Article:
Benini, Stefano, Haouz, Ahmed, Proux, Florence et al. (2 more authors) (2019) The crystal structure of Rv2991 from Mycobacterium tuberculosis : An F 420 binding protein with unknown function. JOURNAL OF STRUCTURAL BIOLOGY. pp. 216-224. ISSN 1047-8477

https://doi.org/10.1016/j.jsb.2019.03.006

The crystal structure of Rv2991 from *Mycobacterium tuberculosis*: An F_{420} binding protein with unknown function

Stefano Benini[a, *]

stefano.benini@unibz.it

Ahmed Haouz[b]

Florence Proux[b, 1]

Pedro Alzari[c]

Keith Wilson[d]

[a]Bioorganic Chemistry and Bio-Crystallography Laboratory (B_2Cl), Faculty of Science and Technology, Free University of Bolzano, Piazza Università 5, Bolzano 39100, Italy

[b]C2RT-Plateforme de cristallographie, Institut Pasteur, CNRS UMR 3528, 75724 Paris Cedex 15, France

[c]Unité de Microbiologie Structurale, Institut Pasteur, CNRS UMR 3528, Université Paris Diderot, Sorbonne Paris Cité, 75724 Paris Cedex 15, France

[d] York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5DD, UK

*Corresponding author.

[1]Present address: Institut de Biologie (IBENS), Section Génomique Fonctionnelle, École normale supérieure, 46 Rue d'Ulm, Paris, 75005, France.

Abstract

The crystal structure of the conserved hypothetical protein Rv2991 from *Mycobacterium tuberculosis* has been solved by SAD using seleno-methionine substituted protein. The dimeric biological assembly and the sequence and fold conservation are typical of F_{420} cofactor binding enzymes. Despite Rv2991 still being of unknown function, sequence and structural comparison with similar proteins enable a role to be proposed for its C-terminal stretch of residues in recognizing and orienting the substrate. In addition, the C-terminus is involved in both protein folding and determining the size of the active site cavity.

Keywords: Structural genomics; *Mycobacterium tuberculosis*; F_{420}; Unknown function; FDOR

1 Introduction

A wide range of redox reactions is catalyzed by enzymes containing flavin cofactors such as flavin mononucleotide, flavin adenine dinucleotide and the deazaflavin cofactor F_{420} (Hemmerich et al., 1970; Daniels et al., 1985; Daniels et al., 1985). While the F_{420} cofactor is not present in humans, F_{420} binding proteins are particularly abundant in the pathogenic bacterium *Mycobacterium tuberculosis*. F_{420} binding proteins are absent in the human gut flora (Selengut and Haft, 2010), and since many still do not have a function assigned, they present an interesting area of research as potential drug targets. For instance, the F_{420} binding deazaflavin-dependent nitroreductase Rv3547 was shown to activate the bicyclic nitroimidazole pro-drug candidate PA-824 by promoting release of toxic NO into nonreplicating *M. tuberculosis* cells (Singh et al., 2008). The crystal structure of Rv3547 (Cellitti et al., 2012), has underlined the importance of investigating F_{420} binding enzymes to identify their substrates and mode of action in order to design new drugs to control tuberculosis.

F_{420} dependent split β-barrel enzymes have been classified as part of the large superfamily of flavin/deazaflavin oxidoreductases (FDORs), which includes proteins able to bind flavin mononucleotide (FMN), flavin adenine dinucleotide (FAD), and haem cofactors (Ahmed et al., 2015). Using sequence similarity networks, Ahmed et al., further divided the two previously determined major clades FDOR-A and FDOR-B (Taylor et al., 2010), into 22 subclades (FDORs A1-A4, AA1-AA6, B1-B12) with Rv2991 belonging to the FDOR-B1 group (Ahmed et al., 2015). In the same work the affinity of Rv2991 for the F_{420} cofactor was determined with a K_d value of $7.8 \pm 0.9\,\mu M$ (with a K_d for FMN of $27.9 \pm 4.3\,\mu M$ and for FAD of $26.4 \pm 1.4\,\mu M$) confirming its cofactor specificity.

Rv2991 was selected as a target of the EU structural genomics project X-TB for its putative involvement in antibiotic resistance. Rv2991 is homologous to reductases in *Bacteroides fragilis* that confer resistance to 5-nitroimidazole antibiotics by reducing their nitro group to a nontoxic amine (Carlier et al., 1997; Manjunatha et al., 2006; Manjunatha et al., 2006). We solved the structure of Rv2991 by SAD and deposited it in the PDB in 2003. Unfortunately, to date no information is available on the role and function of this enzyme in *M. tuberculosis.* Here, we report the crystal structure, the sequence and structure similarity and the gene conservation amongst *Mycobacterium* spp. of Rv2991.

2 Materials and methods

2.1 Cloning, protein expression and purification

The Rv2991 coding sequence was amplified by a two-step PCR procedure from *M. tuberculosis* H37Rv (Cole et al., 1998) genomic DNA using primers containing the attB sites of the Gateway recombination system (Invitrogen). In the first step, the Rv2991 gene was amplified by PCR using the high fidelity *Pfx* DNA polymerase (Invitrogen), in presence of 5X enhancer solution for GC-rich templates, and both specific forward and reverse primers (Table 1). A DNA sequence encoding the TEV protease cleavage site was attached at the 5′ end of the gene. In the second step, generic forward and reverse primers were used and the first PCR product served as template (Table 1). The final PCR product was visualized on an agarose gel, and purified (PCR purification kit, Qiagen). The expression clone was assembled in the Gateway recombination cloning system using the 'one-tube' protocol following the supplier's instructions, except that all components were used at half of the recommended volume, using pDONR201 as a cloning vector and pDEST17 as a His_6-expression vector (pDEST17-Rv2991). The plasmid was first propagated in *Escherichia coli* DH5α and the correctness of the insert was checked by DNA sequencing. The final protein construct contains a single M to G substitution at the N-terminus. The expression vector pDEST-Rv2991 was then transformed into competent *E. coli* BL21(DE3)pLysS cells. For protein expression, cells were grown in 500 ml of LB medium containing 100 µg/ml ampicillin and 25 µg/ml chloramphenicol. The culture was grown for ≈3.5 h at 30 °C and when the OD_{600} reached 0.8–0.9 induced with 1 mM isopropopyl β-D-thiogalactopyranoside (IPTG). After 1.5 hr of growth, cells were harvested by centrifugation, washed twice in 50 mM Tris-HCl buffer containing 150 mM NaCl and 10 mM imidazole, pH 8.0 (buffer A) and frozen. The cell pellet was thawed, and the cells lysed using a French press at 14,000 psi. After centrifugation, the supernatant was loaded onto a chromatographic column packed with 1 ml Ni-NTA resin (Qiagen) and previously equilibrated with buffer A. A linear gradient was developed from 20 mM imidazole to 500 mM at 0.5 ml min^{-1} flow rate in 20 column volumes, with the Rv2991eluting in a sharp peak at 120 mM imidazole. The fractions containing Rv2991 were pooled and diluted to 0.5 mg/ml with 50 mM Tris-HCl buffer containing 0.1 M NaCl and 1 mM β -mercaptoethanol, pH 8.0. The protein solution was incubated with His_6-tagged rTEV protease (Invitrogen) at a 1:7 protein to rTEV ratio (w/w) for 6 hr at 30 °C. The solution was centrifuged to remove the aggregated cleaved His-tags and the supernatant loaded onto Ni-NTA resin, equilibrated as described above, to remove uncleaved Rv2991 and rTEV. The flow through, containing the tag-free protein was collected, concentrated to 5 ml volume and loaded onto a gel filtration Superdex 75 26/60 (GE Healthcare) previously equilibrated with 20 mM Tris-HCl + 150 mM NaCl pH 8.0 at a flow rate of 1 ml min^{-1}. After gel filtration, the pure protein, which contains a single Met-Gly substitution at the N-terminus as a consequence of the cloning procedure, was concentrated to 10 mg/ml for crystallisation.

