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

The molecular structure of Rv1873, a conserved hypothetical protein from *Mycobacterium tuberculosis*, at 1.38 Å resolution

Craig R. Garen,[a] Maia M. Cherney,[a] Ernst M. Bergmann[a,b] and Michael N. G. James[a,b]*

[a]Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada, and [b]Alberta Synchrotron Institute, University of Alberta, Edmonton, Alberta T6G 2E1, Canada

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

Received 20 September 2006
Accepted 7 November 2006

PDB Reference: Rv1873, 2d2y, r2d2ysf.

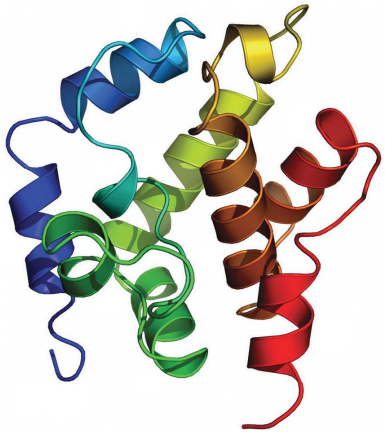
The X-ray crystal structure of the gene product encoded by open reading frame Rv1873 of *Mycobacterium tuberculosis* has been determined by single isomorphous replacement with anomalous scattering (SIRAS) phasing techniques at 1.38 Å resolution from monoclinic crystals with unit-cell parameters $a = 33.44$, $b = 31.63$, $c = 53.19$ Å, $\beta = 90.8^\circ$. The 16.2 kDa Rv1873 is a monomer that adopts a primarily α-helical fold with limited structural similarity to previously determined tertiary structures. It has been annotated as a conserved hypothetical protein of unknown function and is classified by the Clusters of Orthologous Groups (COG) database as belonging to COG5579. The three-dimensional structure of the Rv1873 gene product reveals limited similarity to a repeated motif that is found in a variety of other proteins. While not a novel fold, it serves as a model for orthologues predicted to be related by sequence and it is hoped that knowledge of the structure of Rv1873 will aid in determining a possible function for this protein.

1. Introduction

The aetiological agent of tuberculosis (TB), *Mycobacterium tuberculosis* (*Mtb*), is a highly successful human pathogen. Each year approximately eight million individuals are newly infected with TB and two to three million die from among the two billion carriers of the disease (Raviglione, 2003). To compound these grim statistics, human immunodeficiency virus (HIV) and *Mtb* form a deadly partnership in patients afflicted with both. TB has become a leading cause of death among HIV-positive patients and the latter are many more times likely to be infected with TB than uninfected persons (Corbett *et al.*, 2003). Moreover, noncompliance with recommended therapeutic regimens among all patients has given rise to TB strains that are resistant to existing anti-mycobacterial chemotherapies (Espinal, 2003).

In order to understand this organism more thoroughly on a molecular basis, the *M. tuberculosis* Structural Genomics Consortium (TBSGC) has been formed (Terwilliger *et al.*, 2003). As part of our contribution to this group, we have targeted many gene products from *Mtb* strain H37Rv for X-ray crystallographic structure determination. Some of these targets are annotated as conserved hypothetical proteins (CHPs) of unknown function (Cole *et al.*, 1998; Camus *et al.*, 2002). As a group, CHPs have been under-targeted by the TBSGC, leading to our selection of candidates from among these for crystal structure determination. The elucidation of novel folds among this group or the functional characterization of gene products of unknown function will lead to a better understanding of this pathogen at the molecular level and is very likely to reveal new drug targets.

