Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structures of *Mycobacterium tuberculosis* pyridoxine 5′-phosphate oxidase and its complexes with flavin mononucleotide and pyridoxal 5′-phosphate

Bichitra K. Biswal, Maia M. Cherney, Meitian Wang, Craig Garen and Michael N. G. James*

CIHR Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton T6G 2H7, Canada

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

Received 11 May 2005
Accepted 19 August 2005

PDB References: pyridoxine 5′-phosphate oxidase, 1xxo, r1xxosf; FMN complex, 1y30, r1y30sf; PLP complex, 2aq6, r2aq6sf.

The X-ray crystal structure of a conserved hypothetical protein of molecular weight 16.3 kDa from *Mycobacterium tuberculosis* corresponding to open reading frame (ORF) Rv1155 has been solved by the multiwavelength anomalous dispersion method and refined at 1.8 Å resolution. The crystal structure revealed that Rv1155 is a dimer in the crystal and that each monomer folds into a large and a small domain; the large domain is a six-stranded antiparallel β-barrel flanked by two small α-helices and the small domain is a helix–loop–helix motif. The dimer interface is formed by residues protruding primarily from five of the six β-strands in each subunit. Based on structural similarity and on ligand binding, it has been established that Rv1155 is a pyridoxine 5′-phosphate oxidase, the *Escherichia coli* and human counterparts of which catalyse the terminal step in the biosynthesis of pyridoxal 5′-phosphate (PLP), a cofactor used by many enzymes involved in amino-acid metabolism. The structures of flavin mononucleotide (FMN) and pyridoxal 5′-phosphate (PLP) bound separately to Rv1155 have been determined at 2.2 and 1.7 Å resolution, respectively. Only one monomer binds non-covalently to one FMN molecule or to one PLP molecule. Arg55 and Lys57 are the key residues making hydrogen bonds and ionic interactions with the phosphate and ribose groups of the FMN molecule, whereas Arg55 and Arg129 provide hydrogen bonds and ionic interactions with the phosphate group of the PLP. Structural comparisons of Rv1155 from *M. tuberculosis* with its *E. coli* and human counterparts demonstrate that the core structure is highly conserved and the FMN-binding site is similarly disposed in each of the structures.

1. Introduction

Tuberculosis (TB) is a serious chronic infectious disease caused by the tubercle bacillus. It is estimated that there are currently 2.2 billion people infected with *Mycobacterium tuberculosis* worldwide, leading to ∼2 million deaths annually (Dye *et al.*, 1999; Corbett *et al.*, 2003; Cosma *et al.*, 2003; Toossi, 2003; Ahmed & Hasnain, 2004; Coker, 2004; Monack *et al.*, 2004; Ojcius, 2004; Tripathi *et al.*, 2005). Furthermore, existing therapies require prolonged periods of administration of the drugs and because of noncompliance multidrug resistance has evolved (Coker, 2004). This has led to further studies designed to elucidate the various factors responsible for the survival of this organism. Although the complete genome sequence of the best-characterized strain of *M. tuberculosis*, H37Rv, has provided a wealth of information on the functional aspects of this organism (Philipp *et al.*, 1996; Cole *et al.*, 1998; Camus *et al.*, 2002), biological functions of many of the gene products are still to be determined. The *M. tuberculosis* structural genomics consortium was formed in 2000 (http://www.doe-mbi.ucla.edu/TB); the consortium aims to provide a

structural basis for the development of therapeutics for tuberculosis (Goulding *et al.*, 2002, 2003; Rupp *et al.*, 2002; Rupp, 2003; Smith & Sacchettini, 2003; Terwilliger *et al.*, 2003).

As a member of the *M. tuberculosis* structural genomics consortium, we have targeted ~150 proteins of unknown function and have begun to determine their X-ray crystal structures in order to help in the elucidation of their biological functions, based on their structural similarity to proteins of known function. In this paper, we report the X-ray crystal structure of a conserved hypothetical protein, Rv1155, and its structures with flavin mononucleotide (FMN) and pyridoxal 5′-phosphate (PLP) bound separately. This structural characterization has provided insights into the prediction of the biological function solely from the three-dimensional structure.

2. Experimental methods

2.1. Cloning, expression and purification

H37Rv genomic DNA in a BAC ordered library from L'Institut Pasteur (Brosch *et al.*, 1998; Gordon *et al.*, 1999) was used as the template to amplify Rv1155 by PCR using the following primers: sense, **GGGGACAAGTTTGTACAAAAAAGCAGGCTCA***GAAAACCTGTATTTTCAGGGC*ATGGCCCGCCAAGTCTTCG; antisense, **GGGGACCACTTTGTACAAGAAAGCTGGGTC**TTAGCGATACCGGGCGGCG. The amplified open reading frame (ORF) was gel purified (Qiagen) and then inserted directionally into a cloning vector using the homologous recombination sites (bold) using the Gateway Cloning System (Invitrogen). Rv1155 was then transferred by recombination into an expression vector encoding an amino-terminal GST-fusion protein. Additionally, a site recognized by recombinant tobacco etch virus (rTEV) protease (Invitrogen) encoded by the forward primer (italicized) was used for removal of the fusion tag. The resulting expression plasmid, pGST-1155, was confirmed by DNA-sequence analysis (DNA Sequence Facility, Department of Biochemistry, University of Alberta, Canada).

Rv1155 was expressed in *Escherichia coli* BL21 (DE3) pLysS cells (Novagen). A 16 h culture of the bacteria transformed with the expression vector in LB supplemented with 100 µg ml^{-1} ampicillin and 34 µg ml^{-1} chloramphenicol (LBAC) was diluted 1:40 into 2 l LBAC. Incubation at 310 K proceeded for 2 h, at which point OD $_{600\,\mathrm{nm}} \simeq 0.9$. The temperature was shifted to 298 K and expression was induced by adding isopropyl thiogalactopyranoside (IPTG) to a final concentration of 1 m*M*. After 16 h incubation, the cells were harvested by centrifugation at 2150*g* for 20 min. Bacterial pellets were resuspended in PBS, EDTA-free complete protease inhibitor (Roche) and 1 m*M* DTT (Fischer Scientific). Resuspended cells were frozen at 193 K. Thawed cells were subjected to sonication on ice. The lysate was cleared by centrifugation (30 min, 30 000*g*) and the supernatant was loaded onto a GSTPrep glutathione Sepharose column (GE Healthcare). The column was washed with PBS until $A_{280} \simeq 0$ and the GST-1155 fusion protein was competitively eluted from the column with 10 m*M* reduced glutathione (Sigma) in 50 m*M* Tris–HCl pH 8.0. The GST tag and the N-terminally encoded recombination site were removed by proteolytic cleavage using the rTEV protease. Following extensive dialysis in PBS, the cleaved protein mix was applied once again onto the GSTPrep column. The flowthrough from this column was retained and then concentrated using an Amicon Ultra 5 kDa cutoff (Millipore). Buffer was exchanged to 10 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl and 1 m*M* DTT during concentration. The final protein concentration was determined to be 10 mg ml^{-1} using a Bradford Protein Assay Kit (BioRad) with BSA (Pierce) as a standard. 16% SDS–PAGE analysis of the concentrated protein revealed a single band of ~16 kDa, consistent with the predicted molecular weight of Rv1155 (16.3 kDa).

