Journal of Structural and Functional Genomics **6**: 245–257, 2005.
DOI 10.1007/s10969-005-9004-6

The crystal structure of Rv0793, a hypothetical monooxygenase from *M. tuberculosis*

M. Joanne Lemieux[1], Claire Ference[1], Maia M. Cherney[1], Metian Wang[2], Craig Garen[1] & Michael N. G. James[1,*]

[1]*Department of Biochemistry, CIHR Group in Protein Structure and Function, University of Alberta, Edmonton, Alberta, T6G 2H7, Canada;* [2]*Present address: Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona, 85287-1604, USA; *Author for correspondence: (e-mail: michael.james@ualberta.ca; fax +780-492-0886)*

Received 8 July 2005; accepted in revised form 13 October 2005

Key words: Mycobacterium tuberculosis, antibiotic biosynthesis, monooxygenase, quinol monooxygenase, X-ray crystallography

Abstract

Mycobacterium tuberculosis infects millions worldwide. The Structural Genomics Consortium for *M. tuberculosis* has targeted all genes from this bacterium in hopes of discovering and developing new therapeutic agents. Open reading frame Rv0793 from *M. tuberculosis* was annotated with an unknown function. The 3-dimensional structure of Rv0793 has been solved to 1.6 Å resolution. Its structure is very similar to that of *Streptomyces coelicolor* ActVA-Orf6, a monooxygenase that participates in tailoring of polyketide antibiotics in the absence of a cofactor. It is also similar to the recently solved structure of YgiN, a quinol monooxygenase from *Escherichia coli*. In addition, the structure of Rv0793 is similar to several structures of other proteins with unknown function. These latter structures have been determined recently as a result of structural genomic projects for various bacterial species. In *M. tuberculosis*, Rv0793 and its homologs may represent a class of monooygenases acting as reactive oxygen species scavengers that are essential for evading host defenses. Since the most prevalent mode of attack by the host defense on *M. tuberculosis* is by reactive oxygen species and reactive nitrogen species, Rv0793 may provide a novel target to combat infection by *M. tuberculosis*.

Introduction

Tuberculosis is the most widely spread disease due to a single microbial agent [1]. The World Health Organization has declared tuberculosis infection by the pathogenic *Mycobacterium tuberculosis* (*M. tuberculosis* or TB) a global epidemic (http://www.who.int/health_topics/tuberculosis/en/). In the year 2002 alone, 8.8 M new cases of tuberculosis were reported worldwide; 1.8 M deaths were also reported. An increase in the incidence of tuberculosis has been seen in the past few decades due to a rise in the number of multi-drug resistant strains of *M. tuberculosis* as well as in its rapid spread in HIV infected patients. It is estimated that one third of the 40 million people living with HIV/AIDS are also infected with TB, with 70% of those living in Africa.

The *M. tuberculosis* Structural Genomics Consortium (TBSGC) has undertaken the task of determining the 3-dimensional structures of all proteins from this pathogen (http://www.doembi.ucla.edu/TB/) [2]. Of the 3924 open-reading frames (ORF) in the *M. tuberculosis* genome, many have an unknown function. Elucidation of the 3-dimensional structures and functions of these novel proteins may provide new potential drug targets to combat this pathogen. The strat-

egy undertaken by our lab was to target small cytoplasmic proteins with unknown function for structural determination. Proteins with predicted TM domains, as determined by TBSGC bioinformatics data for each ORF, were rejected. ORFs targeted by other TBSGC members were also eliminated to avoid duplication of effort. In addition, the number of methionine residues was considered for help in the determination of phases. The final criterion applied was that TB-specific unknown ORFs would take precedence over those with homologues in other species.

Rv0793 is an 11 kDa protein having 101 amino acids. It was initially identified as a conserved hypothetical protein with unknown function according to the TBSGC. A subsequent functional annotation search of the Protein Family Database (Pfam) with the sequence of Rv0793 from *M. tuberculosis* resulted in a match with the antibiotic biosynthesis monooxygenase (ABM) domain. A 3-dimensional crystal structure of a protein containing this domain, a monooxygenase from the gene ActVA-Orf6 of *Streptomyces coelicolor*, had already been determined [3, 4]. This homodimeric protein has been shown to play a role in the biosynthesis of aromatic polyketides, specifically the antibiotic actinorhodin, by oxidizing phenolic groups to quinones [5]. The biosynthesis of polyketides involves four phases: initiation, cyclization, elongation and tailoring. ActVA-Orf6 monooxygenase participates in the tailoring phase by its oxidation function.

The synthesis of polyketides takes place in many bacteria, fungi and plants [6] and the mechanism for biosynthesis of these secondary metabolites has been studied in great detail [7]. The synthesis of complex antibiotics by microorganisms provides a means for self-defense, especially for those organisms that are non-motile [8]. Interestingly, a recent study has identified a polykeytide-derived phenolic glycolipid moiety in *M. tuberculosis* whose production was shown to inhibit the host's immune response [9].

Interest in polyketides also stems from their role in drug development. While many polyketide drugs have been isolated from *Streptomyces*, no aromatic polyketide products or antibiotics have been isolated in *M. tuberculosis* thus far, even though both bacteria belong to the *Actinomyces* phylogeny [10]. The identification of enzymes that play a role in antibiotic biosynthesis in *M. tuberculosis* not only may provide a means for targeting attack against the bacteria, but also may provide new therapeutic drugs to combat other organisms. This latter issue is important in light of the increasing instances of multi-drug resistance in microorganisms.

The recently determined structure of YgiN from *Escherichia coli* [11] revealed a structure similar to both Rv0793 and ActVa-Orf6. YgiN is a quinol monooxygenase that participates in the quinol redox cycle. In addition, it is proposed to act as a free-radical scavenger in *E. coli*.

Both ActVA-Orf6 and YgiN do not require a metal cofactor for their activity, making them interesting and unique monooxygenases. Thus far, two other similar proteins have been identified, TcmH from *Streptomyces glaucescens* [12, 13] and ElmH from *Streptomyces olivaceus* [14]. These enzymes also do not require a cofactor for their monooxygenase activities.

The function of *M. tuberculosis* Rv0793 protein has yet to be determined conclusively. The 3-dimensional structure of Rv0793 from *M. tuberculosis* was sought in order to determine whether Rv0793 is similar in structure to ActVA-Orf6 and to establish a function for this conserved protein. The X-ray crystal structure of Rv0793 has been solved to 1.6 Å resolution. The overall structure of Rv0793 is similar to the structures of ActVA-Orf6 (Protein Data Bank code 1LQ9.pdb) from *S. coelicolor*, and YgiN (Protein Data Bank code 1R6Y.pdb) from *E. coli*. The fact that these two homologous structures are of proteins that function as monooxygenases suggests that Rv0793 may also carry out a similar function in *M. tuberculosis*.