Table 1 Gene cloning information.

Source organism	*Mycobacterium tuberculosis* (strain: H37Rv)
DNA source	Cosmid
Forward primer*	CCTGTATTTTCAGGGCGGAACCAAACAGCGCGCC
Reverse primer*	**GTACAAGAAAGCTGGGT**TTACGGGGCGGTCGAGCCACC
Forward primer#	**GGGGACAAGTTTGTACAAAAAAGCAGGCT**CGGAAAACCTGTATTTTCAGGGC
Reverse primer#	**GGGGACCACTTTGTACAAGAAAGAAAGCTGGGT**
Cloning vector	pDONR201
Expression vector	pDEST17-Rv2991
Expression host	*E. coli* BL21(DE3) pLysS cells

The attB recombination sites are in bold, underlined is the rTEV cleavage site.

* denotes primers used in the first PCR step.

# denotes primers used in the second PCR step.

The selenomethionine (SeMet)-labeled protein was produced in BL21(DE3) *E. coli* cells according to published protocols (Van Duyne et al., 1993) and purified as the native Rv2991 following the protocol described above.

2.2 Protein crystallization

Crystallization experiments were performed immediately after protein purification. Vapor diffusion sitting drop crystallization screens were performed at 290 K in 96 well Greiner plates using a CARTESIAN crystallisation robot. The screens used were: Crystal screen 1 and 2 (Hampton Research), Clear strategy screen CSS1 and CSS2 (Molecular Dimensions) and JBS 1 to 8 (Jena Bioscience). Reservoir solutions were 150 µl in volume and crystallization drops were composed of 200 nl of protein and 200 nl of reservoir solution. The crystallization plates were sealed with Crystalclear tape from Hampton Research and stored at 290 K.

To increase the size and the quality of the crystals the initial conditions were reproduced and refined by hand using vapor diffusion hanging drop in Linbro plates. Crystallization drops were made by mixing 2 µl of 12 mg/ml protein solution and 1 µl of the reservoir solution containing 25%(w/v) PEG 4000, 0.1 M tris HCl pH 8.5, 10 mM $MgCl_2$, 200 mM sodium acetate, and 25% (v/v) ethylene glycol. Crystals grew within a week at 290 K with dimensions of 0.1 mm × 0.1 mm × 0.5 mm.

Selenomethionine-substituted protein concentrated to 11.2 mg/ml was crystallized using the hanging drop method by mixing 1 µl of protein with 1 µl of reservoir solution containing 30%(w/v) PEG 4000, 0.1 M Tris HCl pH 8.5, 20 mM magnesium acetate. Drops were equilibrated against 500 µl reservoir solution.

2.3 Data collection and processing

Data were collected for a crystal of the SeMet variant vitrified at 100 K with the mother liquor supplemented with 25% (v/v) of glycerol. As radiation damage was likely, the wavelength was fixed at the *K* edge absorption peak (0.979 Å), in order to maximize the anomalous scattering factors (Se theoretical values: f' 3.84 and f'' −8.32, according to: http://skuld.bmsc.washington.edu/scatter/AS_periodic.html), to collect a SAD dataset (Gonzalez et al., 1999). Data were indexed and integrated using DENZO (Otwinowski and Minor. 1997) and scaled using SCALEPACK (Otwinowski et al., 2003) (Table 2).

Table 2 Data collection, processing and phasing statistics. Values for the outer shell are given in parentheses.

Diffraction source	ESRF beamline ID29
Wavelength (Å)	0.979
Temperature (K)	100
Detector	ADSC quantum 210 CCD
Crystal-detector distance (mm)	230
Rotation range per image (°)	1
Total rotation range (°)	180
Exposure time per image (s)	1
Space group	$P4_322$
a, *b*, *c* (Å)	49.31, 49.31, 132.59
α, β, γ (°)	90, 90, 90
Mosaicity (°)	1.53
Resolution range (Å)	49.39–2.0 (2.03–2.0)
Total No. of reflections	127,485 (2394)
No. of unique reflections	11,425 (921)
Completeness (%)	98.8 (76.7)
Redundancy	11.16 (2.6)
$\langle I/\sigma(I)\rangle$	11.68 (1.23#)

$R_{r.i.m.}$ †	0.186 (1.007)
Overall B factor from Wilson plot (Å^2)	24.4
Phasing statics	46.22–2.30 (Å)
Phasing power (ano)	1.492
R-Cullis for acentric reflexions	0.707
FOM acentric/centric reflexions	0.361/0.043
FOM after DM	0.63

This shell was included as 46% of the reflections had $I/\sigma(I)$ above 2. In the shell 2.15–2.11 the average $I/\sigma(I)$ was 2.25 and in the shell 2.11–2.07 the average $I/\sigma(I)$ was 1.77.

† Estimated $R_{r.i.m.} = R_{merge}[N/(N-1)]^{1/2}$, where N = data multiplicity.

2.4 Structure solution and refinement

The positions of the Se-atoms were determined using SHELXD (Schneider and Sheldrick, 2002) and phase improvement and solvent flattening were carried out using SHARP (Bricogne et al., 2003). Phasing statistics are provided in Table 2. Density modification confirmed the space group to be $P4_322$ and the improved map was used for automatic model building using ARP/wARP (Morris et al., 2004) using Hendrickson Lattman coefficients during the first cycles of refinement as experimental phase information. Model inspection and final manual model building were done with COOT (Emsley and Cowtan, 2004), and refinement with REFMAC5 (Vagin et al., 2004). Structure solution and refined model parameters are provided in Table 3.