2.2. Crystallization

The initial crystallization of Rv1155 was developed using the sitting-drop vapour-diffusion method in 96-well Intelliplates (Hampton Research). Rv1155 was dialysed and concentrated against 50 m*M* Tris–HCl pH 8.0, 25–50 m*M* KF and 10 m*M* glutathione to a final concentration of 9 mg ml^{-1}. A high-throughput crystal screen (Hampton Research) was performed using equal volumes (0.5 µl) of protein and precipitating solution. From this initial screen, relatively poor crystals were obtained using 10% PEG 8000, 0.1 *M* HEPES pH 7.5 and 8% ethylene glycol as precipitating agent. Optimization of the crystallization conditions was accomplished using a macroseeding technique (Stura & Wilson, 1991). Hanging drops containing equal volumes of protein and precipitating solution (12% PEG 8000, 8% ethylene glycol and 30 m*M* KF) were pre-equilibrated against 1 ml precipitating solution overnight prior to seeding. After seeding, crystals grew to 0.05 × 0.2 × 0.3 mm within 1–2 weeks. For data collection, crystals were cryocooled to 100 K after a quick dip into cryoprotectant consisting of 30% ethylene glycol, 12% PEG 8000 and 50 m*M* KF.

2.3. Heavy-atom derivative preparation

Rv1155 crystals were transferred to a drop containing 10 m*M* K_3IrCl_6 in the precipitating solution and soaked for 10 min. Crystals were then cryocooled as described above.

2.4. Preparation of complexes of Rv1155 with FMN and PLP

Crystals of Rv1155 were transferred into 2 m*M* FMN or PLP in the precipitating solution and soaked overnight. They were then flash-cooled using cryoprotectant as described above. FMN or PLP were not present in the cryoprotectant solution.

2.5. Data collection

Crystals belong to the monoclinic space group $P2_1$ with approximate unit-cell parameters $a = 47$, $b = 55$, $c = 55$ Å, $\beta = 108^\circ$ (Table 1). The crystal asymmetric unit contains two molecules with $V_M = 2.2$ Å^3 Da^{-1} and the crystals have a

solvent content of 44% (Matthews, 1968). Multiwavelength anomalous dispersion (MAD) data at the peak anomalous signal and high-energy remote were collected from a K_3IrCl_6-soaked crystal at beamline 8.3.1 at the Advanced Light Source, Berkeley, CA, USA. Data sets from the native crystals and PLP-soaked crystals were collected at the same beamline. Data from the crystal soaked with FMN were collected in-house using an R-AXIS IV^{++} image-plate detector with Cu $K\alpha$ radiation generated by a Rigaku RU-300 rotating-anode X-ray generator. All data sets were integrated and scaled with the programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). Details of the resulting data-collection statistics are given in Table 1.

Table 1
Data-collection and refinement statistics.

Value in parentheses are for the highest resolution shell.

	Peak	Remote	Native	FMN complex	PLP complex
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
a (Å)	47.37	47.48	46.94	46.82	46.87
b (Å)	55.36	55.39	55.38	55.25	55.35
c (Å)	55.31	55.41	54.84	54.87	54.52
α (°)	90	90	90	90	90
β (°)	108.68	108.86	108.17	108.36	107.77
γ (°)	90	90	90	90	90
Z	4	4	4	4	4
Data collection					
Temperature (K)	100	100	100	100	100
Detector	ADSC Q210	ADSC Q210	ADSC Q210	R-AXIS IV	ADSC Q210
Wavelength (Å)	1.1055	1.0332	1.3397	1.5418	1.1158
Resolution (Å)	40–2.2	40–2.2	40–1.8	40–2.2	40–1.7
Highest resolution shell (Å)	2.28–2.2	2.28–2.2	1.86–1.8	2.28–2.2	1.76–1.7
Unique reflections	13965	13626	23158	12534	28603
Multiplicity	5.3	5.4	2.9	2.5	2.9
$I/\sigma(I)$	11.0 (2.7)	10.4 (2.6)	13.3 (1.8)	8.2 (3.5)	13.7 (2.2)
Completeness (%)	99.7 (99.9)	99.8 (99.9)	93.2 (73.6)	94.4(92.0)	97.5 (90.1)
R_{sym}† (%)	12.1 (48.4)	11.9 (49.2)	6.7 (45.1)	13.0 (42.5)	6.6 (30.4)
R_{anom}‡ (%)	3.1	3.4	—	—	—
Refinement					
Refinement resolution (Å)	—	—	40–1.8	40–2.2	40–1.7
$R_{working}/R_{free}$§	—	—	0.19/0.23	0.19/0.25	0.191/0.194
No. of atoms					
Protein	—	—	2232	2232	2232
Water	—	—	260	227	379
Average B factor (Å^2)					
Protein	—	—	16.9	18.3	14.1
Water	—	—	27.4	25.6	28.6
R.m.s. deviations from ideality					
Bond lengths (Å)	—	—	0.005	0.006	0.005
Bond angles (°)	—	—	1.2	1.3	1.3
Residues in Ramachandran plot (%)					
Most favoured regions	—	—	91.5	89.4	93.5
Additional allowed regions	—	—	7.7	10.2	6.5

† $R_{sym} = \sum_{\mathbf{h}}\sum_i[|I_i(\mathbf{h}) - \langle I(\mathbf{h})\rangle|]/\sum_{\mathbf{h}}\sum_i I_i(\mathbf{h})$, where $I_i(\mathbf{h})$ is the ith intensity measurement and $\langle I(\mathbf{h})\rangle$ is the weighted mean of all measurements of $I(\mathbf{h})$. ‡ $R_{anom} = \sum_{\mathbf{h}}|I_{\mathbf{h}}^+ - I_{\mathbf{h}}^-|/\sum_{\mathbf{h}}|I_{\mathbf{h}}^+ + I_{\mathbf{h}}^-|$ over all $\mathbf{h}$, where $I_{\mathbf{h}}$ is the intensity of reflection $\mathbf{h}$. § $R_{working}$ and $R_{free} = \sum_{\mathbf{h}}||F(\mathbf{h})_o| - |F(\mathbf{h})_c||/\sum_{\mathbf{h}}|F(\mathbf{h})_o|$, where $|F(\mathbf{h})_o|$ and $|F(\mathbf{h})_c|$ are observed and calculated amplitudes, respectively. R_{free} was calculated using 5% of data.

