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

The molecular structure of Rv2074, a probable pyridoxine 5′-phosphate oxidase from *Mycobacterium tuberculosis*, at 1.6 Å resolution

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

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

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

Received 26 April 2006
Accepted 28 June 2006

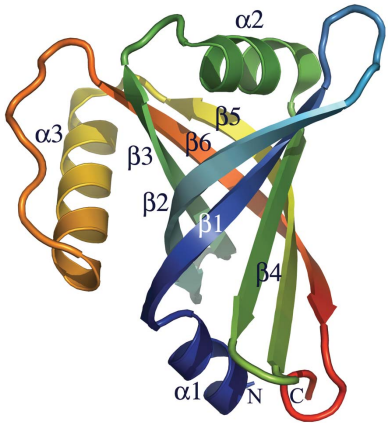
PDB Reference: Rv2074, 2asf, r2asfsf.

The crystal structure of a conserved hypothetical protein corresponding to open reading frame Rv2074 from *Mycobacterium tuberculosis* (*Mtb*) has been solved by the two-wavelength anomalous dispersion method. Refinement of the molecular structure at 1.6 Å resolution resulted in an R_{work} of 0.178 and an R_{free} of 0.204. The crystal asymmetric unit contains an Rv2074 monomer; however, the crystallographic twofold symmetry operation of space group $P4_32_12$ generates dimeric Rv2074. Each monomer folds into a six-stranded antiparallel β-barrel flanked by two α-helices. The three-dimensional structure of Rv2074 is very similar to that of *Mtb* Rv1155, a probable pyridoxine 5′-phosphate oxidase (PNPOx), which corroborates well with the relatively high sequence similarity (52%) between the two. A structural comparison between Rv2074 and Rv1155 revealed that the core structure (a six-stranded β-barrel) is also well conserved; the major differences between the two lie in the N- and C-termini and in the small helical domain. Two citric acid molecules were observed in the active site of Rv2074, the crystals of which were grown in 0.2 *M* sodium citrate buffer pH 5.0. The citric acid molecules are bound to Rv2074 by hydrogen-bonding interactions with Thr55, Gln60 and Lys61. One of the two citric acid molecules occupies the same spatial position that corresponds to the position of the phosphate and ribose sugar moieties of the flavin mononucleotide (FMN) in the *Mtb* Rv1155–FMN, *Escherichia coli* PNPOx–FMN and human PNPOx–FMN complex structures. Owing to its extensive structural similarity with *Mtb* Rv1155 and to the *E. coli* and human PNPOx enzymes, Rv2074 may be involved in the final step in the biosynthesis of pyridoxal 5′-phosphate (PLP; a vitamin B_6).

1. Introduction

Mycobacterium tuberculosis (*Mtb*) is the etiological agent of the highly contagious disease tuberculosis. Currently, one third of the world′s population (around two billion people) are infected by *Mtb* and around two million deaths occur annually (World Health Organization, 2003). The inadequacy of the current treatments together with the increasing emergence of multidrug resistance and the HIV pandemics emphasize the need to understand the biological functions of virtually every gene product of this pathogenic bacterium. Three-dimensional structure determinations of every gene product will help in a more complete understanding this organism at the molecular level and should facilitate the design of potent inhibitors against the various virulence factors responsible for the survival of *Mtb* in the host cell.

Structural genomics, a nascent and rapidly growing field in biology, aims to determine the three-dimensional structures of all proteins found in nature; the resulting data will provide a broad foundation for a fundamental understanding of biology. The *Mtb* structural genomics consortium was formed in 2000 (http://www.doe-mbi.ucla.edu/TB) and 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). The complete genome sequence of the best-characterized strain of *Mtb*, H37Rv, has provided a wealth of information on the functional aspects of 52% of the genome (Philipp *et al.*, 1996; Cole *et al.*, 1998; Camus *et al.*, 2002); however, the biological functions of many of the



gene products remain to be determined. As a member of the *Mtb* structural genomics consortium and in continuation of our efforts in predicting the biological functions of the hypothetical proteins that are conserved among the *Actinomycetes* subgroup of bacteria from their three-dimensional structures, we have determined the crystal structure of a conserved hypothetical protein, Rv2074, from *Mtb*. The details of the structure determination and insight into a possible biological function of Rv2074 are reported in this paper.

2. Experimental methods

2.1. Cloning, expression and purification of Rv2074

The H37Rv genomic DNA in an ordered BAC library from L'Institut Pasteur (Brosch *et al.*, 1998; Gordon *et al.*, 1999) was used as the template to amplify Rv2074 by PCR using the following primers: forward, **GGGGACAAGTTTGTACAAAAAAGCAGGCTCA-***GAAAACCTGTATTTTCAGGGC*GTGGCGATGGTCAACACC; reverse, **GGGGACCACTTTGTACAAGAAAGCTGGGTC**TCA-GGCCCGGTCGAGCAGATCCG.

The amplified ORF was gel purified (Qiagen) and then inserted directionally into a cloning vector by homologous recombination (recombination sites shown in bold in primers) using the Gateway cloning system (Invitrogen). Rv2074 was then transferred by recombination into an expression vector encoding an amino-terminal GST-fusion protein. A site recognized by recombinant tobacco etch virus (rTEV) protease (Invitrogen) is encoded by the forward primer (italicized) for removal of the fusion tag. The resulting expression plasmid, pGST-2074, was confirmed by DNA-sequence analysis (DNA Sequence Facility, University of Alberta).