Materials and methods

Cloning and expression

The Rv0793 ORF was amplified by PCR using BAC supplied by the Pasteur Institute, France. The ORF was cloned into pDEST15 using the Gateway system, (Invitrogen), in frame with an N-terminal GST tag. BL21(DE3)-pLysS cells, freshly transformed with pDEST15-Rv0793, were

grown in 50 ml of LB medium supplemented with 100 μg/ml ampicillin and 34 μg/ml chloramphenicol (LBAC), for 5 h at 37 °C. One liter LBAC was then subinoculated with the 50 ml culture and grown at 37 °C until the OD_{600} reached 0.4–0.6. The cells were then induced with 0.4 mM IPTG and the temperature was reduced to 28 °C. Cells were harvested at 4 °C after 20 h and frozen at −80 °C. Selenomethionine substituted Rv0793 was generated according to the method of Doublie [15]. Briefly, pDEST15-Rv0793 was transformed into the methionine auxotroph, B834 (Novagen). A 50 ml culture of pDEST15-Rv0793 grown for 5 h in LBAC was centrifuged and resuspended in 1 l of minimal media plus the appropriate antibiotics and supplemented with 50 mg selenomethionine. After reaching an OD_{600} of 0.5, the cells were induced with 0.4 mM IPTG and harvested after growth at 28 °C for 20 h.

Purification

Cells expressing GST-Rv0793 were thawed and resuspended in 5 volumes of PBS containing 1 mM each of EDTA and freshly prepared PMSF. Cells were lysed by sonication. Unbroken cells and inclusion bodies were pelleted by centrifugation at $12{,}000 \times g$. The supernatant was subjected to a GSTPrep 16/10 column (GE Healthcare, formerly Amersham Biosciences) equilibrated with PBS. GST-Rv0793 was eluted with 4 column volumes of 50 mM Tris, 10 mM glutathione, pH 8.0. The protein concentration was determined by the Bradford assay (BioRad). In order to separate Rv0793 from the GST tag, GST-Rv0793 was subjected to rTEV protease (Invitrogen) at a ratio of 50 Units/mg of GST-Rv0793. Digestion was carried out at room temperature overnight with rocking. The digested protein was dialyzed in PBS at 4 °C and subsequently subjected to a GSTPrep column in order to remove the free GST as well as any undigested fusion protein. Flow-through fractions containing Rv0793 were pooled and concentrated to 10 mg/ml. A similar protocol was carried out for the purification selenomethionine substituted Rv0793; 1 mM DTT was included in all lysis, purification and crystallization steps. After rTEV cleavage, the expressed Rv0793 contained an extra glycine residue preceding the initial methionine.

Oligomeric state analysis

Purified Rv0793, 50 μg, was injected into a Bio-Rad Bio-Sil 125-5 column (300 mm × 7.8 mm) pre-equilibrated with 0.1 mM phosphate and 0.1 mM KCl, pH 6.8. The column was run at a rate of 0.5 ml/min for 25 min.

Crystallization

The protein was dialyzed either in H_2O or in 10 mM Hepes and 5 mM ammonium acetate, pH 7.0. The final aliquot was filtered using a Millipore Ultrafree-MC centrifugal filter, 0.65 ml capacity and 0.22 pore size (Millipore), to remove any impurities or aggregates. Crystallization conditions were screened using the sitting drop method with a Robbins Scientific *Hydra96* crystallization robot and the High Throughput crystallization screen (Hampton research). Using vapor drop diffusion, crystallization conditions were optimized, with the best crystals produced from hanging drops with a 1:1 protein:mother liquor ratio. The final mother liquor contained 12% PEG20K, 100 mM Hepes, 50 mM ammonium acetate, 30% EG, 15% isopropanol, pH 7.0. Crystals were frozen directly in liquid nitrogen. Crystals having both spacegroups P1 and $P3_121$ appeared in the above conditions.

Co-crystallization of Rv0793 in the presence of various aromatic polyketide analogs was attempted. The analogs include 1 mM menadione, 1 mM acetyl dithranol, 1 mM tetracycline, 1 mM 8-Anilino-1-naphthalenesulfonic acid, and 1 mM 5-Hydroxy-1,4-naphthoquinone (jugalone). In addition, strongly-diffracting crystals of Rv0793 were soaked into solutions containing 1–10 mM concentrations of the above mentioned analogs.

Structural determination

Crystals containing Se-substituted Rv0793 were subjected to synchrotron radiation at ALS, Lawrence Berkeley National Lab, beamline 8.3.1. SAD data were collected at the peak wavelength, 0.979323 Å (Table 1). The data were reduced with Denzo [16]. The programs SOLVE [17] and RESOLVE [18] were used to calculate phases and generate the electron density map, (http://www.solve.lanl.gov). Refinement was carried out with Refmac5 [19] in CCP4 (http://www.ccp4.

Table 1. X-ray diffraction data collection and atomic refinement for Rv0793 from *M. tuberculosis*.

Data reduction	*SeMet*
Space group	P1
Wavelength (Å)	0.979232
Z[a]	2/AU
Unit cell dimensions	
a (Å)	30.4
b	34.6
c	42.8
α (°)	103.4
β	107.9
γ	94.3
Resolution range (Å)	40–1.6
Unique reflections	20461
Avg multiplicity	3.6
Completeness[b]	96.6 (95.4)
$\langle I/\sigma I\rangle$[c]	8.5 (4.8)
R_{merge}[d]	0.098 (0.485)
Phasing	
# Se sites/AU	4
Resolution range	40–1.6
Phasing power[e]	4.1
Figure of merit-SAD	0.26
Figure of merit-DM	0.74
Refinement	
R_{work}/R_{free}[f]	0.17/0.21
Number of atoms	
Protein	1572
Water	208
Average B-factor (Å^2)	
Protein	16.8
Water	31.4
r.m.s. dev.	
Angles (°)	1.219
Bonds (Å)	0.008
Ramachandran plot	
Most favoured	163 (95.3%)
Allowed	8 (4.7%)
Generously allowed	0 (0%)

[a] Z, the number of molecules in the unit cell.
[b] Statistics for the highest resolution shell are in brackets.
[c] $\langle I/\sigma I\rangle$ = ratio between the mean intensity and the mean error of the intensity.
[d] $R_{merge} = \sum_{hkl}\sum|I_j(hkl) - \langle I(hkl)\rangle|\sum_{hkl}\sum_j\langle I(hkl)\rangle$, with I_j(hkl) representing the intensity of measurement j and $\langle I(hkl)\rangle$ the mean of measurements for the reflection hkl.
[e] Phasing power = $\langle FH\rangle$/E, where $\langle FH\rangle$ is the root mean square of the heavy-atom structure factor.
[f] $R_{work} = \sum_{hkl}||F_{obs}(hkl)| - F_{calc}(hkl)||/\sum_{hkl}|F_{obs}(hkl)|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. R_{free} is calculated in the same manner on 5% of structure factors that were not used in the model refinement.

ac.uk/main.html); the maps were generated with Xfit [20] in the XtalView package (http://www.msg.ucsf.edu/local/programs/xtalview/xtalview.html). Figures were prepared with Pymol (http://pymol.sou-rceforge.net/index.php).

Sequence and structural alignment

Sequence alignments were carried out with ClustalW [21]. The DALI server was used to identify proteins with similar structural elements [22]. The programs Swiss-PDB viewer [23] and ALIGN [24] were used to calculate r.m.s.d. for structure similarity. Sequence alignments based on the structural components of Rv0793 were carried out with 3DTcoffee [25] (http://igs-server.cnrs-mrs.fr/Tcoffee/tcoffee_cgi/index.cgi) while the sequence alignment figure was prepared with ESPript [26]. The atomic coordinates, including X-ray amplitudes and phases, for Rv0793 have been deposited to the Protein Data Bank: 1Y0H. Residue numbering is according to the native Rv0793 sequence. The model contains an additional N-terminal glycine due to cloning. Surface area calculations were carried out with AreaIMol [27].

