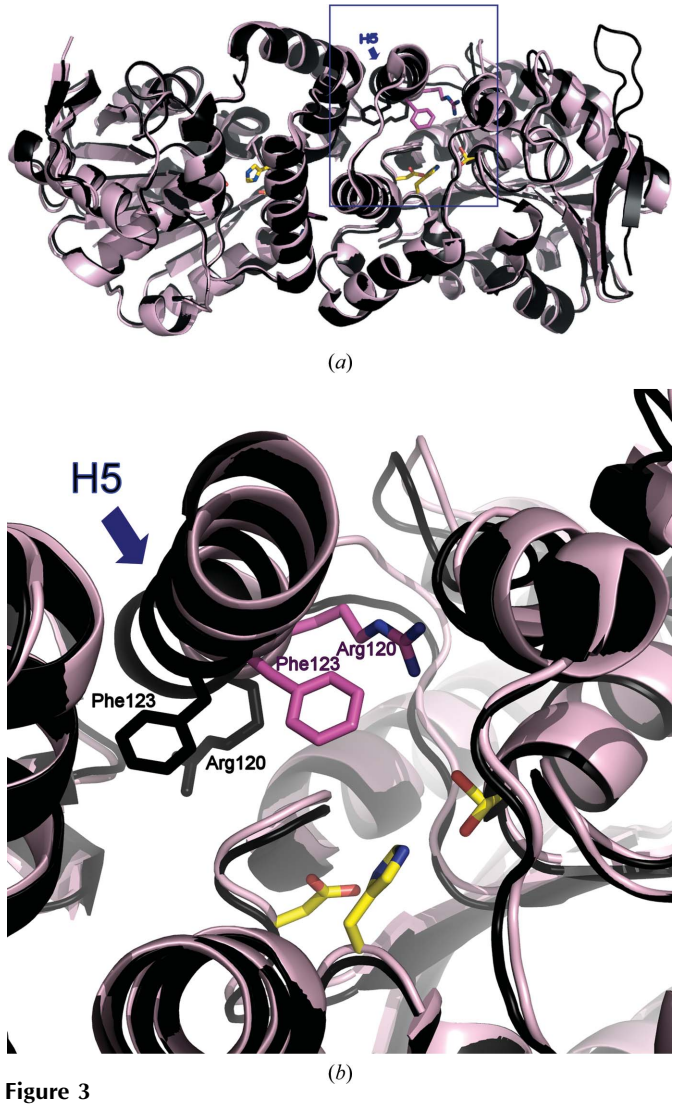

Figure 3
Active-site loop rearrangement in Rv0554. Molecule *A* of the $P2_12_12_1$ structure (black) is superimposed on molecule *B* of the $P4_12_12$ structure (pink), illustrating the change in the position of helix 5 and the preceding loop, involving residues 115–125. Arg120 and Phe123 are shown as sticks (black for $P2_12_12_1$; magenta for $P4_12_12$), as are the active-site catalytic triad residues for the $P4_12_12$ structure (yellow; note that the serine has two alternate conformations). For clarity the malonic acid is not drawn in these figures. (*a*) Overall view. (*b*) Close-up view of the area in the blue box in (*a*).

lases although its components are variable (Heikinheimo *et al.*, 1999; Holmquist, 2000). In Rv0554 it encloses a large cavity above the active site and also participates in dimerization, as described below. The residues of the His_6 tag that are visible in three of the four monomers form an additional β-strand (S0) that extends the core β-sheet, running antiparallel to strand S1. Other α/β-hydrolases have similar extensions, which seem to be well tolerated by the fold. In the case of Rv0554 this may explain why the cleavage of the His_6 tag by rTEV protease was poor.

3.2.2. Active site. The active-site catalytic triad residues Ser87, His239 and Asp211 are present on loops of the core α/β domain at the characteristic positions found in other α/β-hydrolases (Ollis *et al.*, 1992). The catalytic Ser87 is located on the $\beta4$–$\alpha5$ loop, which has the characteristic 'nucleophile elbow' configuration with unusual Ramachandran (φ, ψ) angles for Ser87. In both chains of the $P4_12_12$ structure the density shows clear evidence of two alternate conformations for the Ser87 side chain. The catalytic His239 is located on the S7–H12/13 loop and Asp211 is located on the S6–H11 loop. The 'oxyanion hole' that is characteristic of most α/β-hydrolase enzymes (Holmquist, 2000) is formed by the main-chain NH groups of Arg21 and Met88 in Rv0554.