Expression and purification of Rv2074 was performed in a very similar manner to that of Rv1155 (Biswal *et al.*, 2005). Briefly, Rv2074 was expressed in *Escherichia coli* BL21(DE3)-pLysS cells (Novagen). An overnight culture of the bacteria transformed with the expression vector in Luria–Bertani (LB) medium 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 until the $OD_{600\,nm}$ reached 0.5–0.7. Subsequently, expression was induced by 1 m*M* isopropyl β-D-thiogalactopyranoside (IPTG). After overnight incubation, the cells were harvested by centrifugation at 2150*g* for 25 min. Bacterial pellets were resuspended in PBS containing EDTA-free Complete protease inhibitor (Roche) and 1 m*M* DTT (Fischer Scientific). For purification, the cells were lysed by sonication, the lysate was cleared by centrifugation (30 000*g*, 30 min) and the supernatant was loaded onto a GSTPrep glutathione Sepharose column (GE Healthcare) equilibrated with PBS. The GST-2074 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. After dialysis in PBS, the cleaved protein mixture was once again loaded onto the GSTPrep column. Flowthrough fractions containing Rv2074 from this column were collected and then concentrated using an Amicon Ultra (5 kDa cutoff; Millipore). The purification was performed at 277 K. Selenomethionine-substituted Rv2074 was produced in a similar way as the native protein but in M9 minimal medium by inhibiting methionine biosynthesis (Doublié, 1997) and then purified. DTT was added to 5 m*M* following concentration of the protein and the preparation was kept in an argon atmosphere to prevent oxidation of the SeMet. 16%

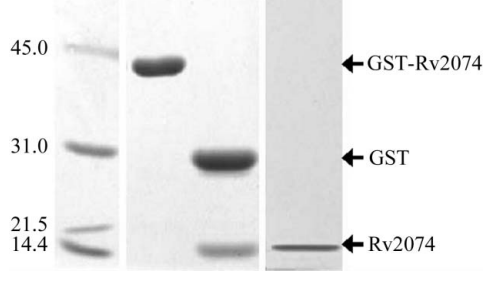


Figure 1
16% SDS–PAGE analysis of purification stages of Rv2074. Lane 1, protein molecular-weight standards (kDa); lane 2, GST-Rv2074 fusion protein; lane 3, fusion protein cleaved with TEV protease; lane 4, purified Rv2074 before concentration.

Table 1
Data-collection and refinement statistics.

Values in parentheses refer to the highest resolution shell.

	Peak	Inflection	High resolution
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$
Unit-cell parameters			
a (Å)	73.125	73.317	73.395
b (Å)	73.125	73.317	73.395
c (Å)	44.936	44.984	44.794
$\alpha = \beta = \gamma$ (°)	90	90	90
Z	8	8	8
Data collection (ALS beamline 8.3.1)			
Temperature (K)	100	100	100
Detector	ADSCQ210	ADSCQ210	ADSCQ210
No. of crystals	1	1	1
Total rotation range (°)	110	108	58
Wavelength (Å)	0.9795	1.0196	1.1158
Resolution (Å)	40.0–2.0	40.0–2.0	40.0–1.6
Highest resolution shell (Å)	2.07–2.00	2.07–2.00	1.66–1.60
Total No. of observations	56760	63919	65558
Unique reflections	8564	8606	16310
Multiplicity	6.6 (4.8)	7.4 (4.9)	4.0 (3.3)
$\langle I/\sigma(I)\rangle$	12.5 (2.7)	15.5 (2.8)	19.8 (2.6)
Completeness (%)	98.5 (92.6)	98.7 (88.6)	97.4 (87.8)
R_{sym}† (%)	13.5 (45.7)	11.4 (40.5)	5.4 (39.8)
$R_{r.i.m.}$‡	14.7 (51.4)	12.3 (45.4)	6.2 (47.7)
$R_{p.i.m.}$‡	5.7 (23.4)	4.5 (20.5)	3.1 (26.2)
Overall B factor from Wilson plot (Å^2)	14.9	15.3	14.4
Refinement			
Refinement resolution (Å)	—	—	40–1.6
Unique reflections (working/test)	—	—	15503/792
R_{work}§	—	—	0.178 (0.244)
R_{free}§	—	—	0.204 (0.328)
No. of atoms			
Protein	—	—	996
Water	—	—	175
Citrate	—	—	2
Sodium			1
Average B factor (Å^2)			
All atoms	—	—	14.5
Protein atoms	—	—	12.1
Solvent atoms	—	—	28.1
R.m.s. deviations from ideal			
Bond lengths (Å)	—	—	0.021
Bond angle (°)	—	—	1.3
Percentage of residues in Ramachandran plot			
In most favored regions	—	—	94.6
In additional allowed regions	—	—	5.4