Table 3 Refinement and final model statistics Values for the outer shell are given in parentheses.

Resolution range (Å)	49.39–2.00 (2.056–2.004)
Completeness (%)	97.3
No. of reflections, working set	10,829 (611)
No. of reflections, test set	542 (34)
Final R_{cryst}	0.202 (0.23)
Final R_{free}	0.281 (0.301)
No. of non-H atoms	
Protein	1237
Ligand	0
Water	125
Total	1362
R.M.S. deviations	
Bonds (Å)	0.020
Angles (°)	1.689
Average B factors (Å^2)	
Protein	28.7
Water	37.3
Ramachandran plot	

Most favoured (%)	98
Allowed (%)	99.8

2.5 Sequence alignments and structural comparison

The Rv2991 sequence (163 residues, NCBI Reference Sequence: NP_217507.1) was used to search non-redundant sequence databases (limiting the search to 500 entries maximum) using BLASTp (Ye et al., 2006). Of the 500 sequences found, 327 were selected for the uniqueness of the represented species, as many entries related to different strains of the same organism. A final alignment was carried out with COBALT (Papadopoulos and Agarwala, 2007) using 27 selected sequences with decreasing sequence identity to Rv2991 (from 100% of the 163 residues long conserved hypothetical protein from *Mycobacterium canettii* to 64% of the 162 residues long F_{420}-dependent oxidoreductase from *Mycolicibacterium gilvum*). Figures of the alignments were prepared with ESPript 3.0 (http://espript.ibcp.fr) (Robert and Gouet, 2014). BLASTp was then run against the PDB, returning only three structures with limited sequence similarity.

A structural similarity search was carried out with PDBeFOLD (http://www.ebi.ac.uk/msd-srv/ssm) (Krissinel and Henrick, 2004) with default parameters in search of unexpected structural similarity as in (Benini et al., 2013). Protein oligomerization was determined using PISA (Krissinel and Henrick, 2007). Figures representing the structures were prepared with CCP4mg (Potterton et al., 2004,McNicholas et al., 2011) or PyMol (Schrödinger: (this is the missing reference in the text.. thanks)).

3 Results and discussion

3.1 Structure of Rv2991 and comparison with other FDOR proteins

There is a single subunit of Rv2991 in the asymmetric unit and the electron density was traceable from residue Thr3 to Ala162. The overall crystal structure is composed of a split β-barrel typical of the flavin/deazaflavin oxidoreductase (FDORs) superfamily (Ahmed et al., 2015). PISA analysis suggests a functional biological dimer as for other members of the superfamily. Analysis of purified Rv2991 by dynamic light scattering indicated that the protein is dimeric in solution (data not shown). The total surface area of the dimer is 15912.7 Å^2 (subunit surface is 9378.9 Å^2) and the buried area upon dimerization is 2845.1 Å^2 (the subunit buried area is 1422.5 Å^2). Fig. 1 shows the dimer.

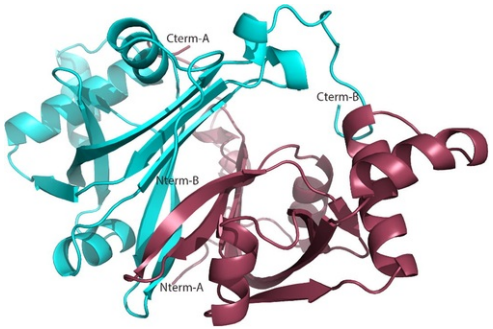


Fig. 1 Ribbon representation of the Rv2991 dimer generated by PISA. Chain A is represented in raspberry and chain B in cyan.

The structural similarity search with PDBeFOLD, limiting the search to 70% secondary structural element, found 97 matches indicating a highly conserved fold. Table 4 reports the first nine non-redundant structures ranked according to the Q score and selected within the first 30 entries found. The most significant structural alignment is with 5JAB, the biliverdin reductase Rv2074 from *M. tuberculosis* in complex with F_{420} (Ahmed et al., 2016) with which it shares a 24% sequence identity (see Table 4).

Table 4 PDBeFold results sorted according to their Q-score. In bold is the protein with the highest sequence identity.

PDB ID	Q	P	Z	RMSD (Å)	Nres	Nalign	Gaps	SI%	SSeQ%
2HTI	0.53	13.8	11.1	1.73	126	119	3	19	91
3DB0	0.51	13.3	10.9	1.74	124	116	5	19	91

5JAB	0.51	14.1	11.2	1.72	127	117	2	**24**	82
3U5W	0.49	10.4	9.6	1.84	126	117	4	16	91
4ZKY	0.47	13.4	10.9	1.63	139	116	4	22	91
2HQ9	0.45	10.8	9.8	1.75	138	116	4	22	91
1VL7	0.45	7.2	8.7	1.89	135	116	6	18	82
2RE7	0.44	11.2	10.5	1.7	132	111	6	19	91
3F7E	0.42	8.9	9.1	2.32	126	116	6	17	82

The Q-score represents the quality function of Cα-alignment, it takes both the alignment length and RMSD into account (Krissinel and Henrick, 2004). 2HTI, flavin-nucleotide-binding protein bh_0577 from *Bacillus halodurans* (JCSG to be published); 3DB0, putative pyridoxamine 5′-phosphate oxidase NP_472219.1 from *Listeria innocua* (JCSG to be published); 5JAB, biliverdin reductase Rv2074 from *M. tuberculosis* in complex with F_{420} (Ahmed et al., 2016); 3U5W uncharacterized protein from *Brucella melitensis* (SSGCID to be published); 4ZKY, F_{420} binding protein, MSMEG_6526, from *Mycobacterium smegmatis* (Ahmed et al., 2015); 2HQ9, hypothetical protein (NP_107146.1) from *Mesorhizobium loti* (JCSG to be published); 1VL7, putative heme oxygenase (alr5027) from *Nostoc* sp. pcc 7120 (JCSG to be published); 2RE7, General Stress Protein COG3871 (YP_263493.1) from *Psychrobacter arcticus* 273-4 (JCSG to be published); 3F7E, MSMEG_3380 F_{420} reductase (to be published). Nres represents the number of residues of the target structure, Nalign represents the number of aligned residues between target and query (Rv2991), Gaps represents the gaps in the structural alignment, SI% represents the sequence identity in percentage and SSeQ% the percentage of secondary structural elements of the query (Rv2991) that match the target structure.