Results

Two crystal forms of Rv0793 having different space groups: P1 and P3_121 were generated. The P1 crystals diffracted to higher resolution and therefore were used for the structural determination (Table 1). Well defined electron density (Figure 1a) was observed for the entire polypeptide chain with the exception of residues 1–3 in the B-subunit. The topologies of the A and B subunits are virtually identical, with a root mean square difference (r.m.s.d) of 0.30 Å for the 97 common Cα atom pairs (Figure 1b). The only significant difference between the two subunits is in the helix α_C on the outside of the cleft (Figure 1c and d). In the B-subunit, this helix extends from residues B72–B84, however, in the A-subunit this helix α_C is a fragmented into an α-helix (α_3) from A72–A78 followed by a short 3_{10}-helix ($\alpha_{3'}$) from A82–A84, as indicated by the smaller circle. The structure of Rv0793 belongs to the α–β sandwich group with two $\beta\alpha\beta$ motifs forming a ferredoxin-like fold shown in Figure 1c and d [28]. The β-sheets from each monomer interact to form the

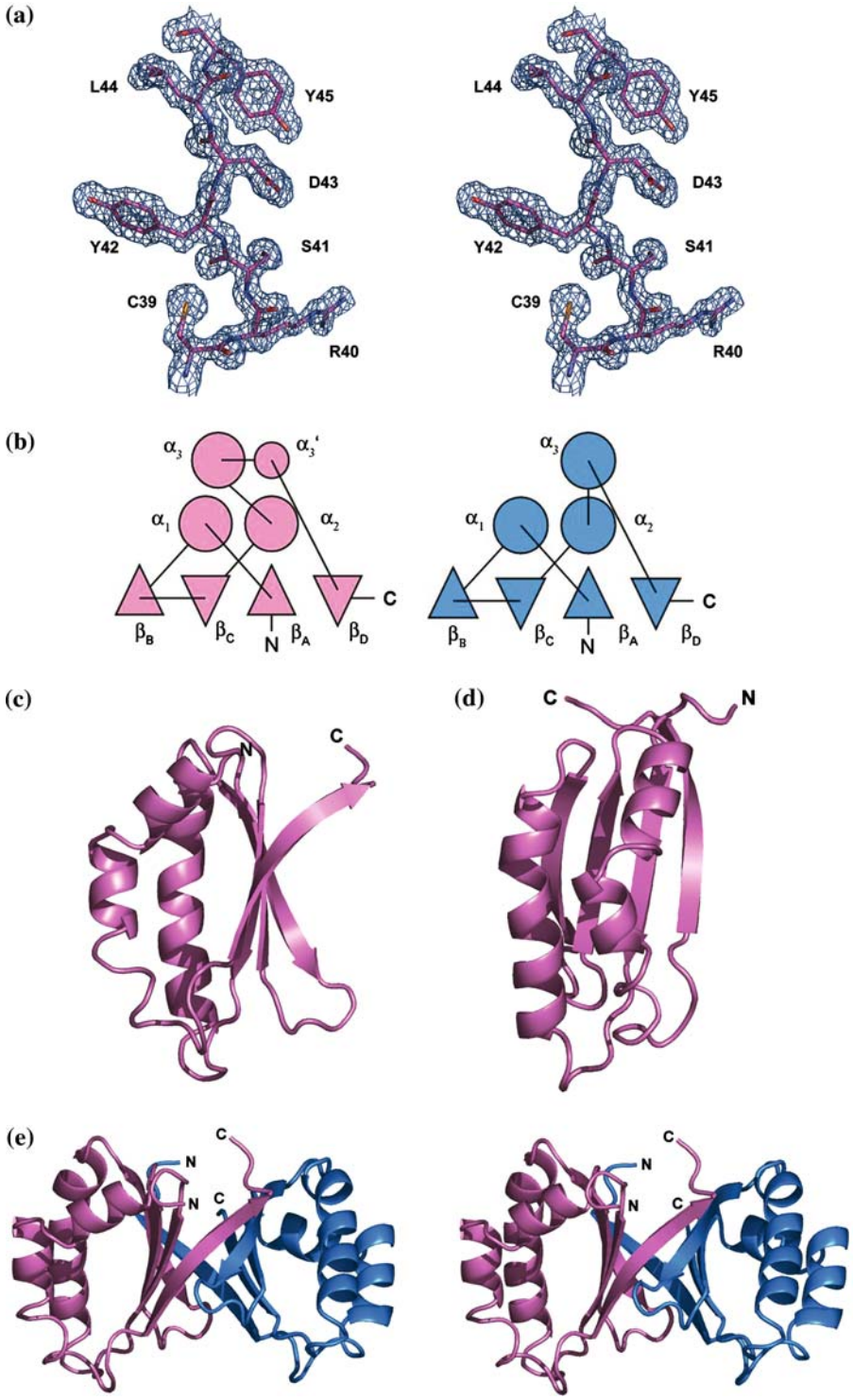

Figure 1. Structure of Rv0793 from *M. tuberculosis*. (a) Residues α39–α45 of Rv0793 with 2Fo-Fc electron density contoured at 1σ. (b) Topology diagram of the structure of the A-subunit (pink) and B-subunit (blue) of Rv0793. (c) Ribbon representation of A-subunit showing the cleft between the helices α2 and α3 and the sheet. (d) Ribbon representation of A-subunit shown in B with 90° rotation. (e) Stereo ribbon representation of both A- and B-subunits of Rv0793. Subunit-A is colored pink, while B is blue. R.m.s.d of Cα overlay between subunits A and B of Rv0793 is 0.30 Å for 97 atom pairs.

dimer interface that is an eight-stranded β-barrel. There is weak domain swapping as each C-terminal β-strand (β_D) interacts with β_B from the other subunit of the dimer.

Rv0793 crystallizes as a homodimer (Figure 1e). The dimer interface of Rv0793 comprises a buried area of 359.7 A^2, with the solvent accessible area for the entire protein being 2886.3 A^2. At the interior of the interface, the Rv0793 dimer is maintained by symmetrical hydrophobic interactions with Tyr44A and Tyr44B as well as Phe55A and Phe55B. The border of the Rv0793 dimer interface is maintained by a symmetrical arrangement of charged arginine residues; Arg10A and Arg10B interact with each other via a water molecule as do Arg57A and Arg57B. When subjected to analytical size exclusion HPLC (see supplemental information), the Rv0793 protein eluted at a retention time of 14.93 min, between the standard chicken egg white albumin, 45 kD (13.2 min) and myoglobin, 17 kD (15.8 min). The estimated molecular weight of Rv0793 being 27 kD suggests that it behaves as a homodimer in solution.

A small hydrophobic cleft is seen between the β-sheet and α-helices in each subunit (Figure 2a). Mostly hydrophobic residues line the cleft that also has a few charged residues as well. Residues in the cleft include Val7, Ala9, Phe11, Tyr41, Glu56, Tyr58, His67, Arg68, Tyr73, Tyr76, Arg77, Val88, Val90 (Figure 2b). Two acetate ions and two water molecules are also located in the cleft of subunit A of Rv0793 (Figure 2c). Only one acetate ion is found in the B-subunit.