† $R_{sym} = \sum_{\mathbf{h}}\sum_i[|I_i(\mathbf{h}) - \langle I(\mathbf{h})\rangle|]/\sum_{\mathbf{h}}\sum_i I_i(\mathbf{h})$, where $I_i(\mathbf{h})$ is the ith intensity measurement and $\langle I(\mathbf{h})\rangle$ is the weighted mean of all measurements of $I_i(\mathbf{h})$. ‡ $R_{r.i.m.}$ (redundancy-independent merging R factor) = $\sum_{\mathbf{h}}[N/(N-1)]^{1/2}\sum_i[|I_i(\mathbf{h}) - \langle I(\mathbf{h})\rangle|]/\sum_{\mathbf{h}}\sum_i I_i(\mathbf{h})$ and $R_{p.i.m.}$ (precision-indicating merging R factor) = $\sum_{\mathbf{h}}[1/(N-1)]^{1/2}\sum_i[|I_i(\mathbf{h}) - \langle I(\mathbf{h})\rangle|]/\sum_{\mathbf{h}}\sum_i I_i(\mathbf{h})$), where N is the redundancy (Weiss, 2001). § R_{work} 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 the observed and calculated structure-factor amplitudes, respectively. R_{free} was calculated using 5% of data.

SDS–PAGE analysis of the purification stages revealed a single band of ∼15 kDa (Fig. 1), consistent with the molecular weight of Rv2074 (14 992 Da). There was no remnant from the tag attached to Rv2074.

2.2. Crystallization

The initial screening of various crystallization conditions for the selenomethionine-substituted Rv2074 was performed at room temperature using the sitting-drop vapour-diffusion technique in 96-well Intelli-plates (Hampton Research). Rv2074 was concentrated to 12 mg ml^{-1} and dialyzed into water. High-Throughput and Index Crystal Screens (Hampton Research) were applied using equal volumes (0.5 µl) of protein and precipitating solution. From these screens, initial crystals were obtained from a variety of conditions. After optimization of the best screen condition, X-ray diffraction-quality crystals were grown in hanging drops containing 1 µl protein solution at 12 mg ml^{-1} and 0.5 µl precipitating solution and the drops were equilibrated against 1 ml precipitating solution (20% PEG 4000, 0.2 *M* sodium citrate buffer pH 5.0 and 3% dioxane). Crystals grew to dimensions of 0.4 × 0.2 × 0.15 mm within 1–2 weeks.

2.3. Data collection

For data collection, crystals were flash-cooled to 100 K in a cryo-protectant containing 30% glycerol, 20% PEG 4000 and 0.2 *M* sodium citrate pH 5.0. Crystals belong to the tetragonal space group $P4_32_12$, with unit-cell parameters $a = b = 73.395$, $c = 44.794$ Å (Table 1). The crystal asymmetric unit contains one molecule of Rv2074; the Matthews coefficient V_M was 2.0 Å^3 Da^{-1} and the solvent content was 38.4% (Matthews, 1968). For structure determination, multiwavelength anomalous dispersion data were collected at the peak and inflection points of the anomalous signal at beamline 8.3.1 at the Advanced Light Source, Berkeley, CA, USA. Data sets were integrated and scaled using the programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). Details of the data-collection statistics are given in Table 1.

2.4. Structure solution and refinement

There are two selenomethionine-substituted residues situated at Met3 and Met25 in the primary structure of Rv2074. Only the position of the Se atom corresponding to SeMet25 was found in the crystal asymmetric unit; it was used to calculate the initial multi-wavelength anomalous dispersion phases and for auto model building of ∼70% of the total amino acids in Rv2074 (programs *SOLVE* and *RESOLVE*; Terwilliger & Berendzen, 1999; Terwilliger, 2000). The resultant partial model was subjected to 100 cycles of rigid-body refinement by the program *CNS* v.1.1 (Brünger *et al.*, 1998) and this model was then refined against the high-resolution (1.6 Å) data. The remainder of the model was completed based on $2|F_o| - |F_c|$ and $|F_o| - |F_c|$, α_c electron-density maps using the program *XtalView* (McRee, 1999). Structure refinement was continued using the program *CNS* v.1.1 with a maximum-likelihood target function. Initially, the completed model was subjected to 3000 K simulated-annealing refinement; subsequently, positional and individual *B*-factor refinements were performed. Once the refinement had converged to an R_{work} value of 0.21, 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) electron-density maps. Refinements after every stage of water identification continued until no significant electron-density peaks ($2|F_o| - |F_c|$ at 1σ and $|F_o| - |F_c|$ at 3σ contour levels) remained in the electron-density maps. The final refinement of the model was performed using the program *REFMAC* v.5.2.0 (Collaborative Computational Project, Number 4, 1994). The first ten residues from the N-terminus and last two residues from the C-terminus were not observed in the electron-density maps. The stereochemical acceptability of the structure was examined by the program *PROCHECK* (Laskowski *et al.*, 1993). Details of the refinement statistics are given in Table 1.