The sequence alignment of Rv2991 and Rv2074 is shown in Fig. 2. Such high sequence identity was not found by the sequence similarity search by BLASTp when Rv2991was compared to the whole PDB. In that search the highest similarity found was with 3F7E, *Mycobacterium smegmatis* Msmeg_3380 F_{420} reductase with a sequence coverage of only 63% and 34% sequence identity (resulting from the alignment of 37/108 residues), followed by 4QVB Rv1155 (Mashalidis et al., 2015; Canaan et al., 2005) with 42% query coverage and 33% sequence identity, and finally with 5BNC, MSMEG_6519 (Ahmed et al., 2015) with a query coverage of 53% and 31% of sequence identity. The multiple alignment confirms a limited sequence conservation as shown in Fig. S1. The structural superposition of Rv2991 and Rv2074 using SSM (Krissinel and Henrick, 2004) reveals a very conserved structural core. The Rv2991 dimer is very similar to that of Rv2074 and suggests a conserved cofactor binding site. Fig. 2 (left panel) shows the structural superposition of Rv2991 and Rv2074 with its cofactor bound, their sequence alignment (bottom panel) and the conserved residues highlighted in Rv2991 cartoon representation.

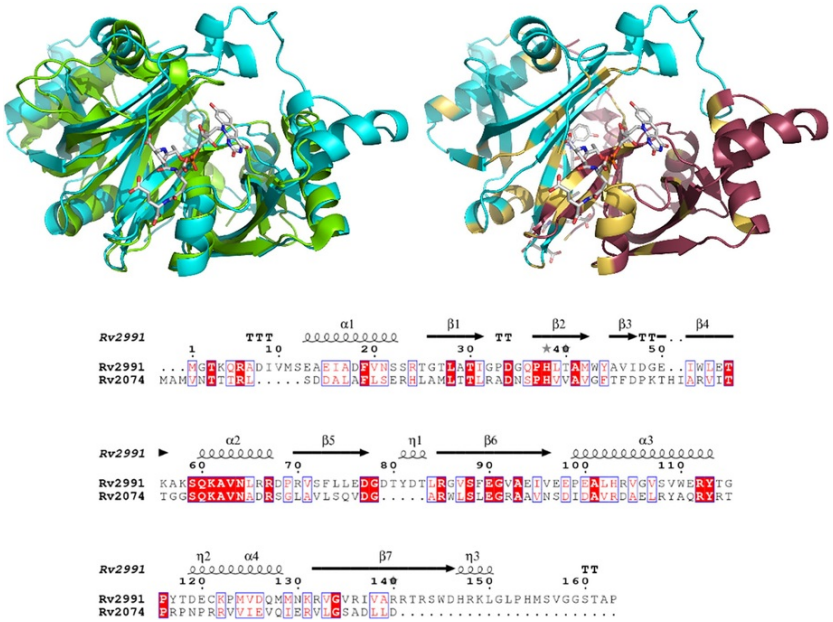


Fig. 2 The left panel shows a cartoon representation of the structural superposition of Rv2991 and Rv2074 with its cofactor bound (Rv2991 in cyan, Rv2074 in green and cofactor represented as sticks with oxygen red, carbon grey, nitrogen blue and phosphate orange), the right panel shows Rv2991 with highlighted the conserved residues in yellow. Chain A is in raspberry and chain B in cyan. Lower panel Rv2991 and Rv2074 sequence alignment.

The residues involved in cofactor binding are mostly conserved or the mutation does not affect binding (e.g., main chain atoms involved in H-bond formation). These are listed in Table 5).

Table 5 Residues involved in F_{420} cofactor binding in Rv2074 and the equivalent ones in Rv2991.

Rv2074	Rv2991
His36 Nε2	His38Nε2
Ala39 O	Ala41 O
Gly41 N	Trp43 N
Ile54 O	Glu54 O
Thr55 Oγ1	Thr55 Oγ1
Gln60 Oε1, Nε2	Gln60 Oε1, Nε2
Lys61N	Lys61N
Lys61Nζ	Lys61Nζ
Arg67 Nη1, Nη2	Arg67 Nη1, Nη2
Trp81 Nε1	Trp145 Nε1*
Arg9 Nη2	Arg6 Nη2?#

* Structurally aligned residues.

Not aligned but could change position upon cofactor binding.

Assuming that binding of the cofactor in Rv2991 is the same as in Rv2074 then the Rv2991 Trp43 (which replaces Gly41 in Rv2074, see Table 5) side chain lies at 90° to the deazaflavin ring system, a peculiarity that could be essential for the enzyme function. The last stretch of non-conserved residues includes α-helix 4 from Asp119 to Met129 and a short turn to connect to the antiparallel β-sheet 7 from Arg132, which contributes to the split β-barrel up to Asp146. The stretch of residues from His147 to Leu150 are part of a 3_{10} helix while the end of the chain from Gly151 to Ala162 forms a long loop which folds back to bind to the other subunit (Fig. 3).

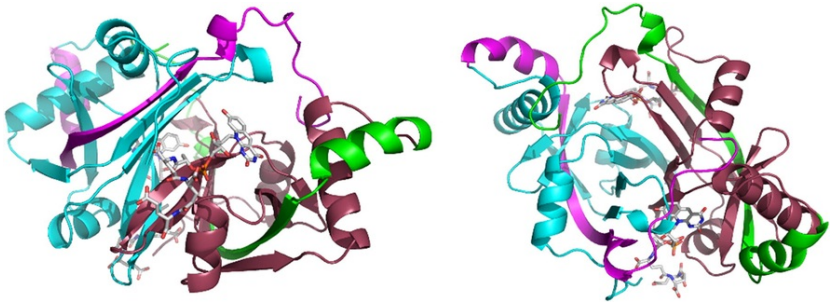


Fig. 3 Ribbon representation of two views of Rv2991 showing the conserved C-terminal. The chain A is in raspberry and the chain B in cyan. The chain A C-terminal is colored in green and chain B C-terminal is colored in magenta. The cofactor superposed is from Rv2074 5JAB. The C-terminal secondary elements include α-helix 4 from Asp119 to Met129, the antiparallel β-sheet 7 from Arg132 to Asp146 and the 3_{10} helix from His147 to Leu150.