The lid domain caps the active site and encloses a large cavity that has a volume of 2332 Å^3 for both molecules of the tetragonal structure, as determined by analysis with *CASTp* (Dundas *et al.*, 2006). This cavity has a similar volume, 2391 Å^3, in molecule *A* of the orthorhombic structure and is only slightly smaller in molecule *B*, with a volume of 2090 Å^3. It comprises one main 'active-site' cavity adjacent to the active-site residues, together with a 'side pocket' that leads off from the active site and is lined by mostly hydrophobic residues from the lid domain. The residues involved in forming this potential substrate-binding region are extensive. A large number of residues from the lid domain (residues 113–188) are involved, as are residues from the 'catalytic' region around Ser87, Asp211 and His239 and the oxyanion hole.

Several differences between monomers can be seen in the active site. Firstly, in the tetragonal structure there is a rearrangement of residues 113–125 (from the start of H5 and the end of the loop preceding it) close to the *B*-chain active site, with an especially large shift in the positions of Arg120 and Phe123 (Fig. 3). This rearrangement, which is not observed in the ligand-free orthorhombic structure, allows Arg120 to hydrogen bond to the malonic acid molecule that is bound in the active site of chain *B* (Fig. 4). It also results in a change in the hydrogen-bonding interactions of the catalytic triad residue Asp211.

Secondly, in the tetragonal structure there are adventitious ligands bound in the catalytic region of the active site, which differ between monomers: a Tris molecule in monomer *A* and a malonic acid molecule in monomer *B* (Fig. 4). These ligands form hydrogen bonds to some of the active-site residues. The malonic acid molecule in the *B*-chain site hydrogen bonds to the peptide NH groups of the oxyanion-hole residues Arg21 and Met88, the side chains of the catalytic residues Ser87 and His239, and Arg120 from the rearranged loop described above (Fig. 4). The Tris molecule seen in the monomer *A* chain also hydrogen bonds to the peptide NH of Met88, as well the main-chain O atoms of Leu184 and Gln183.

The 'side' pocket is often occupied by small molecules in the Rv0554 structures: MPD in the orthorhombic structure and either malonic acid or ethylene glycol in the tetragonal structures (Fig. 4). These small molecules all bind in virtually the same place, bounded by Thr27, Arg146, Phe151, Trp165, Met168, Phe169, Trp172, Ile174, Leu240, Phe242 and Phe243, packing against Trp172 and hydrogen bonding to the side chain of Arg146. In all structures, the narrow channel joining this 'side pocket' to the active site is occupied by solvent molecules, despite the number of hydrophobic residues. In molecule *A* of the tetragonal structure, for example, two ethylene glycol molecules and a number of waters form a chain from the Tris molecule bound at the active site to the malonic acid bound in the 'side pocket'.

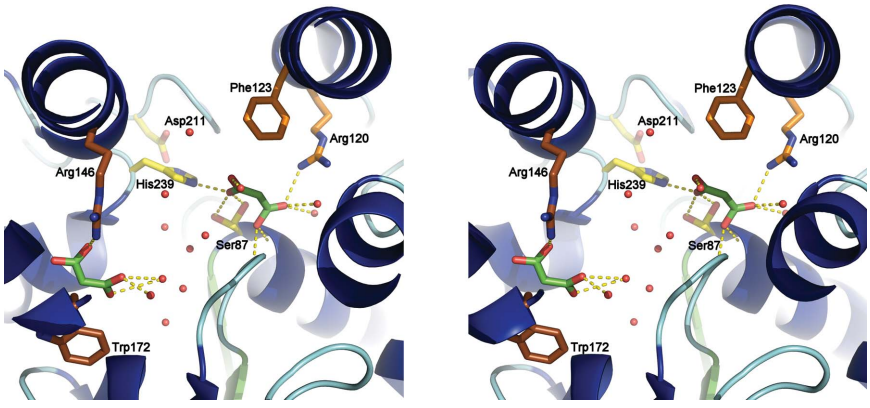

Figure 4
Stereo diagram showing both malonic acids (green, in stick mode) bound in the molecule *B* $P4_12_12$ active site and side pocket (shown in blue/green ribbon representation). The active-site residues are also shown in stick mode (yellow), with Ser87 having two alternate conformations. Hydrogen bonds between the malonic acid and the protein are shown as dashed yellow lines. Arg120 and Phe123, which are rearranged in the *B*-molecule active site, are shown in stick mode (orange), with Arg120 making hydrogen-bonding interactions with the active-site malonic acid. Trp172 and Arg146 are also shown in stick mode (brown); Arg146 interacts with the second malonic acid (in the side pocket), which packs against Trp172. The water molecules filling the cavity between these two ligands are shown as red spheres.