3. Results

3.1. Molecular structure of Rv2074

Rv2074 contains 137 amino acids and folds into a six-stranded antiparallel β-barrel flanked by two α-helices (Fig. 2). The structure is comprised of six β-strands and three α-helices. The secondary-structural elements of Rv2074 relative to the amino-acid sequence are shown in Fig. 3. Two helices [α1 (Asp12–Phe17) and α2 (Gln60–Ser68)] flanking the six-stranded β-barrel are separated by 27.5 Å; this represents the approximate height of the β-barrel. The small structural segment consisting of helix α3 (Ile94–Arg107) and the loop (Tyr108–Arg116; Fig. 2) connecting the strands β5 and β6 is referred to as a small domain to facilitate the subsequent structural compar-

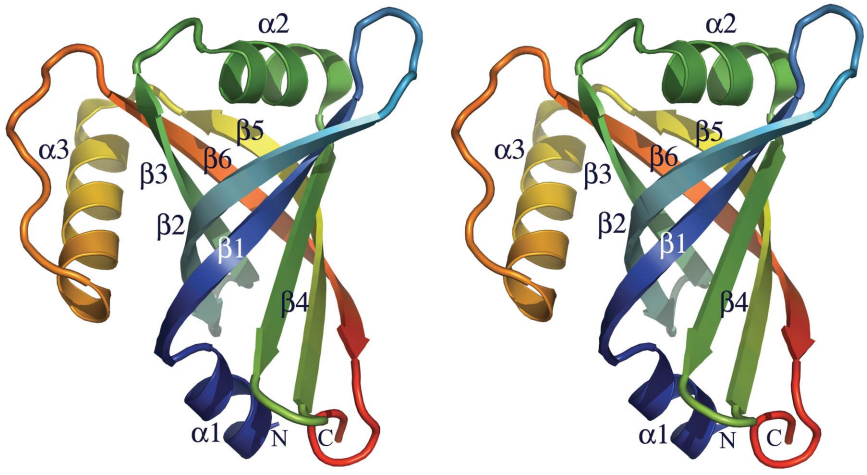


Figure 2
Stereoview of the six-stranded antiparallel β-barrel structure of Rv2074. This figure and Figs. 5 and 6 were prepared using the program *PyMOL* (http://pymol.sourceforge.net).

ison with homologous structures. The crystal asymmetric unit contains a monomer; however, as was anticipated, the crystallographic twofold symmetry operation ($-y$, $-x$, $\frac{1}{2} - z$) generates dimeric Rv2074. The residues Leu23, Ala24, Met25, Thr27, Leu29, Arg30, Asn33, Pro35, Val37, Val38, Ala39, Leu70, Val72, Ser74, Gly78, Trp81, Ser83, Glu85, Arg126 and Leu128 protruding from the strands $\beta1$, $\beta2$, $\beta4$, $\beta5$ and $\beta6$ of each subunit form the dimer interface. The interactions at the dimer interface are primarily hydrophobic and the dimer interface buries 1800 Å^2 of solvent-accessible surface area, as compared with 6970 Å^2, the total solvent-accessible surface area of each monomer, calculated by the program *CNS* v.1.1 with a probe radius of 1.4 Å. There are 75 main chain–main chain, 24 main chain–side chain and 17 side chain–side chain hydrogen bonds in each monomer of Rv2074 (see supplementary material[1]). Hydrogen bonds were determined using the program *HBPLUS* (McDonald & Thornton, 1994).

3.2. Prediction of a possible biological function for Rv2074 and structural comparisons

To learn more about the possible biological function of Rv2074 based on its three-dimensional structural similarity with proteins of known functions, the atomic coordinates of Rv2074 were submitted to the *DALI* server (Holm & Sander, 1998; EMBL–EBI) to search for structural homologues in the PDB. The best five matches whose definitive or probable biological functions are known were *Mtb* Rv1155 (PDB code 1xxo), an antibiotic resistance protein from *Deinococcus radiodurans* (PDB code 1w3o), an FMN-binding protein from *Desulfovibrio vulgaris* (PDB code 1axj), pyridoxamine 5′-phosphate oxidase (PMPOx) from *Pseudomonas aeruginosa* (PDB code 1t9m) and pyridoxine 5′-phosphate oxidase (PNPOx) from *Saccharomyces cerevisiae* (PDB code 1ci0). Except for the second one, they are all known to bind FMN as a cofactor. PMPOx and PNPOx enzymes catalyse the terminal step in the biosynthesis of pyridoxal 5′-phosphate (PLP), a cofactor that is used by many enzymes involved in amino-acid biosynthesis (Zhao & Winkler, 1995; Safo *et al.*, 2000, 2001; di Salvo *et al.*, 2002, 2003; Musayev *et al.*, 2003). A conserved-domain search based on the primary structure of Rv2074 using the National Center for Biotechnology Information (NCBI) conserved-domain search server (Marchler-Bauer & Bryant, 2004) also predicts that Rv2074 resembles the human PNPOx enzyme.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: EN5175).

The three-dimensional structure of Rv2074 is very similar to that of Rv1155, which is also a probable PNPOx (Biswal *et al.*, 2005; Canaan *et al.*, 2005), which correlates well with the relatively high sequence similarity (52%) between the two (Figs. 3 and 4*a*). The root-mean-square deviation (r.m.s.d.) between the overall dimeric structures is 1.97 Å (215 C$^\alpha$ atoms with a maximum deviation of 5.32 Å). The core six-stranded β-barrel structures of Rv2074 and Rv1155 agree much more closely (r.m.s.d. 1.61 Å for 160 C$^\alpha$ atoms; Fig. 4*a*). Furthermore, superimposition of the structures of Rv2074 and human and *E. coli* PNPOx enzymes revealed that the core six-stranded β-barrel structure is also conserved (Fig. 4*b*). Major differences, however, are found in the N- and C-terminal regions and in the small domain. The small domains are a three-helix bundle, a helix–loop–helix motif and a helix and a loop in the human and *E. coli* PNPOx enzymes, Rv1155 and Rv2074 structures, respectively. Understandably, the structural difference in this region can be attributed to the difference in the length of their primary structures. The human and *E. coli* PNPOx enzymes, Rv1155 and Rv2074 are comprised of 261, 218, 147 and 137 amino acids, respectively. This suggests that the small domain may not be an essential structural element of the enzyme. Structural superpositions were carried out using the program *ALIGN* (Cohen, 1997).