A DALI [22] search for molecules structurally similar to Rv0793 resulted in several matches, with twenty proteins giving Z-scores over 5.0. The top 5 comparisons are listed in Table 2. The protein giving the highest score, $Z = 15.0$, was, Pa3566 (1X7V.pdb), a protein of unknown function from *Pseudomonas aeruginosa*. The closest structural match for a protein with known function was to the crystal structure of ActVA-Orf6 (1LQ9.pdb), a monooxygenase from *Streptomyces coelicolor* with a Z-score of 9.7. *E. coli* YgiN was not found by the DALI search because it was only deposited in the pdb database recently.

A functional annotation search of the Protein Family Database (Pfam) [29] with the sequence of Rv0793 from *M. tuberculosis* resulted in a match with the antibiotic biosynthesis monooxygenase (ABM) domain. In addition to Rv0793, three other proteins in *M. tuberculosis* have sequences resembling that of the ABM domain: Rv1117, Rv2749, and Rv3592. Sequence alignment with ClustalW (data not shown) revealed the weak sequence homology between Rv0793 and Rv1117, (Identity: 20.7%, Similarity: 37.9%), Rv2749, (Identity: 24.5%, Similarity: 40.0%), Rv3592, (Identity: 22.2%, Similarity: 32.5%), and ActVA-Orf6, (Identity: 17.9%, Similarity: 25.4%).

A search using the sequence for Rv0793 in the Cluster of Orthologous Groups database (COG) resulted in a match with COG1359, an ancient conserved region found in the phylogenetic branches of archaea, bacteria and eukarya. Three of the ABM homologs present in *M. tuberculosis*, Rv0793, Rv1117, and Rv2749 are associated with COG1359. Rv3592, however, is associated with COG2329, which is annotated as conserved proteins involved in polyketide biosynthesis related to monooxygenase. The *Bacillus subtilis* YjcS, the *Pseudomonas aeruginosa* Pa3566 and the *E. coli* monooxygenase YgiN are also found in COG1359. Although sequencing of the entire genome for *S. coelicolor* has been carried out, its annotation has not been completed. Therefore, it is unclear whether the ActVA-Orf6 belongs to COG1359 or COG2329. Not surprisingly, some proteins belonging to the COG1359 listed above were also found in the DALI search for structural homologs (Table 2 and Figure 3).

A sequence alignment based on the 3-dimensional structure of six proteins belonging to COG1359 was carried out with 3D-Tcoffee (Figure 4) [25]. This structurally based sequence alignment shows that only one residue is conserved throughout the group, namely, Gly37 on Rv0793. In the strands βA and βC as well as the loop between α1 and βB, some identity exists among the six proteins. The structurally-based sequence alignment with other members of COG1359 indicates conservation of some of the residues in the cleft of Rv0793, such as Glu56 and His67 (Figure 4). Many of the hydrophobic residues found in the cleft of Rv0793 are similarly found in other members of COG1359 as indicated by the structural sequence alignment.

Electrostatic surface representations of Rv0793 (Figure 5a), ActVA-Orf6 (Figure 5b) and YgiN (Figure 5c) reveal the different shapes and charges associated with each cleft. The cleft in Rv0793

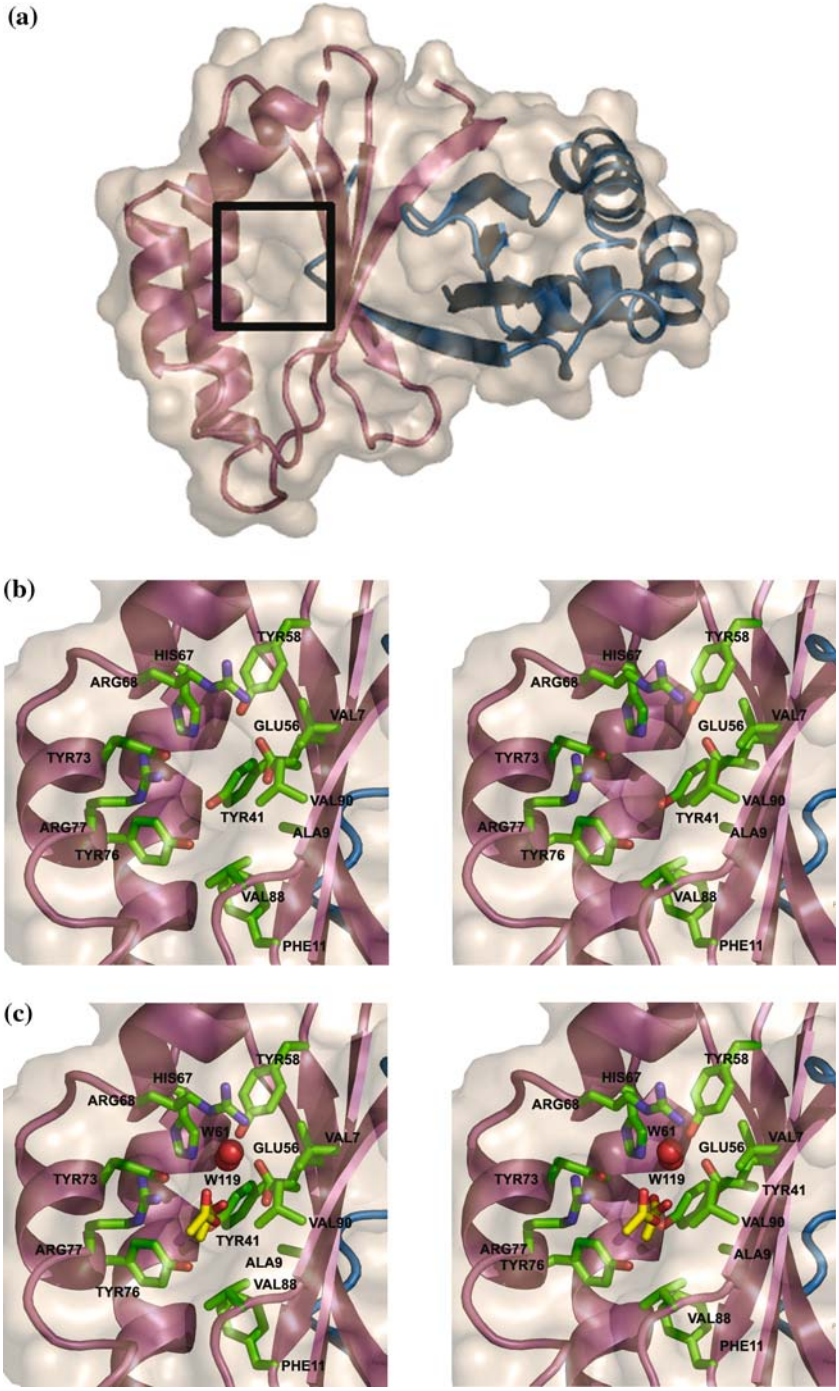

Figure 2. The cleft region of Rv0793 from *M. tuberculosis*. (a) Surface and ribbon representation of Rv0793. The box surrounding the cleft is accessible to solvent and possibly represents a substrate binding site. (b) Stereo-view of residues lining the cleft region. Residues are numbered according to the native Rv0793 protein. (c) Stereo-view of cleft from subunit-A with acetate ions and bound waters.