3.2.3. The Rv0554 dimer. Dynamic light-scattering (DLS) analysis (Wilson, 2003) suggested that Rv0554 is dimeric in solution, giving an estimated molecular weight of 59–66 kDa compared with the monomer value of ∼31 kDa. Elution from size-exclusion columns was also consistent with a dimer in solution (data not shown). Analysis of the crystal structures using *PISA* (Krissinel & Henrick, 2007) also suggested that the protein is likely to form a stable dimer in solution.

The Rv0554 dimer is the same in both crystal forms and is formed primarily by packing of the helical lids of the two mole-

cules against each other and against a small part of the core domain (Fig. 5). This interface buries ∼13% of the solvent-accessible surface area upon dimer formation (1741 Å^2 for the orthorhombic structure and 1575 Å^2 for the tetragonal structure). The residues involved in this dimer interface comprise 115–166 from the helical lid and 209–217 from the core domain. The resulting dimer could be described as a lid-to-lid dimer.

Importantly, this dimer structure means that some residues in the active-site regions are also involved in the dimer interface. This is clearly seen in the tetragonal structure, where the movement of residues 113–125 in monomer *B* removes Arg120 from the dimer interface (where it is located in monomer *A* and in the orthorhombic structure) to take up a position in the active site, where it binds to the malonic acid molecule. This could provide a mechanism for feedback between the two active sites of the dimer. Residues 141 and 144 from the *A* chain also contribute to the *B*-chain active-site pocket, while residues 141, 144 and 148 from the *B* chain contribute to the *A*-chain active-site pocket, further emphasizing the intimate linkage between dimerization and active-site structure. Several other small pockets are found in the dimer interface. One of these, formed by residues 115–122 and 209–218 of chain *A* and 144–159 of chain *B*, contains a bound ethylene glycol molecule in the $P4_12_12$ dimer (Fig. 5). Among the residues that contribute to the lining of this pocket is the catalytic Asp211, again emphasizing the relation between the active site and dimer formation.

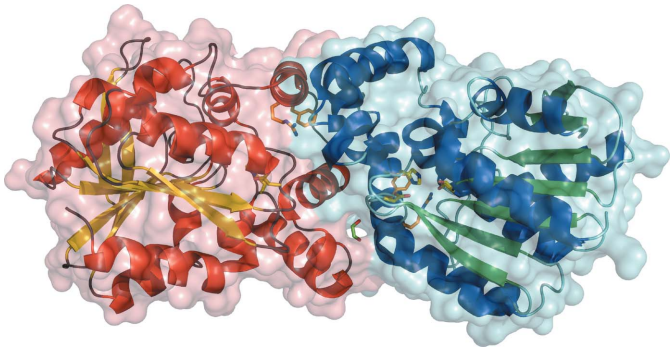

Figure 5
The Rv0554 dimer. Surface diagram for the $P4_12_12$ dimer, with ribbon diagram overlaid showing how the lid domains of the two monomers of the $P4_12_12$ dimer pack against one another (the *A* chain is shown in yellow/red and the *B* chain in blue/green). Active-site residues are shown in stick mode (yellow) with Ser87 in two alternate conformations. Arg120 and Phe123 which are rearranged in the *B* chain are also shown in stick mode (orange). Note the ethylene glycol molecule (green) in a pocket in the dimer interface. This pocket does not exist in the $P2_12_12$ dimer.

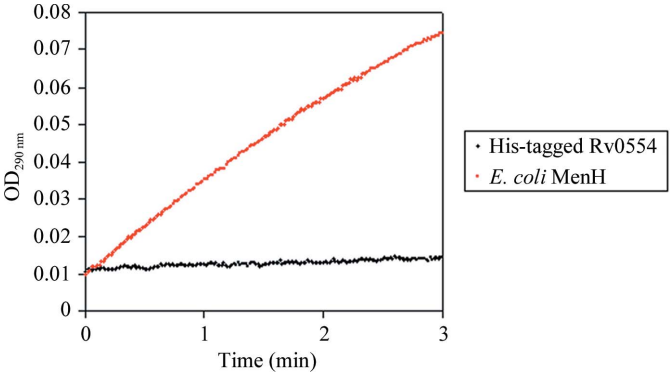

Figure 6
SHCHC synthase activity assay. The two traces show real-time monitoring of SHCHC synthase activity for His_6-Rv0554 compared with *E. coli* MenH (positive control).