Recently, the crystal structure of a putative PNPOx from *Mtb*, Rv2607, has been reported (Pedelacq *et al.*, 2006). Like the three-dimensional structures of Rv1155 and Rv2074, Rv2607 is also a six-stranded antiparallel β-barrel flanked by two α-helices (Pedelacq *et al.*, 2006). Furthermore, the core structure is conserved among the three (Fig. 4*c*). This suggests that the *Mtb* genome may contain multiple genes that code for enzymes with a similar function. The FMN-binding site is conserved among the Rv2074, Rv1155, Rv2607, human and *E. coli* PNPOx enzymes (Fig. 5). Understandably, the essential catalytic residue Lys61 is conserved among them. On the other hand, the PLP-binding site does not seem to be very well conserved.

3.3. Fold similarity with serine proteases

As mentioned earlier, the six-stranded antiparallel β-barrel structure of Rv2074 is similar to those of *Mtb* Rv1155 and Rv2607 as well as to the human and the *E. coli* PNPOx enzymes. A similar fold is also observed in viral, bacterial and mammalian serine proteinases (Read & James, 1988; Fujinaga *et al.*, 1995; Kim *et al.*, 1996), suggesting a common pathway of evolution of the oxidases and proteinases (Fig. 6). However, the domain organization to form the biological entity is significantly different between the oxidases and proteases and is clearly associated with their different biological functions. In

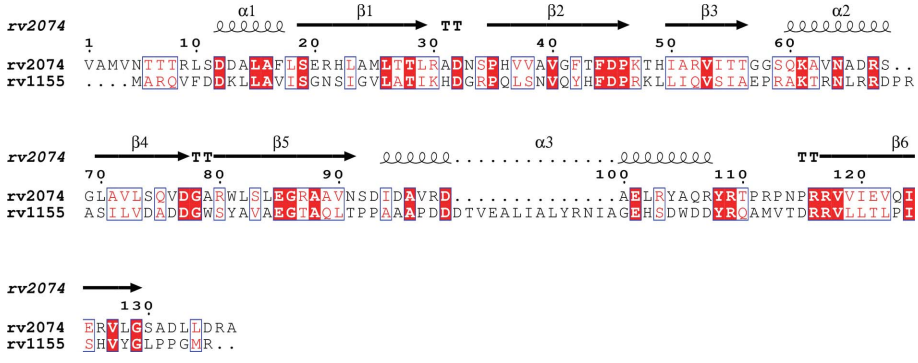


Figure 3
Sequence alignment between Rv2074 and Rv1155. The secondary-structural elements (helices represented by coils and strands by arrows) corresponding to the Rv2074 structure are shown. The numbering above the alignment corresponds to the Rv2074 sequence. Sequence alignment was performed using the program *ClustalW* (Thompson *et al.*, 1994) and the figure was generated by the program *ESpript* (Gouet *et al.*, 1999).