Table 2. DALI structural similarity search.

Rank	PDB	*Z*-score	r.m.s.d.	# Res	Identity	Protein
1	1x7v-A	15.0	1.6	94	31	Pa3566, *Pseudomonas aeruginosa*, structural genomics, unknown function
2	1q8b-A	10.1	2.7	86	15	YjcS, *Bacillus subtilis*, structural genomics, unknown function
3	1sqe-A	9.7	2.7	88	7	Pg130, *Staphylococcus aureus*, structural genomics, unknown function
4	1lq9-A	9.7	3.0	93	15	ActVA-Orf6 monooxygenase, (Oxidoreductase), *Streptomyces coelicolor*
5	1tz0-A	8.0	2.9	87	13	Anthrax toxin protective antigen Heptameric Prepore, Pa-83, *Bacillus anthracis*

Z-score represent the statistical significance of the best domain–domain alignment. r.m.s.d. represents the root mean square deviation of C atoms in rigid body superimposition. # Res represents the number of matching C atoms used to calculate the r.m.s.d. Identity reflects number of identical residues in the structurally equivalent region.

is narrow in comparison to those of ActVA-Orf6 and YgiN. Both Rv0793 and YgiN have an overall negative charge associated with their cleft, due to a conserved Glu residue (Glu56 in Rv0793 and Glu64 in YgiN, Figures 3 and 5e), whereas the cleft of ActVA-Orf6 is electroneutral. Superimposition of the residues in the cleft region of Rv0793 with those in ActVA-Orf6 (Figure 5d) and with YgiN (Figure 5e) reveals which residues contribute to the similarities and differences in the different clefts. Helix α_1 in Rv0793 is closer to the β-sheet compared to α_1 in ActVA-Orf6 and YgiN, resulting in a more narrow cleft for Rv0793 (Figures 3 and 5a, b, c). In addition, the 3_{10}-helix at the front on the cleft, α_3', constricts the cleft of Rv0793.

Residues in ActVA-Orf6 involved in polyketide substrate binding are Tyr51, Asn62, Trp66,

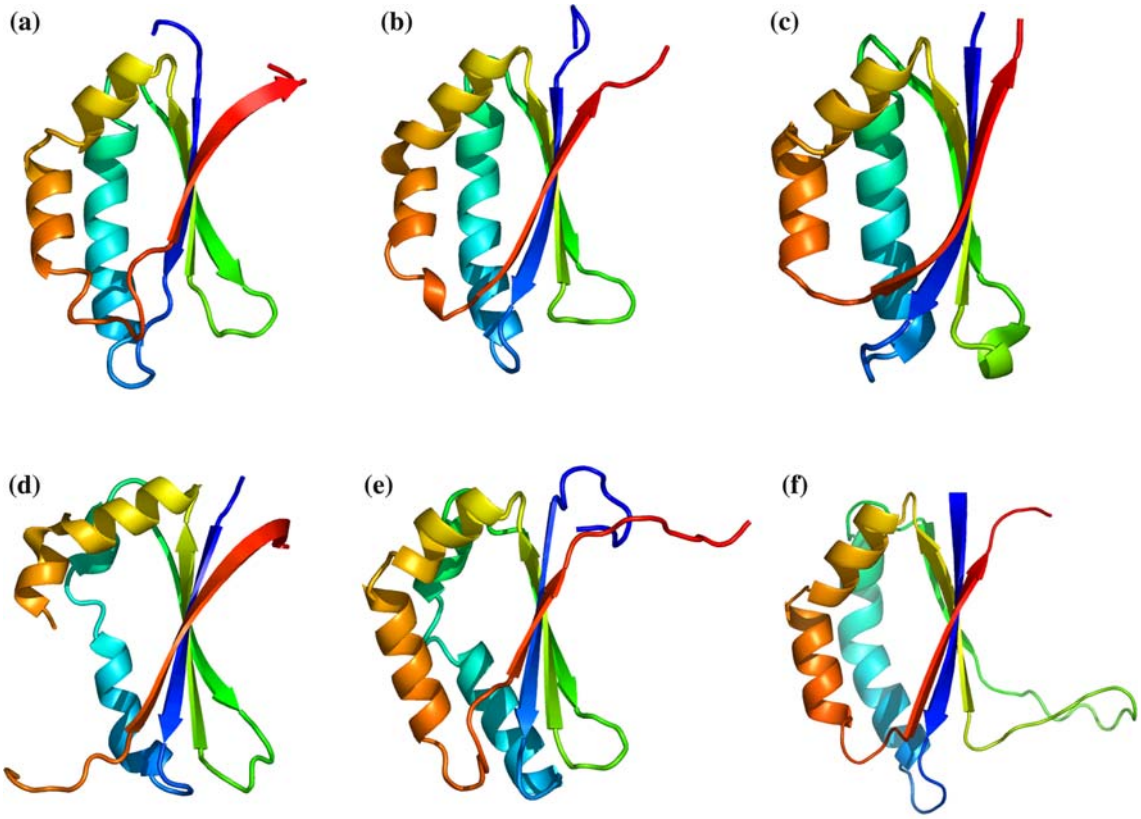

Figure 3. Ribbon representation of the individual subunit for Rv0793 from *M. tuberculosis* with top four DALI results and *E. coli* homolog YgiN. (a) 1Y0H, Rv0793; (b) 1X7V, Pa3566 from *P. aeruginosa*, structural genomics, unknown function; (c) 1Q8B, YjcS, *B. subtilis*, structural genomics, unknown function; (d) 1SQE, Pg130, *S. aureus*, structural genomics, unknown function; (e) *S. coelicolor* 1LQ9, ActVA-Orf6; (f) 1R6Y, *E. coli* YgiN monooxygenase protein.

Tyr72, and Arg86 [4]. These residues can be seen in the acetyl dithranol bound structure of ActVA-Orf6 (Figure 5d). Structurally, Arg86 in ActVA-Orf6, which has a flexible conformation [4], is in roughly the same location as Arg77 in Rv0793. Otherwise, there is little structural agreement of the residues in the cleft of Rv0793 and ActVA-Orf6 (Figure 4).

When the structure of Rv0793 is superimposed with that of YgiN (Figure 5e), structural agreement exists with the residues in the cleft; for example, His67 in Rv0793 and His75 from YgiN aligned. In addition, residues Val7, Tyr41, Glu56 and Tyr76 in the cleft of Rv0793 align structurally with residues Val4, Tyr40, Glu64, and Tyr84 in the cleft of YgiN, respectively. In YgiN, residues proposed to be important in catalysis are Tyr40, Met62, His75, Met81 and Tyr84 [11] (Figure 5e).

Co-crystallization and soaking of Rv0793 in the presence of various aromatic polyketide analogs, including acetyldithranol, tetracycline, 8-Anilino-1-naphthalenesulfonic acid, 5-Hydroxy-1,4-naphthoquinone (jugalone), and menadione, were carried out. No new density representing the analogs were seen with any of the data collected. Only acetate ions and water molecules were present in the cleft between the β-sheet and the α-helices. In the absence of acetate in the crystallization conditions, the cryoprotectant, ethylene glycol, was found in the cleft (data not shown).