3.2.4. Comparisons with homologous structures. Searches for structural homologues with *SSM* (Krissinel & Henrick, 2004) or *DALI* (Holm *et al.*, 2006, 2008) emphasize the ubiquity of the α/β-hydrolase fold and the large number of different functions that it is able to support. The best structural matches to Rv0554, with r.m.s. differences in C^α positions in the range 1.6–1.8 Å (over 226–242 residues), are to a range of α/β-hydrolases including aryl esterases, nonhaem bromo/chloroperoxidases, γ-lactamases and a range of C—C-bond hydrolases. There are also many more matches with r.m.s. differences between 1.8 and 2.5 Å (over 207–251 residues).

Potentially more instructive is a search for homologues that also share the protein interface used in the Rv0554 dimer. A search with *PISA* reveals three such matches: the response regulator RSBQ (Kaneko *et al.*, 2005; PDB codes 1wom and 1wpr), deacetyl cephalosporin C acetyltransferase (Lejon *et al.*, 2008; PDB codes 2vax, 2vat and 2vav) and serine transacetylase (Mirza *et al.*, 2005; PDB code 2b61). The latter two form predicted biological dimers with the same interface as Rv0554. Both bind acetyl-CoA, but the binding residues are generally not conserved in Rv0554 and a tryptophan residue in Rv0554, Trp172, largely blocks the acetyl-CoA binding site seen in the two acetylases; this makes it difficult to infer any functional similarity.

3.2.5. Functional assays. Assays of possible thioesterase activity for Rv0554 proved negative. No activity above control levels was detected for Rv0554 in the generic thioesterase assays, so further more specific DHNA-CoA thioesterase assays were subsequently undertaken using BSA as a negative control and the *E. coli* thioesterase EntH as a positive control. Continuous monitoring of the unique absorbance of DHNA-CoA at 392 nm showed that the level of thioesterase activity for Rv0554 was negligible and comparable with that of BSA, whereas the positive control EntH acts as a good DHNA-CoA hydrolase, as expected.

Assays for possible SHCHC synthase activity of Rv0554 were also negative. No activity was detected when 200 n*M* His-tagged Rv0554 (the concentration was determined by the Bradford method assuming 100% purity of the protein) was assayed (Fig. 6). The same negative result was obtained when freshly prepared His_6-tagged Rv0554 was used in the activity test. In comparison, 5 n*M* His_6-tagged EcoMenH showed high SHCHC synthase activity (2.4 μ*M* min^{-1}) under the same conditions, confirming the validity of the assay.

4. Discussion

Genomic context, coupled with sequence similarity, can provide a powerful indicator of genes encoding enzymes from

biosynthetic pathways. However, such evidence does have to be moderated against the possibility that pathways may differ in detail between different organisms. Ultimately, there is no substitute for experimental testing.

We identified the open reading frame Rv0554 from *M. tuberculosis* as a likely *menH* gene on the basis of a number of indications. It is located between the *menC* and *menD* genes in the *Mtb* genome, within a larger gene cluster that contains almost all of the other *men* genes. The same arrangement is also found within the degenerate *M. leprae* genome. In *M. leprae* there is clear evidence that this homologue is not a pseudogene as the protein has been detected in the cell wall and the cytosol (Marques *et al.*, 2008). The Rv0554 protein was predicted from sequence comparisons to belong to the α/β-hydrolase superfamily and was tentatively annotated as a cofactor-free bromoperoxidase BpoC. Given its genomic context, however, and the fact that the best-characterized MenH protein, the *yfbB* gene product from *E. coli*, is known to be an α/β-hydrolase protein, we hypothesized that Rv0554 might be the *M. tuberculosis* MenH.

The crystal structure of Rv0554 shows that it does indeed have the predicted α/β-hydrolase fold, with all of the associated functional elements, notably a catalytic triad comprising Ser87, His239 and Asp211 and a putative oxyanion hole. Functional assays, however, give no support for either possible MenH activity. The original thioesterase activity suggested for MenH, involving cleavage of CoA from DHNA-CoA, was shown not to be present in *E. coli* MenH and is also not supported by our assays of Rv0554. On the other hand, we also show here that the SHCHC synthase activity shown for *E. coli* MenH is likewise absent from Rv0554. We are forced to conclude that Rv0554 is not the *M. tuberculosis* MenH enzyme. Its genomic association and co-regulation with other menaquinone-biosynthesis enzymes may still be explained if the pathway differs in *Mtb* when compared with other better-characterized organisms. Here, we note that phylogenetic comparisons have shown that the *men* gene ordering in Actinobacteria, to which *Mtb* belongs, is the least similar to other organisms and that some of the individual genes are very hard to identify.