The chain A conserved C-terminus from Gly158 to Ala162 establishes H-bonds with chain B α-helix 1 and α-helix 3 thus contributing to dimer stabilization. In detail, Gly158A O binds to Arg24B Nη1 (2.95 Å, α-helix 1), SerA160 N binds to GluB111 O (2.78 Å, α-helix 3), Thr161A Oγ1 simultaneously binds to ArgB112 O (2.76 Å, α-helix 3) and ArgB24 Nη1 (2.91 Å, α-helix 1). The conserved C-terminal of Rv2991 is a feature shared by all the 27 aligned protein

sequences with high identity as shown in the alignment of Fig. 4.

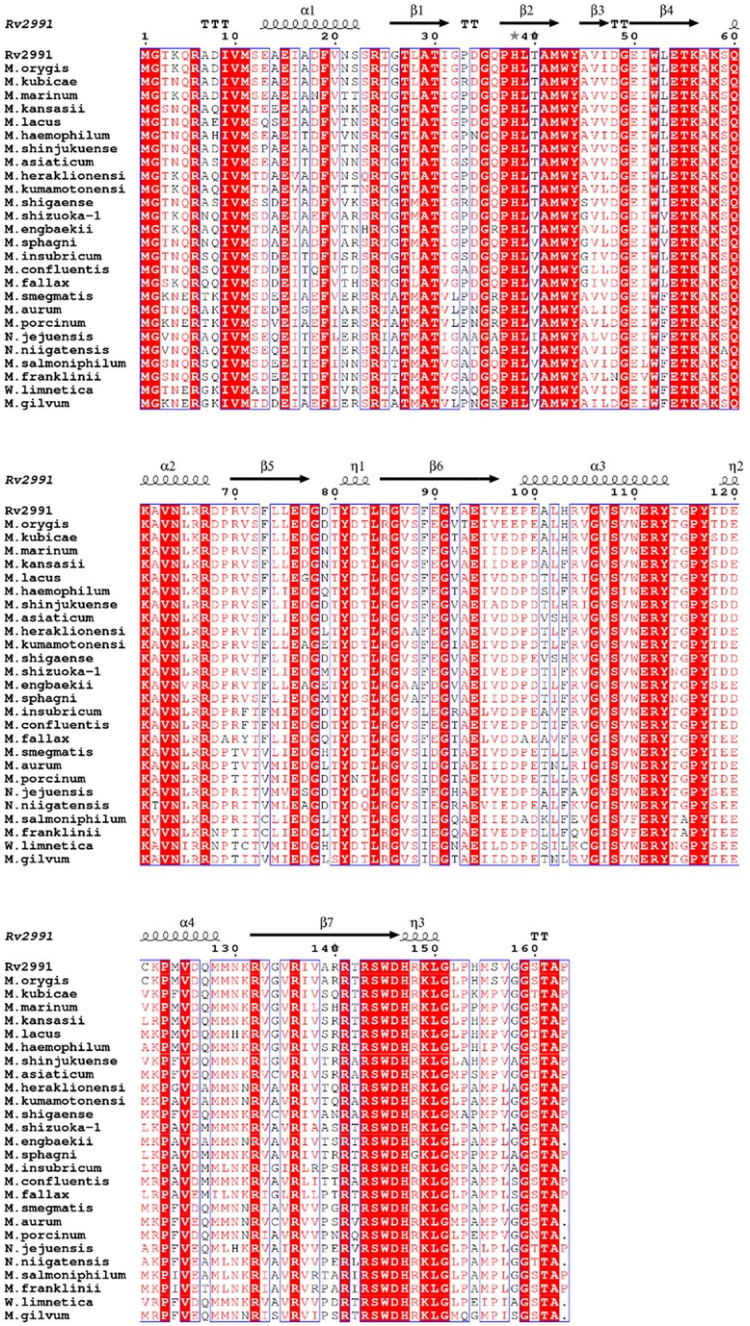


Fig. 4 Multiple sequence alignment of the 27 selected sequences. Rv2991 and homologue from *Mycobacterium canettii, Mycobacterium orygis, Mycobacterium kubicae, Mycobacterium marinum, Mycobacterium kansasii, Mycobacterium lacus, Mycobacterium haemophilum,*

Mycobacterium shinjukuense, Mycobacterium asiaticum, Mycolicibacter heraklionensis, Mycolicibacter kumamotonensis, Mycobacterium shigaense, Mycobacterium sp. shizuoka-1, Mycolicibacter engbaekii, Mycolicibacterium sphagni, Mycolicibacterium insubricum, Mycolicibacterium confluentis, Mycolicibacterium fallax, Mycolicibacterium smegmatis, Mycolicibacterium aurum, Mycolicibacterium porcinum, Nocardia jejuensis, Nocardia niigatensis, Mycobacteroides salmoniphilum, Mycobacteroides franklinii, Williamsia limnetica, Mycolicibacterium gilvum,

Arg24, Glu111 and Arg112 together with Thr161, are conserved in all the species considered.

The structural alignment of the structures selected by PDBeFold shows that, besides the conserved core involved in cofactor binding, the major differences are always in α-helix 4 (from residue 119 to 129) and the C-terminal stretch of residues (from residue 147 to 162). We propose that α-helix 4 and the stretch of residues from His147 to Gly158 are likely to be involved in substrate binding and specificity, regulating its entrance by direct interaction and steric hindrance possibly accommodating the compounds, in respect to the F_{420} cofactor, in a position as to allow the reaction to occur. Fig. 5 compares all the nine structures superposed on Rv2991 by SSM and grouped in Table 4.