Discussion

The overall structure of Rv0793 consists of a homodimer with each subunit having a ferrodoxin fold, one of the most common folds amongst protein structures. This fact is supported by the large number of similar structures identified in the DALI search. As a result, functional characterization of this protein is difficult.

There are no genes that are very close to the Rv0793 open reading frame on the same strand that may suggest its function. Rv0788, which is 3285 bases upstream is annotated as a probable phosphoribosylformylglycinamide synthase I (FGAM SYNTHASE I); whereas Rv0795, which is 1948 bases downsteam from Rv0793 is annotated as a putative transposase for insertion sequence elements. On the complementary strand, Rv0792c is annotated as a probable transcriptional regulatory protein (probably of the GNTR-family), while Rv0794c is annotated as a probable oxidoreductase and has been described as a mercuric reductase/glutathione reductase/dihydr-olipoamide dehydrogenase.

The most likely function for Rv0793 may be related to that of YgiN or ActVA-Orf6, which have the closest structural match for proteins with known function. ActVA-Orf6 of *S. coelicolor* has been shown to play a role in the biosynthesis of an aromatic polyketide, actinorhodin, whereas YgiN plays a role in quinol oxidation. Both ActVA-Orf6 [5] and YgiN [11] have been shown to act as monooxygenases that function without a cofactor. The top two matches from the DALI search, 1X7V (Pa3566) and 1Q8B (YjcS) in addition to the *E. coli* YgiN are proteins that belong to the same COG1359 as Rv0793. There are members of COG1359 for which no current structure exists. Moreover, structures have been revealed for the orthologs in COG1359 via structural genomics consortia where no functions have yet been ascribed. These proteins in COG1359 with similar tertiary structures but unknown function may represent a class of proteins that function as monooxygenases. Although a structure of Rv0793 in complex with a substrate analog has not been obtained, this does not rule out the possibility that Rv0793 may play a role as an aromatic hydrocarbon monooxygenase.

Ribbon representations of the A-subunit of Rv0793, the top four DALI matches and YgiN show the structural similarity among these proteins (Figure 3). Variations in the sizes of the clefts located between the β-sheets and the α-helices exist amongst the homologous structures. Interestingly, the hydrophobic cleft for polyketide biosynthesis, also known as the polyketide tunnel [30], has been shown to be of different depth depending on the protein and the species. This reflects the enzymes ability to accommodate the polyketides with varying lengths as found in nature. Variations in residues comprising the clefts of these homologous structures reveal a group of proteins having similar functions but most likely with different native substrates.

Many of the residues found in the cleft of Rv0793 are conserved in some of the structural matches and members of the COG1359 and the

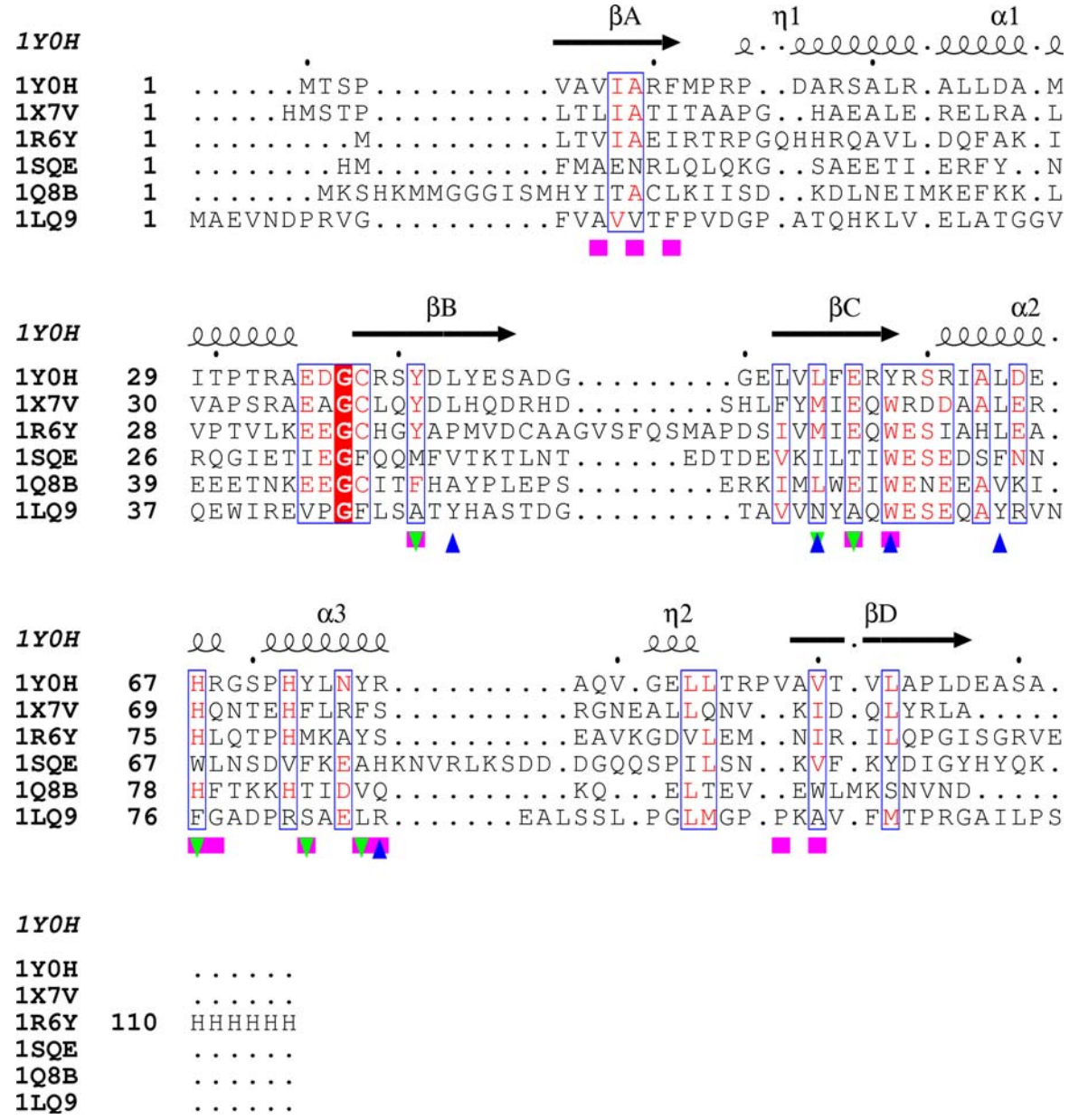

Figure 4. Sequence alignment of Rv0793 from *Mycobacterium tuberculosis* (1Y0H) based on structural imposition with other members of the COG1359. Secondary structure for Rv0793, labeled according to the topology diagram, (Figure 1), is depicted on top of the alignment. Residues located in cleft of Rv0793 are indicated by pink boxes. Residues located in the cleft of ActVA-Orf6 from *Streptomyces coelicolor* (1LQ9) are indicated by blue arrows. Residues located in the cleft of YgiN from *E. coli* (1R6Y) are indicated by green arrows. 1X7V, Pa3566, *Pseudomonas aeruginosa*. 1SQE.pdb, Pg130, *Staphylococcus aureus*. 1Q8B, Yjcs, *Bacillus subtilis*. R.m.s.d. between 1Y0H and 1X7V is 1.14 Å with 78 Cα atoms and 29 identical residues. R.m.s.d. between 1Y0H and 1R6Y is 1.45 Å with 75 Cα atoms and 21 identical residues. R.m.s.d. between 1Y0H and 1SQE is 1.38 Å with 51 Cα atoms and 5 identical residues. R.m.s.d. between 1Y0H and 1Q8B is 1.22 Å with 65 Cα atoms and 13 identical residues. R.m.s.d. between 1Y0H and 1LQ9 is 1.35 Å with 59 Cα atoms and 13 identical residues. (Swiss-PDB-Viewer alignment).