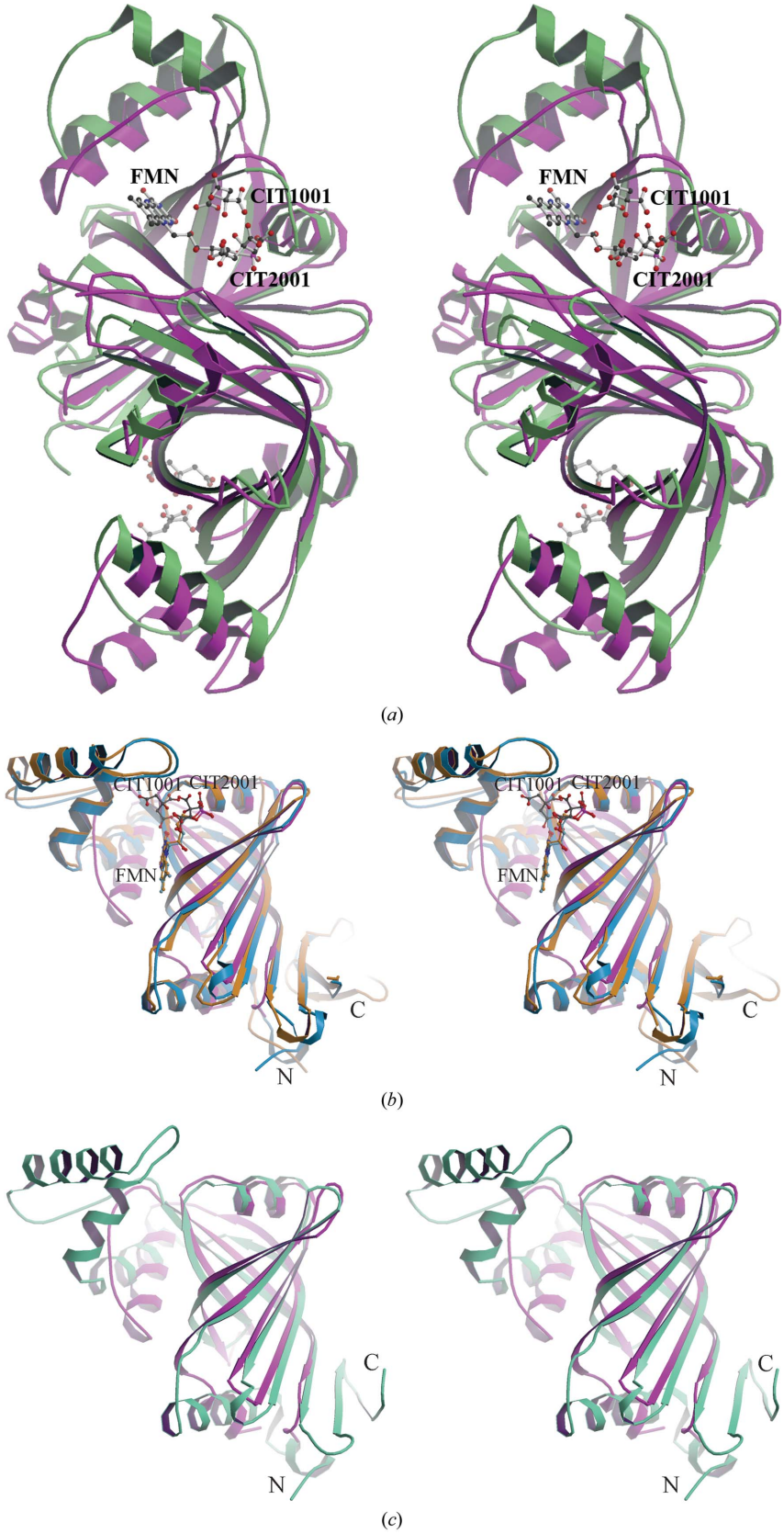


Figure 4
(*a*) Stereoview of the superimposition of the dimeric structures of Rv2074–CIT (magenta) and Rv1155–FMN (light green; PDB code 1y30). FMN and citric acid molecules are shown in ball-and-stick representation. This figure and Figs. 4(*b*), 4(*c*) and 7(*b*) were prepared using the programs *MOLSCRIPT* (Kraulis, 1991) and *RASTER3D* (Merritt & Murphy, 1994). (*b*) Stereoview of superimpositions of the monomeric structures of Rv2074–CIT (magenta), human PNPOx–FMN (orange; PDB code 1nrg) and *E. coli* PNPOx–FMN (cyan; PDB code 1dnl). FMN and citrate molecules in the active site are shown in ball-and-stick representation. (*c*) Stereoview of superimposition of the monomeric structures of Rv2074 (magenta) and Rv2607 (aquamarine; PDB code 2a2j), showing the conserved β-barrel.

PNPOx enzymes (Safo *et al.*, 2000, 2001; di Salvo *et al.*, 2002, 2003; Musayev *et al.*, 2003; Biswal *et al.*, 2005; Canaan *et al.*, 2005; Pedelacq *et al.*, 2006) the domains of the dimer are related by twofold symmetry, whereas the two domains of the serine proteases are related approximately by a translation followed by a 90° rotation (Read & James, 1988). The active site of the clan SA serine proteinases is found in a groove between the two β-barrel domains.

3.4. Citric acid molecules in the possible active site

During the structure refinement of Rv2074, $2|F_o| - |F_c|$ (1σ contour level) as well as $|F_o| - |F_c|$ (3σ contour level) electron-density peaks were observed in the possible active-site region of Rv2074. Interpretation of the electron-density peaks was initially difficult. Careful examination of the electron-density peaks, together with knowledge of the crystallization conditions, however, led us to a novel conclusion. The observed electron-density peaks correspond to two closely associated bound citric acid molecules (Fig. 7*a*). Crystals of Rv2074 were grown in the presence of 0.2 *M* sodium citrate buffer pH 5.0. Atom O3 of CIT1001 is hydrogen bonded to Thr55 OG1 and Lys61 NZ of Rv2074. Gln60 N and Lys61 NZ of Rv2074 are involved in hydrogen-bond formation with the O1 and O2 atoms of the CIT2001 ion (Fig. 7*b*). A constellation of water molecules, providing hydrogen-bonding interactions with the two citrate molecules, is observed in the active site. CIT2001 in Rv2074 occupies the corresponding spatial position of the phosphate and ribose sugar moieties of the flavin mononucleotide bound to *Mtb* Rv1155 and human and *E. coli* PNPOx enzymes (Figs. 4*a* and 4*b*).

A sodium ion coordinating with the O1 and O7 atoms of CIT1001, with the O3 atom of CIT2001 and with two symmetry-related water molecules W56 and W127 was also observed in the electron-density maps (Figs. 7*a* and 7*b*). The mean length of the these coordination bonds is 2.34 Å, which agrees well with the observed mean Na···O distance of 2.42 Å (Harding, 2002). Although sodium is generally observed to have six coordinating ligands, the sixth coordinating atom was not observed in the Rv2074 structure, presumably because of crystal packing.

Rv2074 crystals grown in the presence of 0.2 *M* sodium citrate buffer were soaked in 5 m*M* FMN or in 5 m*M* PLP in crystallization buffer at room temperature for 10–12 h, anticipating that FMN or PLP might be able to displace the citric acid molecules. The electron-density maps obtained from the data collected from the FMN-soaked crystal revealed that only one citric acid molecule (CIT2001) can be placed in the electron-density peaks that appeared in the active site. However, there were no clear electron-density peaks that would correspond to an FMN molecule. On the other hand, the electron-density peaks in the PLP-soaked crystals were identical to those observed in the native crystal.