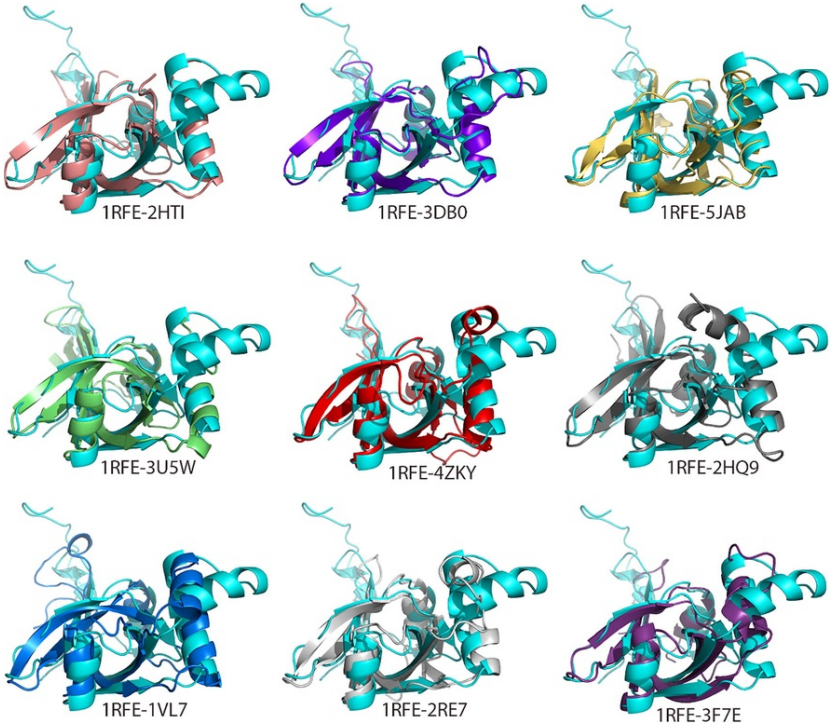


Fig. 5 Structural superposition between Rv2991 (1RFE, cyan) and the structures grouped in Table 4. 2HTI (salmon), flavin-nucleotide-binding protein bh_0577 from *B. halodurans* (JCSG to be published), 3DB0 (dark purple), putative pyridoxamine 5′-phosphate oxidase NP_472219.1 from *L. innocua* (JCSG to be published), 5JAB (yellow), biliverdin reductase Rv2074 from *M. tuberculosis* in complex with F_{420} (Ahmed et al., 2016), 3U5W (green), uncharacterized protein from *B. melitensis* (SSGCID to be published), 4ZKY (red), F_{420} binding protein, MSMEG_6526, from *M. smegmatis* (Ahmed et al., 2015), 2HQ9 (grey) hypothetical protein (NP_107146.1) from *M. loti* (JCSG to be published), 1VL7 (blue), putative heme oxygenase (alr5027) from *Nostoc* sp. pcc 7120 (JCSG to be published), 2RE7 (light grey) General Stress Protein COG3871 (YP_263493.1) from *P. arcticus* 273-4 (JCSG to be published), 3F7E (violet) MSMEG_3380 F_{420} reductase (to be published).

The electrostatic potential surface shows that a patch of negatively charged residues is located on the face opposite to the cofactor binding site. This patch is likely to be the site of interaction of a positively charged moiety of either another unknown protein or a ligand and could represent a putative allosteric regulation site. The left view of Fig. 6 shows the F_{420} cofactor (taken from Rv2074, 5JAB) bound to the binding groove while, the right view of Fig. 6, which is rotated of 180° along the vertical axis, shows that the F_{420} could be reached through a gap in the surface at the top of the negatively charged pocket.

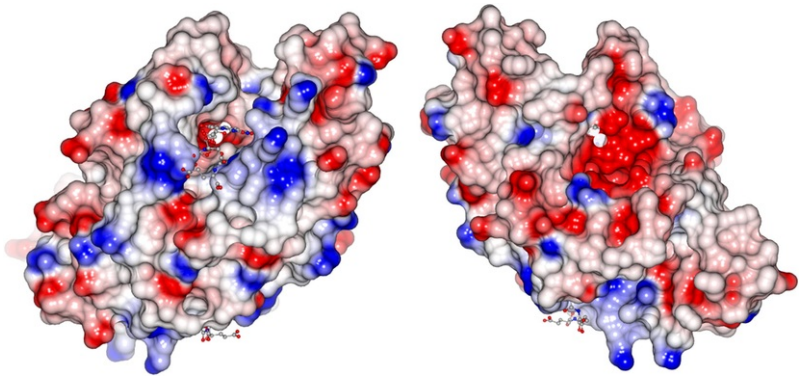


Fig. 6 The electrostatic potential surface generated by CCP4mg (McNicholas et al., 2011). Negative charges are in red while positive charges in blue. The left panel shows the surface on the side of the F_{420} cofactor binding site. The right panel shows the protein rotated by 180° along the vertical axis. The cofactor superposed is from Rv2074 5JAB and is represented in ball and stick (carbon gray, nitrogen blue, oxygen red, phosphorous brown).

3.2 Conservation of Rv2991 in *Mycobacterium* spp.

The conservation of Rv2991 in *Mycobacterium* spp. has been analyzed as reported previously for the biliverdin reductase Rv2074 (Ahmed et al., 2016). Table 6 reports the results of a BLASTp search of the Rv2991 sequence against non-redundant databases compared to the results of Rv2074. An overall decreased presence of Rv2991 homologues is observed. Rv2991 is present in 25 out of the 38 (66%) species considered while Rv2074 is conserved in 30 out of 38 (79%). There is a drastic reduction of conservation amongst pathogenic strains, indeed Rv2991 homologues are only found in two (*M. tuberculosis* and *M. canettii*) out of the five pathogenic species (40%) in contrast with Rv2074 which is conserved in four out of five pathogenic species (80%).

Table 6 Conservation of Rv2991 in *Mycobacterium* spp. in comparison with Rv2074 (list of genomes taken from (Ahmed et al., 2016)).

Mycobacterium species	Rv2074 homologue	Rv2991 homologue	Common habitat
Pathogenic			
Mycobacterium tuberculosis	Yes	Yes	Human infection
Mycobacterium bovis	Yes	No	Human/animal infection
Mycobacterium africanum	Yes	No	Human infection
Mycobacterium canettii	Yes	Yes	Human infection
Mycobacterium leprae	No	No	Human infection
Opportunistic/commensal			
Mycobacterium abscessus	Yes	Yes	Soil, water, infection
Mycobacterium avium	Yes	Yes	Water, soil, infection
Mycobacterium intracellulare	Yes	Yes	Water, soil, infection
Mycobacterium marinum	Yes	Yes	Water, soil, infection
Mycobacterium kansasii	Yes	Yes	Human infection
Mycobacterium fortuitum	Yes	Yes	Soil, water, infection
Mycobacterium xenopi	Yes	Yes	Soil, water, infection

Mycobacterium neoaurum	Yes	No	Soil, infection
Mycobacterium ulcerans	Yes	No	Water, soil, infection
Mycobacterium chelonae	Yes	Yes	Water, soil, infection
Mycobacterium genavense	Yes	Yes	Human infection
Mycobacterium iranicum	Yes	Yes	Human infection
Mycobacterium yongonense	Yes	No	Human infection
Mycobacterium simiae	Yes	Yes	Rhesus monkey
Mycobacterium haemophilum	Yes	Yes	Human infection
Mycobacterium lepromatosis	No	No	Human infection
Mycobacterium setense	Yes	Yes	Human infection
Mycobacterium kyorinense	No	Yes	Human infection
Mycobacterium vulneris	Yes	Yes	Human infection
Mycobacterium celatum	No	Yes	Human infection
Rarely pathogenic/commensal			
Mycobacterium phlei	Yes	Yes	Soil
Mycobacterium smegmatis	Yes	Yes	Soil
Mycobacterium rhodesiae	Yes	Yes	Soil
Non-pathogenic			
Mycobacterium rutilum	No	Yes	Soil
Mycobacterium pallens	No	No	Soil
Mycobacterium crocinum	No	No	Soil
Mycobacterium rufum	Yes	Yes	Soil
Mycobacterium aromaticivorans	Yes	No	Soil
Mycobacterium indicus pranii	Yes	No	Soil
Mycobacterium bovis (BCG vaccine)	No	No	Laboratory
Mycobacterium vanbaalenii	Yes	No	Mineral oil, soil
Mycobacterium gilvum	Yes	Yes	Water, soil
Mycobacterium chubuense	Yes	Yes	Soil

Rv2991 Overall 25/38 (66%), Pathogenic 2/5 (40%), Opportunistic/commensal 16/20 (80%), Rarely pathogenic/commensal 3/3 (100%), Non-pathogenic 4/10 (40%).