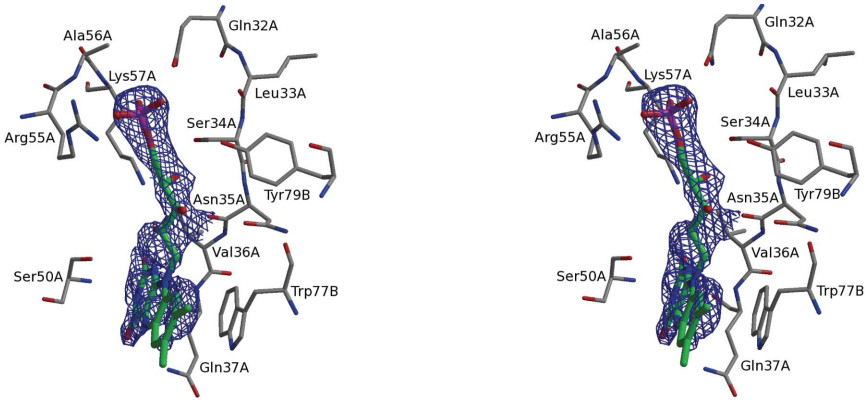

Figure 1
Stereoview of $2|F_o| - |F_c|$ electron-density map contoured at the 1σ level showing the bound FMN. Active-site residues are shown in ball-and-stick representation, with C, N and O atoms in grey, blue and red colours, respectively. This figure was prepared using the program *BOBSCRIPT* (Esnouf, 1999).

2.6. Phasing, model building and refinement

The position of a single Ir atom in the crystal asymmetric unit was determined and was used to calculate multi-wavelength anomalous diffraction phases using the program *SOLVE* (Terwilliger & Berendzen, 1999; Terwilliger, 2000). The phases were improved by density modification and the resultant electron-density map at 2.2 Å was used to build 80% of the model using the program *RESOLVE* (Terwilliger & Berendzen, 1999; Terwilliger, 2000). The resultant model was subjected to 50 cycles of rigid-body refinement using the program *CNS*1.1 (Brünger *et al.*, 1998), treating each monomer as a rigid group. At this stage, the model was refined against the high-resolution (1.8 Å) native data and the remainder of the model was fitted into the calculated $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ electron-density maps using the program *XtalView* (McRee, 1999). Structure refinement was continued using the program *CNS* v.1.1 using maximum likelihood as the target function. Initially, the model was subjected to 3000 K simulated-annealing refinement, followed by positional and individual B-factor refinements. After ensuring that all amino-acid residues fit correctly into the electron-density contours, the identification of water molecules began. This was completed in several stages based on $2|F_o| - |F_c|$ (1σ contour level) and $|F_o| - |F_c|$ (3σ contour level)

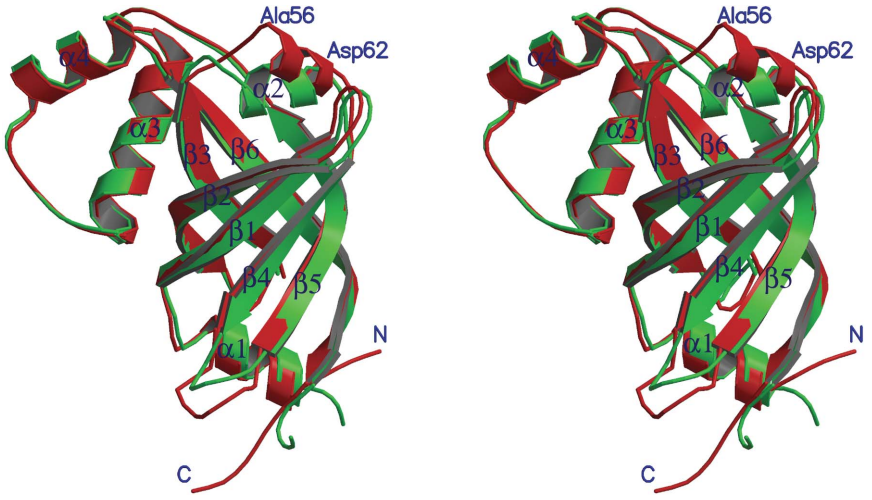

Figure 2
Stereoview of the superimposition of both subunits of the dimer showing the differences between the subunits in the region Ala56–Arg62. Green and red colours represent the subunits *A* and *B*, respectively. Figs. 2–7 were prepared using the programs *MOLSCRIPT* (Kraulis, 1991) and *RASTER3D* (Merritt & Murphy, 1994).