ABM family (Figure 4). His67 in Rv0793 is found in COG1359 members except for ActVA-Orf6 which has Phe76 at the equivalent position. In addition, Glu56 in Rv0793 occupies the same location in the cleft as Ala64 in ActVA-Orf6. ActVA-Orf6 is able to bind polyketides having

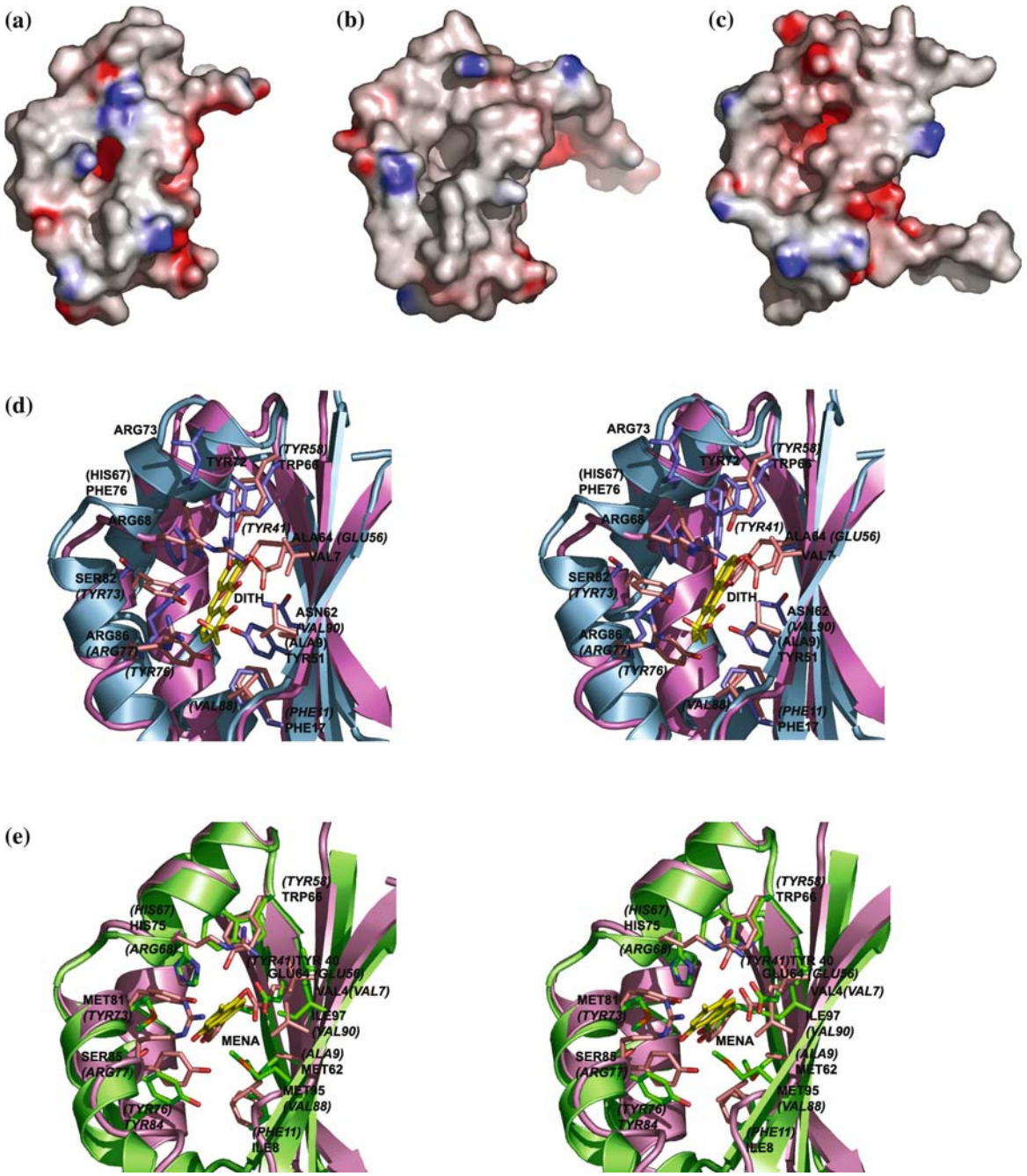

Figure 5. Cleft comparison of Rv0793 with ActVA-Orf6 and YgiN. Electrostatic surface representation showing cleft of (a) Rv0793 (1Y0H, A), (b) ActVA-Orf6 (1N5S, B), and (c) YgiN (obtained via email from Dr. Z. Jia and Ms. M. Adams, Queen's University, Canada). (d) A stereo image of residues found in the cleft of Rv0793 (ribbon in pink and resides in salmon) are overlaid with the residues in the cleft of ActVA-Orf6 (blue). Bound in the cleft of ActVA-Orf6 is acetyl-dithranol (yellow). (e) A stereo image of residues found in the cleft of Rv0793 (ribbon in pink and resides in salmon) are overlaid with the residues in the cleft of YgiN (green). Bound in the cleft of YgiN is menadione, MENA (yellow). Italic labels represent those equivalent residues found in Rv0793.

three rings. Rv0793 most likely could not accommodate such a large substrate due to the presence of Glu56 which makes the cleft shallower. These differences in cleft residues may reflect the differences in the native substrates for each monooxygenase.

A catalytic mechanism of ActVA-Orf6 monooxygenase has been proposed based on its structure with various substrate analogs bound [4]. In this mechanism, Trp66 is proposed to act as an electron donor stabilizing a hydroxyl group in the substrate. Deprotonation of the hydroxyl group occurs via the network of hydrogen bonds eminating from the indole N^{ε} H of Trp66. Arg86 and Tyr72 are also proposed to be involved in the deprotonation event, resulting in a C-6 carbanion that can react with molecular oxygen; Asn62 and Tyr51 stabilize the peroxy intermediate. The substrate is then protonated and dehydrated. Trp66 in ActVA-Orf6 is replaced by a Tyr59 in Rv0793; His67 in Rv0793 may serve to deprotonate and stabilize an intermediate. Arg86, also proposed to be involved in the deprotonation event in ActVA-Orf6, is structurally conserved as Arg77 in Rv0793.

Residues in the cleft of Rv0793 are more similar to those in YgiN than to those of ActVA-Orf6. In YgiN, Met62 is proposed to stabilize the peroxy intermediate, whereas Tyr40 and Tyr84 may be involved in catalysis. Although Met62 in YgiN is not conserved in Rv0793, Tyr40 and Tyr84 in YgiN are conserved as Tyr41 and Tyr76 in Rv0793. The conservation of Tyr residues in the cleft of Rv0793 may support their role in creating a hydrogen bonding network, promoting deprotonation of a quinol hydroxyl group. Thus, the residues present in the cleft support the role of monooxygenase for Rv0793. The exact substrate for this protein, however, remains to be determined.