A literature survey suggests that oxidoreductase enzymes bind citric acid in their active site. In particular, human aldose reductase

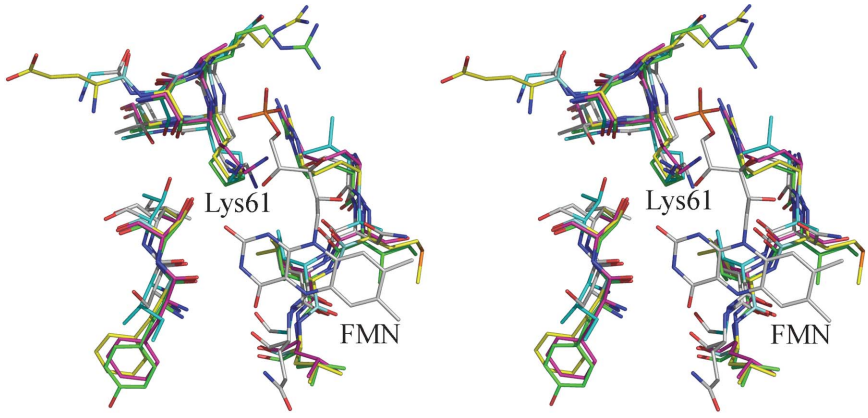


Figure 5
Stereoview of superimposition of the FMN-binding site. The C atoms of Rv2074, Rv1155, Rv2607 and human and *E. coli* PNPOx enzymes are shown in cyan, grey, magenta, yellow and green, respectively. The FMN molecule of the *E. coli* PNPOx–FMN structure is shown in a stick representation with C atoms in grey.

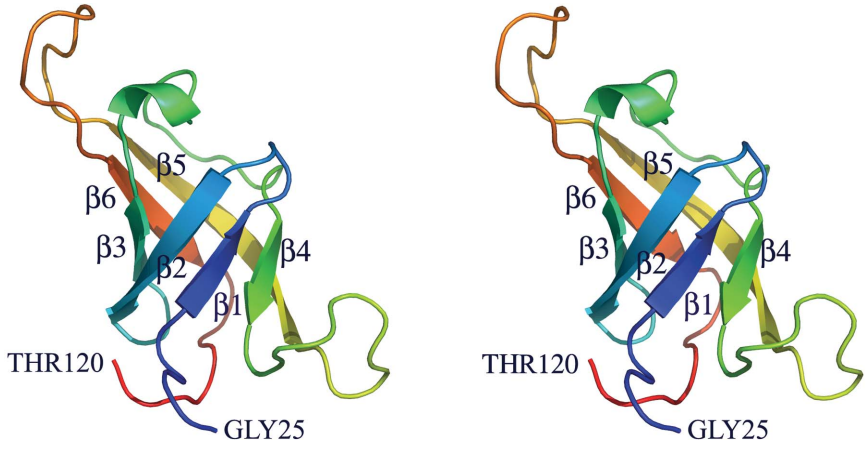


Figure 6
Stereoview of the six-stranded antiparallel β-barrel chymotrypsin fold. This figure was prepared from the crystal structure of SGT, a *Streptomyces griseus* serine proteinase (Read & James, 1988; PDB code 1sgt).