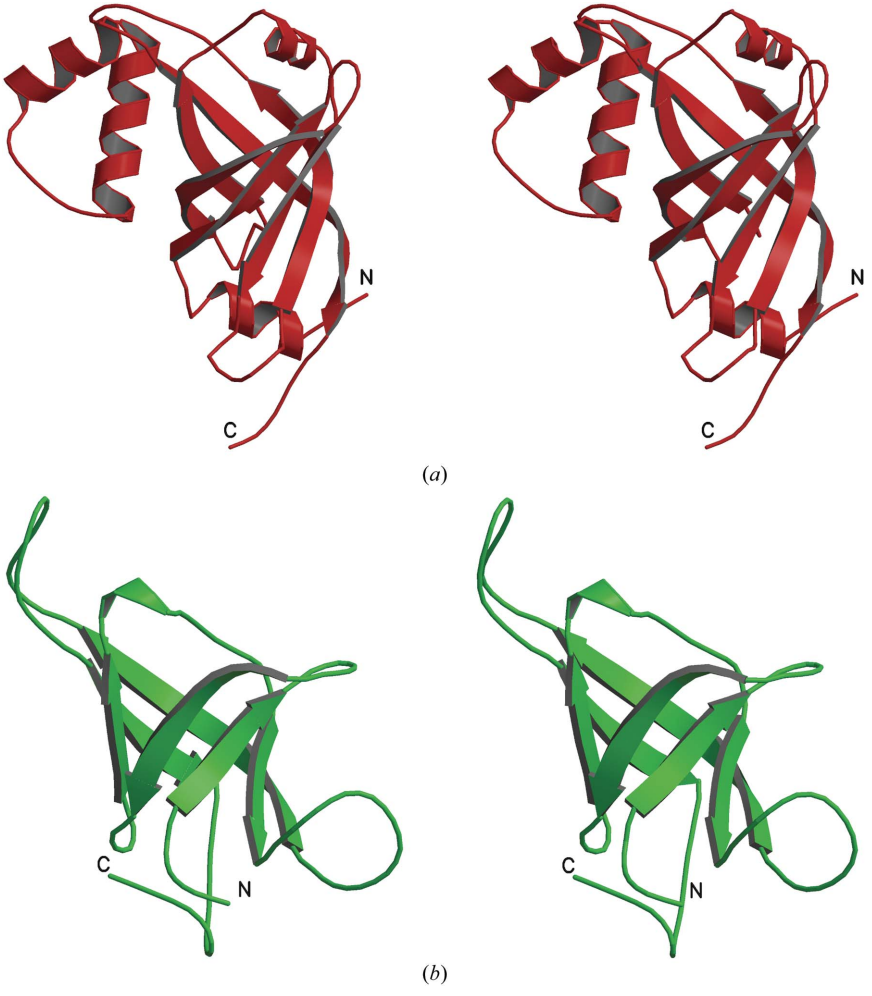

Figure 3
Stereoviews of (*a*) the Rv1155 monomer and (*b*) the α domain of chymotrypsin (Gly25–Ala120) showing their common structural feature of a six-stranded Greek-key β-barrel.

electron-density maps. Cycles of position and individual *B*-factor refinements after every stage of water identification continued until no significant electron-density peaks remained in the electron-density maps. R_{work} and R_{free} were monitored closely throughout the refinement. NCS restraints were not used in the refinement. The structures of the FMN and PLP complexes were solved by molecular replacement using the program *CNS* v.1.1. A difference Fourier map, $|F_{PC}| - |F_P|$, α_{calc} ($|F_{PC}|$ are the structure-factor amplitudes from the protein–FMN or protein–PLP complexes and $|F_P|$ are the structure-factor amplitudes from the apo protein) permitted an initial positioning of the FMN and PLP molecules into the respective difference electron-density maps. Subsequently, both complex structures were refined in the same manner as the native structure. The first four residues from the N-terminus of each monomer were not observed in the electron-density maps. The stereochemical acceptability of the structures was checked by the program *PROCHECK* (Laskowski *et al.*, 1993). Details of the refinement statistics are given in Table 1. A $2|F_o| - |F_c|$ electron-density map corresponding to the bound FMN molecule is shown in Fig. 1.

3. Results

3.1. Structure of Rv1155

The crystallographic asymmetric unit of the $P2_1$ space group contains a dimer of Rv1155. Each Rv1155 monomer consists of 147 amino-acid residues folded into two domains. The larger domain comprises amino-acid residues Val5–Leu88 and Arg129–Arg147; it adopts a six-stranded β-barrel flanked by two short α-helices (Fig. 2). The smaller domain resembles a helix–loop–helix motif. Although both subunits of the dimer have the same fold, subtle structural differences exist between the subunits. The root-mean-square deviation (r.m.s.d.) between the two subunits of the dimer is 0.77 Å (129 C^{α} pairs with a maximum deviation of 1.88 Å). The major differences between the subunits lie mainly in the N- and C-termini in strands β1 and β2 and in helix α2 (Fig. 2). Surprisingly, residues Ala56–Arg62 form an α-helix in one of the subunits, whereas the corresponding region in the other subunit

conforms more closely to a 3_{10}-helix (Fig. 2). Furthermore, the r.m.s. deviation of C^{α} atoms in this region (Ala56–Arg62, not included among the 129 residues above) is 4.73 Å. The dimer interface consists of residues mainly from strands $\beta1$, $\beta2$, $\beta4$, $\beta5$ and $\beta6$ of both subunits and buries 2350 Å^2 surface area compared with the total surface area of 8000 Å^2 of each subunit, determined using a probe radius of 1.4 Å. Residues from each subunit that contribute to formation of the dimer interface are: Ile19, Val21, Ala23, Ile25, Lys26, Asp28, Gly29, Arg30, Pro31, Gln32, Leu33, Ser34, Asn35, Val36, Gln37, Arg66, Ala67, Leu70, Trp77, Tyr79, Val81, Glu83, Gly84 and Tyr140.

Interestingly, the monomeric fold of Rv1155 resembles a chymotrypsin fold with one six-stranded Greek-key β-barrel (Fig. 3), suggesting a common pathway of evolution of chymotrypsin and Rv1155. Although both Rv1155 and chymotrypsin have two domains and each domain has the structural characteristics of one six-stranded Greek key β-barrel, the organizations of these domains to form the biologically functional entity are significantly different, implicating different biological functions. In Rv1155, both domains of the dimer are related by twofold symmetry and presumably function like an oxidase, as discussed later in this paper, whereas the two domains of the chymotrypsin are approximately related by a translation followed by a 90° rotation and it has therefore acquired a protease activity (Read & James, 1988).

The crystal structure of Rv1155 of *M. tuberculosis* has recently been reported by another research group (PDB code 1w9a; Canaan *et al.*, 2005). Superimposition of both structures gives an r.m.s.d. value of 0.14 Å for 268 C^{α} atoms. The two structure determinations are exceedingly similar; however, the Canaan group stated that they were unable to bind FMN. They also carried out a fluorescence titration experiment that showed no changes in the presence of FMN. These results are at variance with our crystallographic results described below; however, a literature survey indicates that flavin-binding proteins quench the fluorescence of their FMN or FAD ligand.