The fact that at least three separate homologous proteins are present in *M. tuberculosis* each having ABM domains, suggests that there may be specific roles carried out by each protein. Elucidation of the structures of other *M. tuberculosis* proteins with the ABM domain fold may shed light on the function of this class of proteins, including Rv0793, in *M. tuberculosis*. If Rv0793 and its homologs are involved in antibiotic biosynthesis, this may provide a unique avenue for drug design targeted against *M. tuberculosis*.

Interestingly, other DALI hits included proteins that bind small antibiotics/drugs such as AcrB, a multi-drug transporter from *E. coli*. The homologous region is a domain located on the large periplasmic domain on the outer edge, opposite to the proposed pore [31]. Thus far, the only site proposed to be involved in drug binding has been in the central vestibule located at the membrane/periplasm interface [32]. It is possible that AcrB may also modify the drugs that it transports.

Alternatively, these ABM homologs in *M. tuberculosis* may act as superoxide scavengers as a defense mechanism against the host. When encapsulated by macrophages, the bacterium is subjected to both reactive oxygen species (ROS) and reactive nitrogen species (RNS) [33] as part of the host self-defense system. Inhibition of these host defenses leads to an increased rate of infection and death. The ABM homologs therefore may be included in the already large number of antioxidant enzymes found in this mycobacterium including two superoxide dismutases. SOD, the structure of the protein encoded by *sodC* has recently been solved [34]. Rv0793, and the members of the ABM family, are unique, however, in that unlike SOD, they do not require a metal cofactor to carry out their function.

More structural and functional studies are needed for the members of the ABM family in order to determine their functions and mechanism of action. Revealing the functions of these orphan proteins may be the next step for future structural genomics studies.

Acknowledgements

This work has been supported by the Canadian Institute for Health Research (CIHR). We would like to thank Ernst Bergman, Jonathan Parrish (Alberta Synchrotron Institute, University of Alberta) and James Holden, (BL8.3.1, Advanced Light Source, Berkeley National Laboratory) for the data collection. X-ray diffraction data were collected at beamline 8.3.1 of the Advanced Light Source (ALS) at the Lawrence Berkeley Lab, under an agreement with the Alberta Synchrotron Institute (ASI). The ALS is operated by the Department of Energy and supported by the National Institute 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) and the Alberta Heritage Foundation for Medical Research (AHFMR) and Western Economic Diversification, Canada. We also thank the Biochemistry DNA Core Lab for sequencing and the Institute for Biomolecular Design for carrying out the HLPC analysis. M.J.L is supported by fellowship scholarships with CHIR and AHFMR. M.N.G.J acknowledges support from the Canada Research Chairs Program.

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s10969-005-9004-6 and is accessible for authorized users.

References

1. Zahrt, T.C. (2003) *Microbes Infect*. **5**, 159–167.
2. Terwilliger, T.C. et al. (2003) *Tuberculosis (Edinb)* **83**, 223–249.
3. Kendrew, S.G., Federici, L., Savino, C., Miele, A., Marsh, E.N. and Vallone, B. (2000) *Acta Crystallogr. D Biol. Crystallogr*. **56**(Pt 4), 481–483.
4. Sciara, G. et al. (2003) *EMBO J*. **22**, 205–215.
5. Kendrew, S.G., Hopwood, D.A. and Marsh, E.N. (1997) *J. Bacteriol*. **179**, 4305–4310.
6. Shen, B. (2003) *Curr. Opin. Chem. Biol*. **7**, 285–295.
7. Rawlings, B.J. (1999) *Nat. Prod. Rep*. **16**, 425–484.
8. Hutchinson, C.R. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 3010–3012.
9. Reed, M.B., Domenech, P., Manca, C., Su, H., Barczak, A.K., Kreiswirth, B.N., Kaplan, G. and Barry, C.E. 3rd (2004) *Nature* **431**, 84–87.
10. Saxena, P., Yadav, G., Mohanty, D. and Gokhale, R.S. (2003) *J. Biol. Chem*. **278**, 44780–44790.
11. Adams, M.A. and Jia, Z. (2005) *J. Biol. Chem*. **208**, 8358–8363.
12. Shen, B. and Hutchinson, C.R. (1993) *Biochemistry* **32**, 6656–6663.
13. Shen, B. and Hutchinson, C.R. (1994) *J. Biol. Chem*. **269**, 30726–30733.
14. Rafanan, E.R. Jr., Le, L., Zhao, L., Decker, H. and Shen, B. (2001) *J. Nat. Prod*. **64**, 444–449.
15. Doublie, S. (1997) *Methods Enzymol*. **276**, 523–530.
16. Minor, Z.O.a.W. (1997) In *Methods in Enzymology Vol. 276: Macromolecular Crystallography, part A* (Eds., Carter, C.W., Jr. and R.M. Sweet), Academic Press, New York, pp. 307–326.
17. Terwilliger, T.C. and Berendzen, J. (1999) *Acta Crystallogr. D Biol. Crystallogr*. **55**(Pt 4), 849–861.
18. Terwilliger, T.C. (2003) *Methods Enzymol*. **374**, 22–37.
19. Pannu, N.S., Murshudov, G.N., Dodson, E.J. and Read, R.J. (1998) *Acta Crystallogr. D Biol. Crystallogr*. **54**, 1285–1294.
20. McRee, D.E. (1999) *J. Struct. Biol*. **125**, 156–165.
21. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) *Nucleic Acids Res*. **22**, 4673–4680.
22. Holm, L. and Sander, C. (1993) *J. Mol. Biol*. **233**, 123–138.
23. Kaplan, W. and Littlejohn, T.G. (2001) *Brief Bioinform*. **2**, 195–197.
24. Cohen, G.H. (1997) *J. Appl. Cryst*. **30**, 1160–1161.
25. O'Sullivan, O., Suhre, K., Abergel, C., Higgins, D.G. and Notredame, C. (2004) *J. Mol. Biol*. **340**, 385–395.
26. Gouet, P., Robert, X. and Courcelle, E. (2003) *Nucleic Acids Res*. **31**, 3320–3323.
27. Lee, B. and Richards, F.M. (1971) *J. Mol. Biol*. **55**, 379–400.
28. Orengo, C.A., Flores, T.P., Jones, D.T., Taylor, W.R. and Thornton, J.M. (1993) *Curr. Biol*. **3**, 131–139.
29. Bateman, A. et al. (2004) *Nucleic Acids Res*. **32**(Database issue), D138–D141.
30. Keatinge-Clay, A.T., Maltby, D.A., Medzihradszky, K.F., Khosla, C. and Stroud, R.M. (2004) *Nat. Struct. Mol. Biol*. **11**, 888–893.
31. Murakami, S., Nakashima, R., Yamashita, E. and Yamaguchi, A. (2002) *Nature* **419**, 587–593.
32. Yu, E.W., McDermott, G., Zgurskaya, H.I., Nikaido, H. and Koshland, D.E. Jr. (2003) *Science* **300**, 976–980.
33. Nathan, C. and Shiloh, M.U. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 8841–8848.
34. Spagnolo, L. et al. (2004) *J. Biol. Chem*. **279**, 33447–33455.