Rv2074 (no indent)Overall 30/38 (79%), Pathogenic 4/5 (80%), Opportunistic/commensal 17/20 (85%), Rarely pathogenic/commensal 3/3 (100%), Non-pathogenic 6/10 (60%).

4 Conclusions

Rv2991 is an F_{420} cofactor binding enzyme with a split β-barrel typical of the flavin/deazaflavin oxidoreductase (FDORs) superfamily, with a highly-conserved cofactor binding region, but also with a divergent C-terminal possibly involved in substrate recognition and active site fold conservation. A patch of negatively charged residues is located on the face opposite to the cofactor binding site and we speculate that this could represent a putative allosteric regulation site. Compared to Rv2074, Rv2991 is less conserved across *Mycobacterium* spp., which hints at a redundant role of Rv2991 in the enzymatic pool of these organisms, confirming its non-essentiality for survival and pathogenicity of *M. tuberculosis* (Sassetti et al., 2001; Sassetti et al., 2003; Sassetti and Rubin, 2003). However, a role in antibiotic activation/resistance cannot be ruled out. Further *in vitro* investigations are required to identify its substrates and reaction products.

Uncited reference

Schrodinger (2019).

Acknowledgements

The work was supported by a grant from the 5th PCRDT program of the European Union (acronym X-TB). The authors thank the EU for support through the grant 'Structural and Functional Genomics of Mycobacterium tuberculosis' (QLK2-CT-2001-02018). We thank the staff of the recombinant protein platform at the Institut Pasteur for the production and purification of Rv2991. We are grateful to Dr. William Shepard for the assistance during the collection of diffraction data at beamline ID29 at the ESRF. Guy Dodson and Eleanor Dodson are acknowledged for the helpful discussions and contribution to the project during structure determination in 2003. SB thanks all the co-authors for their patience during manuscript preparation.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsb.2019.03.006.

References

Ahmed F.H., Carr P.D., Lee B.M., Afriat-Jurnou L., Mohamed A.E., Hong N.S., Flanagan J., Taylor M.C., Greening C. and Jackson C.J., Sequence-structure-function classification of a catalytically diverse oxidoreductase superfamil in *Mycobacteria*, *J. Mol. Biol.* **427**, 2015, 3554–3571.

Ahmed F.H., Mohamed A.E., Carr P.D., Lee B.M., Condic-Jurkic K., O'Mara M.L. and Jackson C.J., Rv2074 is a novel F_{420} H_2–dependent biliverdin reductase in *Mycobacterium tuberculosis*, *Protein Sci.* **25**, 2016, 1692–1709.

Benini S., Chechik M., Ortiz Lombardia M., Polier S., Leech A., Shevtsov M.B. and Alonso J.C., The 1.58 ~~A~~Å resolution structure of the DNA-binding domain of bacteriophage SF6 small terminase provides new hints on DNA binding, *Acta Crystallogr Sect. F. Struct. Biol. Cryst. Commun.* **69**, 2013, 376–381.

Bricogne G., Vonrhein C., Flensburg C., Schiltz M. and Paciorek W., Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0, *Acta Crystallogr. D Biol. Crystallogr.* **59**, 2003, 2023–2030.

Canaan S., Sulzenbacher G., Roig-Zamboni V., Scappuccini-Calvo L., Frassinetti F., Maurin D., Cambillau C. and Bourne Y., Crystal structure of the conserved hypothetical protein Rv1155 from *Mycobacterium tuberculosis*, *FEBS Lett.* **579**, 2005, 215–221.

Carlier J.P., Sellier N., Rager M.N. and Reysset G., Metabolism of a 5-nitroimidazole in susceptible and resistant isogenic strains of *Bacteroides fragilis*, *Antimicrob. Agents Chemother.* **41**, 1997, 1495–1499.

Cellitti S.E., Shaffer J., Jones D.H., Mukherjee T., Gurumurthy M., Bursulaya B., Boshoff H.I., Choi I., Nayyar A., Lee Y.S., Cherian J., Niyomrattanakit P., Dick T., Manjunatha U.H., Barry C.E., 3rd, Spraggon G. and Geierstanger B.H., Structure of Ddn, the deazaflavin-dependent nitroreductase from *Mycobacterium tuberculosis* involved in bioreductive activation of PA-824, *Structure* **20**, 2012, 101–112.

Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E., 3rd, Tekaia F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver K., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton J., Squares R., Squares S., Sulston J.E., Taylor K., Whitehead S. and Barrell B.G., Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, *Nature* **393**, 1998, 537–544.

Daniels L., Bakhiet N. and Harmon K., Widespread distribution of a 5-deazaflavin cofactor in actinomyces and related bacteria, *System. Appl. Microbiol.* **6**, 1985, 12–17.

Emsley P. and Cowtan K., Coot: model-building tools for molecular graphics, *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2004, 2126–2132.

Gonzalez A., Pedelacq J., Sola M., Gomis-Ruth F.X., Coll M., Samama J. and Benini S., Two-wavelength MAD phasing: in search of the optimal choice of wavelengths, *Acta Crystallogr. D Biol. Crystallogr.* **55**, 1999, 1449–1458.

Hemmerich P., Nagelschneider G. and Veeger C., Chemistry and molecular biology of flavins and flavoproteins, *FEBS Lett.* **8**, 1970, 69–83.

Krissinel E. and Henrick K., Inference of macromolecular assemblies from crystalline state, *J. Mol. Biol.* **372**, 2007, 774–797.

Krissinel E. and Henrick K., Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2004, 2256–2268.

Manjunatha U.H., Boshoff H., Dowd C.S., Zhang L., Albert T.J., Norton J.E., Daniels L., Dick T., Pang S.S. and Barry C.E., 3rd, Identification of a nitroimidazo-oxazine-specific protein involved in PA-824 resistance in *Mycobacterium tuberculosis*, *Proc. Natl. Acad. Sci. USA* **103**, 2006, 431–436.