3.2. Elucidation of biological function

In an attempt to elucidate the biological function of Rv1155 from its three-dimensional structure, a search for close structural homologues of Rv1155 using *DALI* (Holm & Sander, 1998; EMBL–EBI) revealed a best match to the FMN-binding protein from *Desulfovibrio vulgaris* (PDB code 1axj). Additionally, a conserved-domain search based on the primary structure of Rv1155 using the NCBI conserved-domain search server (Marchler-Bauer & Bryant, 2004) predicted that Rv1155 resembles human pyridoxine 5′-phosphate oxidase (PNPOx), the enzyme that catalyses the terminal step in the biosynthesis of pyridoxal 5′-phosphate (PLP), a cofactor used by many enzymes involved in amino-acid metabolism (Zhao & Winkler, 1995; Safo *et al.*, 2000, 2001; di Salvo *et al.*, 2002, 2003; Musayev *et al.*, 2003). The PNPOx enzyme uses FMN as a cofactor. Encouraged by these predictions, we soaked the crystals of Rv1155 in 2 m*M* solutions of FMN or PLP. As anticipated, we found that Rv1155 bound both FMN and PLP.

The FMN-binding pocket is located on the surface of the β-barrel and is constituted of amino-acid residues mainly from helix $\alpha2$ and strands $\beta2$ and $\beta3$ (Fig. 4*a*). The FMN phosphate moiety is positioned near the N-termini of helix $\alpha2$ and strand $\beta2$ and the isoalloxazine ring points down between strands $\beta2$ and $\beta3$ (Fig. 4*a*). Understandably, the general location of the FMN-binding pocket in Rv1155 is the same as that in the *E. coli* and human PNPOx enzymes. In Rv1155, Gln32*A*, Leu33*A*, Ser34*A*, Asn35*A*, Val36*A*, Gln37*A*, Ser50*A*, Arg55*A*, Ala56*A*, Lys57*A*, Trp77*B*, Tyr79*B* and Gly141*B* form the binding pocket of the FMN molecule (Fig. 4*b*). The O1P atom of the phosphate group of the FMN is hydrogen bonded to Glu32*A* OE1 and Lys57*A* N of the Rv1155 (Fig. 4*b*). Arg55*A* NE and NH2 make favorable electrostatic interactions with the phosphate O1P atom. The O2P atom of the phosphate moiety of FMN is hydrogen bonded to the residues Gln32*A*, Leu33*A*, Try79*B* and Gly141*B* by water-mediated interactions. The O atoms of the ribose sugar are hydrogen

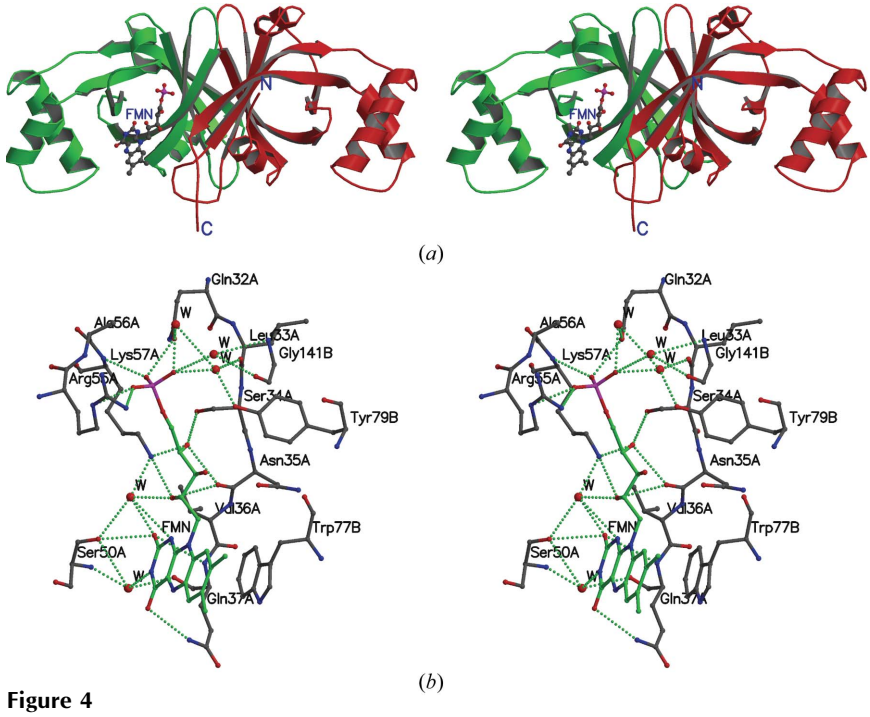

Figure 4
(*a*) Stereoview of Rv1155 with bound FMN in the active site. (*b*) Rv1155 and FMN interactions. Protein residues are shown in ball-and-stick representation with C, N and O atoms in grey, blue and red colours, respectively. FMN is shown in ball-and-stick representation with C, N and O atoms in green, blue and red colours, respectively. Hydrogen bonds are shown in green. A and B after the residue name correspond to the subunits *A* and *B*, respectively.

bonded to Asn35*A* O and Lys57A NZ. Both O atoms of the flavin ring are hydrogen bonded to Ser50*A* O and Gln37*A* NE2. The indole group of Trp77*B* protruding from the other subunit provides stacking interactions with the dimethyl benzyl ring of the flavin moiety. The detailed interactions are depicted in Fig. 4(*b*).

PLP binds in the same pocket but with a completely different orientation of the phosphate group and pyridine ring (Fig. 5*a*). The phosphate group of PLP is located towards strands $\beta3$ and $\beta6$, whereas it is the substituted pyridine ring that points towards the N-termini of helix $\alpha2$ and strand $\beta2$ (Fig. 5*a*). Many of the residues involved in FMN binding also contribute to PLP binding. Residues Gln32*A*, Leu33*A*, Ser34*A*, Asn35*A*, Arg55*A*, Ala56*A*, Lys57*A*, Arg129*A* and Tyr79*B* form the PLP-binding site (Fig. 5*b*). Arg55A NH1 and NH2 and Arg129A NH1 are hydrogen bonded to the phosphate O atoms of PLP. The O atoms of the pyridine ring are hydrogen bonded to Ser34*A* OG, Asn35A O and Tyr79B OH. The similarity between the interactions of Rv1155 with FMN and with PLP involves a constellation of water molecules in the binding pocket providing water-mediated hydrogen bonds between Rv1155 and the two ligands. Only one molecule (subunit *A*) of the dimer binds non-covalently to one FMN or PLP molecules; the other does not, presumably because of crystal packing and the altered architecture of the FMN-binding domain. In particular, Lys57, a key residue at the N-terminus of helix $\alpha2$ that is involved in FMN as well as PLP binding, adopts very different conformations in each subunit (Fig. 2). Similar observations of dimer asymmetry and cofactor binding to one monomer only have also been observed in the case of an enzyme thymidylate synthase (Anderson *et al.*, 1999).