We have examined the Rv0554 structure in the light of functional studies on *E. coli* MenH and identify several differences that would explain the lack of SHCHC synthase activity in Rv0554. Sequence comparisons identified ten nonglycine residues that are conserved in all MenH enzymes (Jiang *et al.*, 2009). Three of these are the catalytic triad residues that are also conserved in Rv0554. Three others, Arg90, Arg124 and Arg168, which are proposed to interact with the substrate in MenH have counterparts in Rv0554 that while not equivalent in sequence location are nevertheless found in similar spatial positions in the active-site cavity. These are Arg113, Arg120 and Arg146 in Rv0554. On the other hand, two other residues shown to be essential for catalysis in *E. coli* MenH, Tyr85 and Trp147, have no counterpart in Rv0554.

Does the structure of Rv0554 give any clue as to its likely true enzymatic function? The fold is unhelpful, since the α/β-hydrolase superfamily encompasses proteins of many different functions that have relatively similar sequence identity (20–30%) to each other. These include proteases, lipases, esterases, C—C-bond hydrolases, enol-lactonases, lactamases, epoxide hydrolases, peroxidases and many others. Some exhibit several activities, for example the cofactor-independent haloperoxidases, which often also exhibit esterase activity (Cheeseman *et al.*, 2004; Holmquist, 2000). In a given α/β-hydrolase it is usually a few key residues removed from the catalytic triad that determine function, coupled with other features such as the presence or absence of a 'lid', the size and shape of any tunnel or cavity associated with the active site, the presence of moveable elements and possibly the oligomerization state.

Here, we note several features that seem likely to be functionally important. The presence of three arginines in the active-site region and the binding of malonic acid both point to the probability of an acidic substrate; malonic acid not only stabilized the crystals but also bound in the active site. In molecule *B* of the tetragonal crystal structure the change in position of residues 113–125 brings Arg120 into the active site where it contacts one of the two malonic acid molecules, suggesting that this structural rearrangement could be triggered by substrate binding. The intimate dimer formed by docking of the lid domains of two molecules also seems likely to be functionally important, especially in view of the relationship between dimerization and active-site structure. The change in position of residues 113–125 means that several residues previously at the dimer interface become part of the active site, suggesting a mechanism for possible feedback between active sites. Very similar dimers are found for two other α/β-hydrolase enzymes, serine transacetylase (PDB code 2b6l) and deacetyl cephalosporin C acetyltransferase (PDB code 2vax). These enzymes both bind CoA, but comparisons with Rv0554 show that the CoA-binding residues are not conserved in Rv0554 and that Trp172 in Rv0554 blocks part of the binding site used by CoA in those enzymes.

5. Conclusions

Despite its genomic context and predicted α/β-hydrolase fold, which both suggested that it was the 'missing' *M. tuberculosis* MenH enzyme, Rv0554 proved not to have the enzymatic activity associated with MenH. The crystal structure showed that it does indeed have the α/β-hydrolase fold, with an intact catalytic triad, but it appears to lack two essential residues that are required for MenH activity. Rv0554 is shown to form an intimate dimer, within which a structural rearrangement can occur that could allow feedback between the two active sites. The binding of small molecules in the active-site cavity points to a likely substrate-binding region, with the favoured binding of malonic acid, together with the presence of three arginine residues in the active-site cavity, further suggesting a preference for acidic substrates. Nevertheless, the true enzymatic activity of Rv0554 cannot yet be ascertained and must await further biological data.

Funding from the Foundation for Research, Science and Technology of New Zealand is gratefully acknowledged.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.

Anishetty, S., Pulimi, M. & Pennathur, G. (2005). *Comput. Biol. Chem.* **29**, 368–378.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Cheeseman, J. D., Tocilj, A., Park, S., Schrag, J. D. & Kazlauskas, R. J. (2004). *Acta Cryst.* D**60**, 1237–1243.

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

Dhiman, R. K., Mahapatra, S., Slayden, R. A., Boyne, M. E., Lenaerts, A., Hinshaw, J. C., Angala, S. K., Chatterjee, D., Biswas, K., Narayanasamy, P., Kurosu, M. & Crick, D. C. (2009). *Mol. Microbiol.* **72**, 85–97.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. & Liang, J. (2006). *Nucleic Acids Res.* **34**, W116–W118.