Mashalidis E.H., Gittis A.G., Tomczak A., Abell C., Barry C.E., 3rd and Garboczi D.N., Molecular insights into the binding of coenzyme F_{420} to the conserved protein Rv1155 from *Mycobacterium tuberculosis*, *Protein Sci.* **24**, 2015, 729–740.

McNicholas S., Potterton E., Wilson K.S. and Noble M.E., Presenting your structures: the CCP4mg molecular-graphics software, *Acta Crystallogr. D Biol. Crystallogr.* **67**, 2011, 386–394.

Morris R.J., Zwart P.H., Cohen S., Fernandez F.J., Kakaris M., Kirillova O., Vonrhein C., Perrakis A. and Lamzin V.S., Breaking good resolutions with ARP/wARP, *J. Synchrotron Radiat.* **11**, 2004, 56–59.

Otwinowski Z. and Minor W., Processing of X-ray diffraction data collected in oscillation mode, In: Carter C.W. and Sweet R.M., (Eds.), *Macromolecular Crystallography Part A,* 1997, Elsevier, 307–326.

Otwinowski Z., Borek D., Majewski W. and Minor W., Multiparametric scaling of diffraction intensities, *Acta Crystallogr. A* **59**, 2003, 228–234.

Papadopoulos J.S. and Agarwala R., COBALT: constraint-based alignment tool for multiple protein sequences, *Bioinformatics* **23**, 2007, 1073–1079.

Potterton L., McNicholas S., Krissinel E., Gruber J., Cowtan K., Emsley P., Murshudov G.N., Cohen S., Perrakis A. and Noble M., Developments in the CCP4 molecular-graphics project, *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2004 2288–2294.

Robert X. and Gouet P., Deciphering key features in protein structures with the new ENDscript server, *Nucl. Acids Res.* **42**, 2014, W320–W324.

Sassetti C.M., Boyd D.H. and Rubin E.J., Genes required for mycobacterial growth defined by high density mutagenesis, *Mol. Microbiol.* **48**, 2003, 77–84.

Sassetti C.M., Boyd D.H. and Rubin E.J., Comprehensive identification of conditionally essential genes in mycobacteria, *Proc. Natl. Acad. Sci. USA* **98**, 2001, 12712–12717.

Sassetti C.M. and Rubin E.J., Genetic requirements for mycobacterial survival during infection, *Proc. Natl. Acad. Sci. USA* **100**, 2003, 12989–12994.

Schneider T.R. and Sheldrick G.M., Substructure solution with SHELXD, *Acta Crystallogr. D Biol. Crystallogr.* **58**, 2002, 1772–1779.

Schrödinger, L., The PyMOL Molecular Graphics System, Version 2.2, Schrödinger, LLC.

Selengut J.D. and Haft D.H., Unexpected abundance of coenzyme F_{420}-dependent enzymes in *Mycobacterium tuberculosis* and other actinobacteria, *J. Bacteriol.* **192**, 2010, 5788–5798.

Singh R., Manjunatha U., Boshoff H.I., Ha Y.H., Niyomrattanakit P., Ledwidge R., Dowd C.S., Lee I.Y., Kim P., Zhang L., Kang S., Keller T.H., Jiricek J. and Barry C.E., 3rd, PA-824 kills nonreplicating *Mycobacterium tuberculosis* by intracellular NO release, *Science* **322**, 2008, 1392–1395.

Taylor M.C., Jackson C.J., Tattersall D.B., French N., Peat T.S., Newman J., Briggs L.J., Lapalikar G.V., Campbell P.M., Scott C., Russell R.J. and Oakeshott J.G., Identification and characterization of two families of F_{420} H_2-dependen reductases from *Mycobacteria* that catalyse aflatoxin degradation, *Mol. Microbiol.* **78**, 2010, 561–575.

Vagin A.A., Steiner R.A., Lebedev A.A., Potterton L., McNicholas S., Long F. and Murshudov G.N., REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use, *Acta Crystallographica Section D: Biol Crystall.* **60**, 2004, 2184–2195.

Van Duyne G.D., Standaert R.F., Karplus P.A., Schreiber S.L. and Clardy J., Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin, *J. Mol. Biol.* **229**, 1993, 105–124.

Ye J., McGinnis S. and Madden T.L., BLAST: improvements for better sequence analysis, *Nucl. Acids Res.* **34**, 2006, W6–9.

Appendix A. Supplementary material

The following are the Supplementary data to this article:

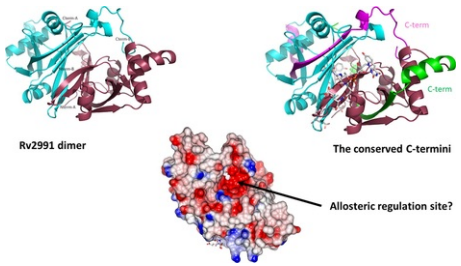
Multimedia Component 1

Supplementary data 1

Multimedia Component 2

Supplementary data 2

Graphical abstract



Highlights

- Rv2991 is an F_{420} cofactor binding enzyme with a split β-barrel.
- Rv2991 shows a divergent C-terminal region possibly involved in substrate recognition.
- Rv2991 features a putative allosteric regulation site on the face opposite to F_{420} binding pocket.
- Compared to Rv2074, Rv2991 is less conserved across *Mycobacterium* spp.

Queries and Answers

Query: Your article is registered as a regular item and is being processed for inclusion in a regular issue of the journal. If this is NOT correct and your article belongs to a Special Issue/Collection please contact p.punniyaseelan@elsevier.com immediately prior to returning your corrections.
Answer: Yes

Query: The author names have been tagged as given names and surnames (surnames are highlighted in teal color). Please confirm if they have been identified correctly.
Answer: Yes

Query: This section comprises references that occur in the reference list but not in the body of the text. Please cite each reference in the text or, alternatively, delete it. Any reference not dealt with will be retained in this section.

Answer: this reference relates to the software PyMol used for the pictures, see section:

2.5 Sequence alignments and structural comparison

end of sentence: Figures representing the structures were prepared with CCP4mg (Potterton et al., 2004,McNicholas et al., 2011) or PyMol (Schrödinger.).

Query: Have we correctly interpreted the following funding source(s) and country names you cited in your article: Institut Pasteur, France? /

Answer: Yes

Query: Note that the color shade has been removed in Table 6. Please check and correct if necessary.

Answer: color shade was not essential so better with white background, thank you for changing it