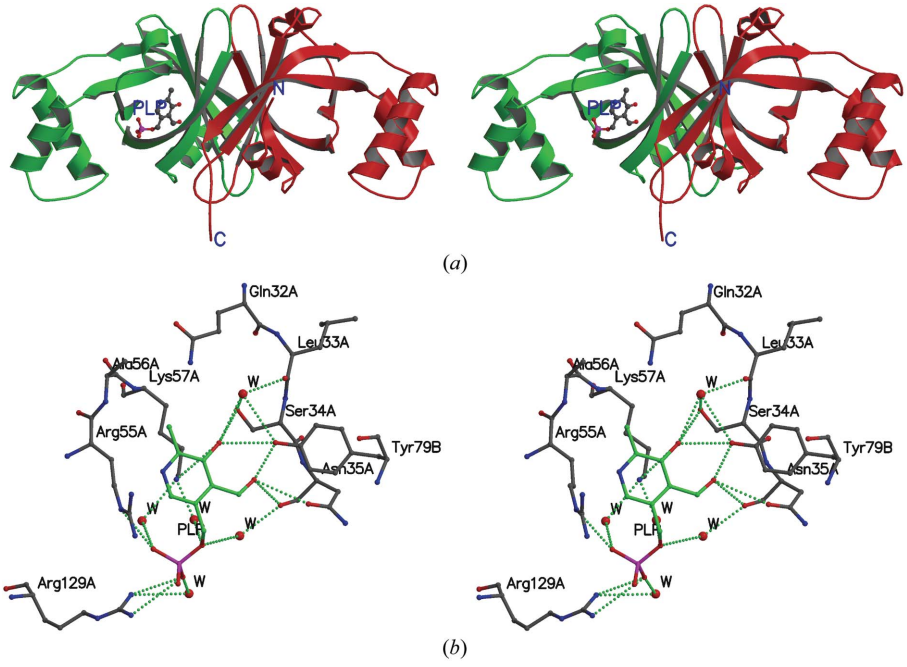

Figure 5
(*a*) Stereoview of the Rv1155 with bound PLP. PLP is shown in ball-and-stick representation. (*b*) Rv1155 and PLP interactions. The colour codes used to represent protein residues and PLP molecules are same as in Fig. 4.

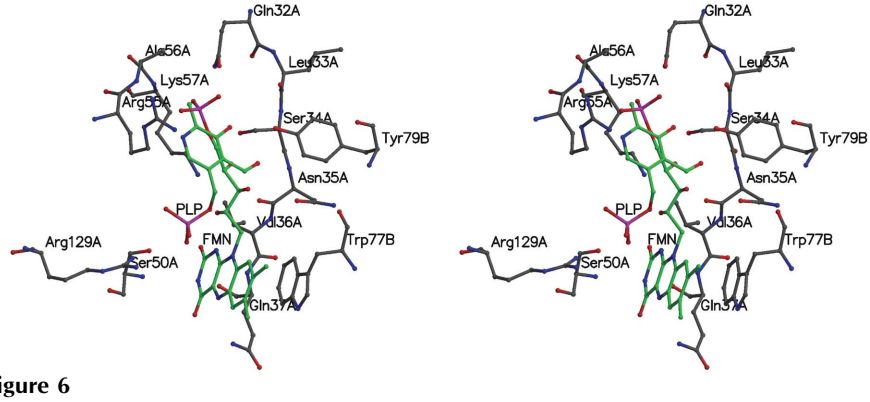

Figure 6
Stereoview of the superimposition of the Rv1155–FMN and Rv1155–PLP complex structures showing the relative orientation of FMN and PLP. The colour codes used to represent protein residues, FMN and PLP are the same as in Figs. 4 and 5.

Overall, the structure of Rv1155 does not undergo major conformational changes upon FMN or PLP binding. The r.m.s. deviations between the structures of unbound Rv1155 and the FMN complex and between unbound Rv1155 and the PLP complex are only 0.12 and 0.09 Å, respectively. However, in the PLP-complex structure the C^{α} atoms of residues Arg55*A*, Ala56*A* and Lys57*A*, part of PLP-binding site, are moved with respect to the unliganded Rv1155 by 0.23, 0.63 and 0.50 Å, respectively. Although these residues also form a part of the FMN-binding site in the Rv1155–FMN complex structure, no such movements of similar magnitude were detected in the FMN-binding site.

As both the FMN- and the PLP-complex structures were determined separately, in order to learn about the relative orientation of FMN and PLP in the Rv1155 structure both complex structures were superimposed and are shown in Fig. 6. It was found that the pyridine ring of PLP and the FMN phosphate moiety occupy the same spatial position (Fig. 6). On the other hand, the complex structures of *E. coli* and human PNPOx enzymes with both FMN and PLP bound simultaneously revealed that the pyridine ring of PLP is stacked parallel against the FMN isoalloxazine ring (Safo *et al.*, 2001; Musayev *et al.*, 2003). We have attempted to form the ternary complex of FMN, PLP and Rv1155; however, the resulting crystals did not diffract well. Further attempts at cocrystallization of the ternary complex are being made.