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

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U. & Sali, A. (2006). *Curr. Protoc. Bioinformatics*, ch. 5, Unit 5.6.

Glasner, M. E., Fayazmanesh, N., Chiang, R. A., Sakai, A., Jacobson, M. P., Gerlt, J. A. & Babbitt, P. C. (2006). *J. Mol. Biol.* **360**, 228–250.

Hecht, H.-J., Sobek, H., Haag, T., Pfeifer, O. & van Pee, K. H. (1994). *Nature Struct. Biol.* **1**, 532–537.

Heikinheimo, P., Goldman, A., Jeffries, C. & Ollis, D. L. (1999). *Structure*, **7**, R141–R146.

Holm, L., Kaariainen, S., Rosenstrom, P. & Schenkel, A. (2008). *Bioinformatics*, **24**, 2780–2781.

Holm, L., Kaariainen, S., Wilton, C. & Plewczynski, D. (2006). *Curr. Protoc. Bioinformatics*, ch. 5, Unit 5.5.

Holmquist, M. (2000). *Curr. Protein Pept. Sci.* **1**, 209–235.

Holsclaw, C. M., Sogi, K. M., Gilmore, S. A., Schelle, M. W., Leavell, M. D., Bertozzi, C. R. & Leary, J. A. (2008). *ACS Chem. Biol.* **3**, 619–624.

Jiang, M., Chen, X., Guo, Z.-F., Cao, Y., Chen, M. & Guo, Z. (2008). *Biochemistry*, **47**, 3426–3434.

Jiang, M., Chen, X., Wu, X.-H., Chen, M., Wu, Y.-D. & Guo, Z. (2009). *Biochemistry*, **48**, 6921–6931.

Kaneko, T., Tanaka, N. & Kumasaka, T. (2005). *Protein Sci.* **14**, 558–565.

Kim, S. Y., Lee, B. S., Shin, S. J., Kim, H. J. & Park, J. K. (2008). *J. Med. Microbiol.* **57**, 1473–1480.

Krawczyk, J., Kohl, T. A., Goesmann, A., Kalinowski, J. & Baumbach, J. (2009). *Nucleic Acids Res.* **37**, e97.

Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Kurosu, M. & Begari, E. (2010). *Molecules*, **15**, 1531–1553.

Kurosu, M., Narayanasamy, P., Biswas, K., Dhiman, R. & Crick, D. C. (2007). *J. Med. Chem.* **50**, 3973–3975.

Lejon, S., Ellis, J. & Valegard, K. (2008). *J. Mol. Biol.* **377**, 935–944.

Lu, X., Zhang, H., Tonge, P. J. & Tan, D. S. (2008). *Bioorg. Med. Chem. Lett.* **18**, 5963–5966.

Marques, M. A., Neves-Ferreira, A. G., da Silveira, E. K., Valente, R. H., Chapeaurouge, A., Perales, J., da Silva Bernardes, R., Dobos, K. M., Spencer, J. S., Brennan, P. J. & Pessolani, M. C. (2008). *Proteomics*, **8**, 2477–2491.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

McPherson, A. & Cudney, B. (2006). *J. Struct. Biol.* **156**, 387–406.

Mdluli, K. & Spigelman, M. (2006). *Curr. Opin. Pharmacol.* **6**, 459–467.

Meganathan, R. (2001). *Vitam. Horm.* **61**, 173–218.

Mirza, I. A., Nazi, I., Korczynska, M., Wright, G. D. & Berghuis, A. M. (2005). *Biochemistry*, **44**, 15768–15773.

Moreland, N., Ashton, R., Baker, H. M., Ivanovic, I., Patterson, S., Arcus, V. L., Baker, E. N. & Lott, J. S. (2005). *Acta Cryst.* D**61**, 1378–1385.

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

Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., Sussman, J. L., Verschueren, K. H. G. & Goldman, A. (1992). *Protein Eng.* **5**, 197–211.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *Acta Cryst.* D**53**, 448–455.

Raman, K., Yeturu, K. & Chandra, N. (2008). *BMC Syst. Biol.* **2**, 109.

Sassetti, C. M., Boyd, D. H. & Rubin, E. J. (2003). *Mol. Microbiol.* **48**, 77–84.

Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.

Terwilliger, T. C. (2004). *Acta Cryst.* D**60**, 2144–2149.

Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.

Wilson, W. W. (2003). *J. Struct. Biol.* **142**, 56–65.

Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.