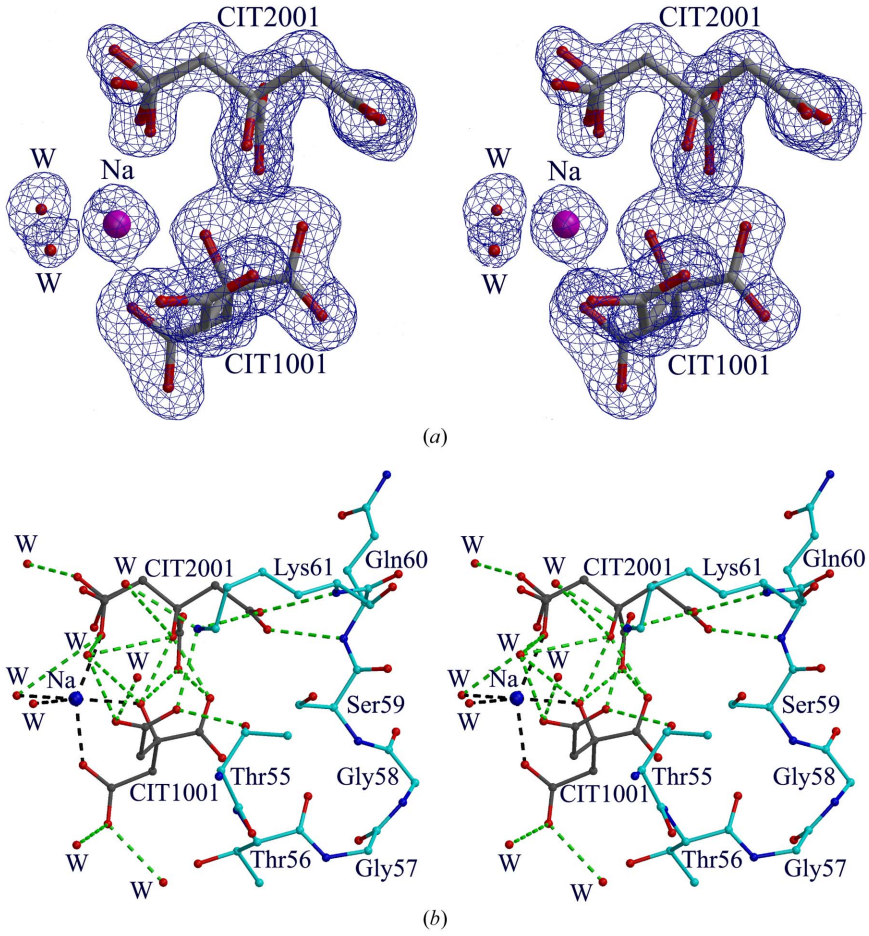


Figure 7
(*a*) Stereoview of the $2|F_o| - |F_c|$ electron-density map (blue) contoured at the 1σ level, showing two citric acid molecules bound in the active site of Rv2074. The citrate C and O atoms are shown in grey and red, respectively. Also shown are the Na atom and coordinating water molecules. This figure was prepared by the program *BOBSCRIPT* (Esnouf, 1999). (*b*) Stereoview of the interactions between the citric acid molecules and Rv2074. The C atoms of citric acid molecules and Rv2074 are shown in grey and cyan, respectively. Hydrogen and coordination bonds are shown in green and black, respectively.

binds citric acid in the active site when crystallized in the presence of 50 m*M* ammonimum citrate buffer at pH 5.0 (Harrison *et al.*, 1997; El-Kabbani *et al.*, 2004). This suggests that in order to obtain more definitive knowledge on the biological function of Rv2074, further experiments crystallizing Rv2074 in the absence of citrate buffer will need to be carried out.

4. Conclusion

The structure of Rv2074 resembles a six-stranded antiparallel β-barrel and is similar to those of *Mtb* Rv1155 and Rv2607 and the human and *E. coli* PNPOx enzymes. In light of its three-dimensional structural similarity to PNPOx enzymes, Rv2074 may also be involved in the terminal step in the biosynthesis of PLP, a vitamin B_6. As with Rv2074, binding of citric acid molecules in the active sites of other oxidoreductases has also been observed. Efforts towards the successful crystallization of Rv2074 in the absence of citrate buffer are being made.

We thank Ernst Bergmann for his help during data collection. MNGJ is grateful for a Canada Research Chair in Protein Structure and Function. X-ray 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 is 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).

References

Biswal, B. K., Cherney, M. M., Wang, M., Garen, C. & James, M. N. G. (2005). *Acta Cryst.* D**61**, 1492–1499.
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.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cohen, G. H. (1997). *J. Appl. Cryst.* **30**, 1160–1161.
Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.

di Salvo, M. L., Ko, T.-P., Musayev, F. N., Raboni, S., Schirch, V. & Safo, M. K. (2002). *J. Mol. Biol.* **315**, 385–397.
di Salvo, M. L., Safo, M. K., Musayev, F. N., Bossa, F. & Schirch, V. (2003). *Biochim. Biophys. Acta*, **1647**, 76–82.
Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.
El-Kabbani, O., Darmanin, C., Oka, M., Schulze-Briese, C., Tomizaki, T., Hazemann, I., Mitschler, A. & Podjarny, A. (2004). *J. Med. Chem.* **47**, 4530–4537.
Esnouf, R. M. (1999). *Acta Cryst.* D**55**, 938–940.
Fujinaga, M., Chernaia, M. M., Tarasova, N. I., Mosimann, S. C. & James, M. N. G. (1995). *Protein Sci.* **4**, 960–972.
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.
Harding, M. M. (2002). *Acta Cryst.* D**58**, 872–874.
Harrison, D. H., Bohren, K. M., Petsko, G. A., Ringe, D. & Gabbay, K. H. (1997). *Biochemistry*, **36**, 16134–16140.
Holm, L. & Sander, C. (1998). *Nucleic Acids Res.* **26**, 316–319.
Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A., Chambers, S. P., Markland, W., Lepre, C. A., O'Malley, E. T., Harbeson, S. L., Rice, C. M., Murcko, M. A., Caron, P. R. & Thomson, J. A. (1996). *Cell*, **87**, 343–355.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
McDonald, I. K. & Thornton, J. M. (1994). *J. Mol. Biol.* **238**, 777–793.
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.
Musayev, F. N., di Salvo, M. L., Ko, T.-P., Schirch, V. & Safo, M. K. (2003). *Protein Sci.* **12**, 1455–1463.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pedelacq, J. D., Rho, B. S., Kim, C. Y., Waldo, G. S., Lekin, T. P., Segelke, B. W., Rupp, B., Hung, L. W., Kim, S. I. & Terwilliger, T. C. (2006). *Proteins*, **62**, 563–569.
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.
Smith, C. V. & Sacchettini, J. C. (2003). *Curr. Opin. Struct. Biol.* **13**, 658–664.
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.
Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.
World Health Organization (2003). *The World Health Organization Global Tuberculosis Program.* http://www.who.int/gtb/.
Zhao, G. & Winkler, M. E. (1995). *J. Bacteriol.* **177**, 883–891.