3.3. Comparison of Rv1155 structure with its *E. coli* and human counterparts

The above results have prompted us to carry out further structural comparisons between Rv1155 and its *E. coli* and human counterparts. The crystal structures of *E. coli* PNPOx (PDB code 1g76) and human PNPOx (PDB code 1nrg) and their ternary complexes with FMN and PLP have been reported (Safo *et al.*, 2001; Musayev *et al.*, 2003). Superimpositions of the monomeric structures of FMN-bound Rv1155 with its *E. coli* and human counterparts using the program *ALIGN* (Cohen, 1997) resulted in r.m.s. deviations of 2.06 Å (for 113 C^{α} pairs with a maximum distance 6.5 Å) and 2.08 Å (for 110 C^{α} pairs with a maximum distance 5.5 Å), respectively. The major differences lie in the N-terminal region, the smaller helix–loop–helix domain and in the C-terminal region. The smaller domain of Rv1155 is a helix–loop–helix motif, whereas the corresponding domain in the *E. coli* as well as in the human PNPOx enzymes is a three-helix bundle. There are two extra β-strands in *E. coli* and human PNPOx compared with Rv1155. The core structures, however, agree much better. Understandably, both *E. coli* and human PNPOx structures can be superimposed with an r.m.s.d. of 0.77 Å (182 C^{α} pairs with a maximum distance of 2.0 Å), indicating a very high degree of structural similarity between the two. Interestingly, the core β-barrel structure of the monomer is conserved in the three organisms (Fig. 7). The three proteins fold into a similar three-dimensional structure despite their low sequence identity (Fig. 8). The *E. coli* and human PNPOx enzymes and Rv1155 contain 218, 261 and 147 amino acids, respectively. The FMN-binding pocket is very similar; however, the bound conformation of the FMN molecule is slightly different (Fig. 7). The binding modes of the phosphate and sugar moieties of the FMN in Rv1155 are similar to those in *E. coli* and human enzymes. On the other hand, the isoalloxazine ring of the FMN adopts a different conformation in Rv1155 compared with that in the *E. coli* and human enzymes (Fig. 7). Superimposition of the *E. coli* and human enzymes on the Rv1155–FMN structure suggests that in the Rv1155–FMN structure the steric hindrance of the indole group of Trp77*B* protruding from the other subunit of the dimer prevents the isoalloxazine ring of FMN from adopting a conformation similar to that observed in the *E. coli* or human enzyme. As a result of this, in Rv1155 the indole group of Trp77*B* makes stacking interactions with the dimethylbenzyl ring of the flavin moiety. Arg55, which is involved in FMN and PLP binding, is conserved across the three organisms (Fig. 8).

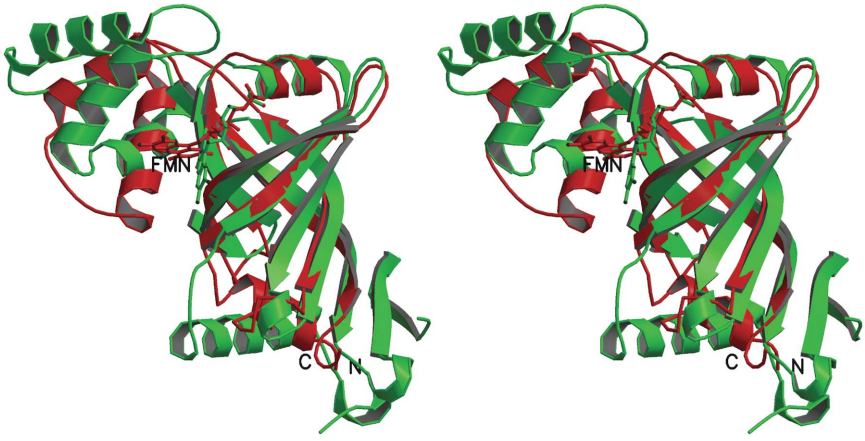

Figure 7
Stereoview of the superimposition of the structures of Rv1155 and *E. coli* PNPOx enzyme. The bound FMN is also shown in ball-and-stick representation. The red and green colours correspond to the Rv1155 and *E. coli* PNPOx enzyme, respectively.

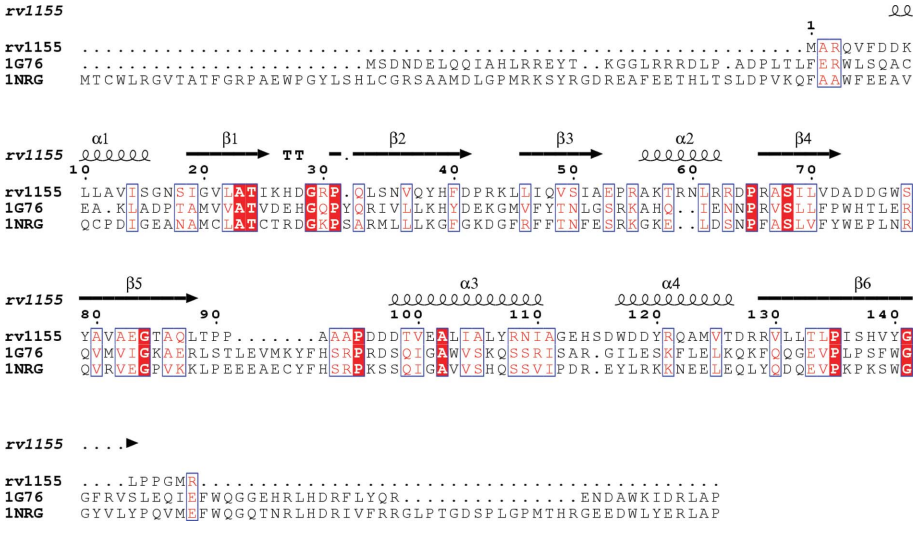

Figure 8
Sequence alignments between the *E. coli*, human and Rv1155 PNPOx enzymes. 1G76 and 1NRG correspond to the *E. coli* and human PNPOx sequences, respectively. Identical residues are shown in red boxes. The secondary-structural elements corresponding to the Rv1155 structure are also shown. Sequence alignment was performed using the program *CLUSTALW* (Thompson *et al.*, 1994) and the figure was generated using the program *ESPRIPT* (Gouet *et al.*, 1999).

4. Conclusions

In *E. coli*, the final step of vitamin B_6 biosynthesis is the oxidation of pyridoxine 5′-phosphate (PNP) to PLP; it is catalysed by PNPOx. Failure to produce a sufficient supply of PLP in mammals usually results initially in neurological disorders (Safo *et al.*, 2000). It is postulated that the enzyme from both *E. coli* and eukaryotic sources catalyzes the oxidation of either the C4′ hydroxyl group or the

amino group of the two substrates pyridoxine 5′-phosphate (PNP) and pyridoxamine 5′-phosphate (PMP) to an aldehyde, thereby forming PLP (Zhao & Winkler, 1995; Safo *et al.*, 2000, 2001; di Salvo *et al.*, 2002, 2003; Musayev *et al.*, 2003). A H atom is removed from C4′ during the oxidation and a pair of electrons are transferred to the tightly bound FMN molecule (Safo *et al.*, 2000, 2001; di Salvo *et al.*, 2002, 2003; Musayev *et al.*, 2003).

Based on the above structural results, Rv1155 may also be involved in the vitamin B_6 biosynthesis pathway in *M. tuberculosis*. Although our structural studies have pointed strongly to the biological function of the gene product from ORF Rv1155, understanding the details of its mechanisms of action needs more definitive structural, biochemical and biophysical experiments. Furthermore, since the three-dimensional structure and FMN-binding site of Rv1155 are very similar to those of the *E. coli* and human PNPOx enzymes, it appears that Rv1155 may not make a good drug target, although this requires more rational studies. In summary, we have predicted a biological function of Rv1155 from its three-dimensional structure. Our study clearly demonstrates that the biological functions of proteins of unknown functions can be determined in light of their three-dimensional structural similarity to proteins of known function.

We thank the Alberta Heritage Foundation for Medical Research (AHFMR) for funding the purchase of the area-detector facilities and the Canadian Institutes of Health Research (CIHR) for operating funds. X-ray diffraction data were collected at beamline 8.3.1 of the Advanced Light Source (ALS) at Lawrence Berkeley Laboratory, Berkeley, CA, USA under an agreement with the Alberta Synchrotron Institute (ASI). The ALS is operated by the Department of Energy and supported by the National Institutes of Health. Beamline 8.3.1 was funded by the National Science Foundation, the University of California and Henry Wheeler. The ASI synchrotron access program is supported by grants from the Alberta Science and Research Authority (ASRA), the Alberta Heritage Foundation for Medical Research (AHFMR) and Western Economic Diversification (WED). MNGJ is grateful for a Canada Research Chair in Protein Structure and Function.

References

Ahmed, N. & Hasnain, S. E. (2004). *Indian J. Med. Res.* **120**, 207–212.
Anderson, A. C., O'Neil, R. H., DeLano, W. L. & Stroud, R. M. (1999). *Biochemistry*, **38**, 13829–13836.
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.
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.
Camus, J. C., Pryor, M. J., Medigue, C. & Cole, S. T. (2002). *Microbiology*, **148**, 2967–2973.
Canaan, S., Sulzenbacher, G., Roig-Zamboni, V., Scappuccini-Calvo, L., Frassinetti, F., Maurin, D., Cambillau, C. & Bourne, Y. (2005). *FEBS Lett.* **579**, 215–221.
Cohen, G. H. (1997). *J. Appl. Cryst.* **30**, 1160–1161.
Coker, R. J. (2004). *Trop. Med. Int. Health*, **9**, 25–40.
Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.
Corbett, E. L., Watt, C. J., Walker, N., Maher, D., Williams, B. G., Raviglione, M. C. & Dye, C. (2003). *Arch. Intern. Med.* **163**, 1009–1021.
Cosma, C. L., Sherman, D. R. & Ramakrishnan, L. (2003). *Annu. Rev. Microbiol.* **57**, 641–676.
Dye, C., Scheele, S., Dolin, P., Pathania, V. & Raviglione, M. C. (1999). *J. Am. Med. Assoc.* **282**, 677–686.
Esnouf, R. M. (1999). *Acta Cryst.* D**55**, 938–940.
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. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Goulding, C. W. *et al.* (2002). *Curr. Drug Targets Infect. Disord.* **2**, 121–141.
Goulding, C. W., Perry, L. J., Anderson, D., Sawaya, M. R., Cascio, D., Apostol, M. I., Chan, S., Parseghian, A., Wang, S. S., Wu, Y., Cassano, V., Gill, H. S. & Eisenberg, D. (2003). *Biophys. Chem.* **105**, 361–370.
Holm, L. & Sander, C. (1998). *Nucleic Acids Res.* **26**, 316–319.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Laskowski, R. J., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Marchler-Bauer, A. & Bryant, S. H. (2004). *Nucleic Acids Res.* **32**, W327–W331.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Merritt, E. A. & Murphy, M. E. (1994). *Acta Cryst.* D**50**, 869–873.
Monack, D. M., Mueller, A. & Falkow, S. (2004). *Nature Rev. Microbiol.* **2**, 747–765.
Musayev, F. N., di Salvo, M. L., Ko, T.-P., Schirch, V. & Safo, M. K. (2003). *Protein Sci.* **12**, 1455–1463.
Ojcius, D. (2004). *Nature Rev. Microbiol.* **2**, 858.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Philipp, W. J., Poulet, S., Eiglmeier, K., Pascopella, L., Balasubramanian, V., Heym, B., Bergh, S., Bloom, B. R., Jacobs, W. R. Jr & Cole, S. T. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 3132–3137.
Read, R. J. & James, M. N. G. (1988). *J. Mol. Biol.* **200**, 523–551.
Rupp, B. (2003). *Acc. Chem. Res.* **36**, 173–181.
Rupp, B., Segelke, B. W., Krupka, H. I., Lekin, T., Schafer, J., Zemla, A., Toppani, D., Snell, G. & Earnest, T. (2002). *Acta Cryst.* D**58**, 1514–1518.
Safo, M. K., Mathews, I., Musayev, F. N., di Salvo, M. L., Thiel, D. J., Abraham, D. J. & Schirch, V. (2000). *Structure Fold. Des.* **8**, 751–762.
Safo, M. K., Musayev, F. N., di Salvo, M. L. & Schirch, V. (2001). *J. Mol. Biol.* **310**, 817–826.
Salvo, M. L. di, Ko, T. P., Musayev, F. N., Raboni, S., Schirch, V. & Safo, M. K. (2002). *J. Mol. Biol.* **315**, 385–397.
Salvo, M. L. di, Safo, M. K., Musayev, F. N., Bossa, F. & Schirch, V. (2003). *Biochim. Biophys. Acta*, **1647**, 76–82.
Smith, C. V. & Sacchettini, J. C. (2003). *Curr. Opin. Struct. Biol.* **13**, 658–664.
Stura, E. A. & Wilson, I. A. (1991). In *Crystallization of Nucleic Acids and Proteins: A Practical Approach*, edited by A. Ducruix & R. Giegé. Oxford: IRL Press.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
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.
Toossi, Z. (2003). *J. Infect. Dis.* **188**, 1146–1155.
Tripathi, R. P., Tewari, N., Dwivedi, N. & Tiwari, V. K. (2005). *Med. Res. Rev.* **25**, 93–131.
Zhao, G. & Winkler, M. E. (1995). *J. Bacteriol.* **177**, 883–891.