Rv1873 has a predicted molecular weight of 16 207 Da and has no sequence homologue in the Protein Data Bank (PDB) structure database (Berman *et al.*, 2000). A *BLAST* search (Altschul *et al.*, 1990) using the predicted Rv1873 amino-acid sequence revealed that it has a possible conserved domain (Marchler-Bauer & Bryant, 2004) belonging to COG5579 in the Clusters of Orthologous Groups (COG) database (Tatusov *et al.*, 2001; Fig. 1). The predicted anno-



tation for COG5579 is a conserved hypothetical protein of unknown function. In order to gain some insight into the molecular function of Rv1873, its three-dimensional structure has been determined by single isomorphous replacement with anomalous scattering (SIRAS) phasing methods using a native crystal soaked in a solution of trimethyllead acetate. As no other structures of proteins from COG5579 have been determined at this time, we anticipate that Rv1873 will serve as the model structure for the group.

2. Materials and methods

PCR primers were designed using sequences for directional cloning of inserts into the Gateway cloning system (Invitrogen). They included homologous recombination sites and a region encoding a tobacco etch virus (TEV) protease cleavage site as published elsewhere (Biswal *et al.*, 2006). The gene-specific sequences used to amplify this open reading frame from *Mtb* strain H37Rv genomic DNA (Brosch *et al.*, 1998) are 1873F, 5′-ATGAAGTCAGCAAGCGACCCGTTCG-3′, and 1873R, 5′-CTATGTGGACCGCCAGTAATGCCACC-3′. The resulting PCR products were inserted into the pDONR-201 vector (Invitrogen) for DNA amplification. The PCR insert was subsequently cloned directionally into an expression vector (pDEST-15, Invitrogen) that encodes an N-terminal glutathione *S*-transferase (GST) fusion protein. The pGST-1873 expression plasmid was confirmed by restriction-endonuclease analysis and DNA sequencing (University of Alberta DNA Core Facility).

The fusion protein was expressed in *Escherichia coli* strain BL21 (DE3) pLysS (Novagen) by inducing expression of T7 RNA polymerase (Studier, 1991) with 0.5 m*M* isopropyl *β*-D-thiogalactopyranoside at 298 K. After 16 h of incubation, bacterial cell pellets were resuspended in phosphate-buffered saline (PBS) containing protease inhibitors in tablet form (Roche) and 1 m*M* dithiothreitol (DTT) pH 7.4 and then frozen at 193 K. Thawed pellets were lysed by sonication and clarified by centrifugation (30 000*g*, 30 min). The supernatant was loaded onto a 5 ml GSTrap FF glutathione Sepharose cartridge (GE Healthcare) and the fusion protein was competitively eluted with 10 m*M* reduced glutathione in 50 m*M* Tris–HCl pH 8.0. Proteolytic removal of the GST tag using recombinant TEV protease (Invitrogen) leaves a single extra glycine residue at the N-terminus of Rv1873. Following extensive dialysis in PBS to remove the glutathione, the cleaved protein mix was applied once again to an equilibrated GSTrap cartridge. A 1 ml HiTrap chelating cartridge

Table 1
Crystal parameters and data-collection statistics.

Values in parentheses are for the outermost resolution shell.

	TMLA peak	TMLA remote	Native
Space group	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters			
a (Å)	33.56	33.55	33.44
b (Å)	31.87	31.86	31.63
c (Å)	53.79	53.78	53.19
α (°)	90	90	90
β (°)	90.79	90.78	90.80
γ (°)	90	90	90
Z†	2	2	2
Temperature (K)	100	100	100
Detector	ADSC Q210	ADSC Q210	ADSC Q210
Wavelength (Å)	0.95116	0.97965	1.11588
Resolution (Å)	50–1.73 (1.79–1.73)	50–1.73 (1.79–1.73)	35–1.38 (1.43–1.38)
Unique reflections	11526 (1124)	11603 (1129)	21820 (1973)
Multiplicity	3.5 (3.5)	3.5 (3.5)	1.9 (1.6)
$I/\sigma(I)$	18.3 (2.46)	20.38 (2.73)	15.5 (1.95)
Completeness (%)	94.9 (93)	95.5 (93.7)	94.2 (85.8)
$R_{\rm merge}$	0.057 (0.505)	0.052 (0.464)	0.030 (0.402)
Phasing power‡, isomorphous	1.78 (1.17)	2.34 (1.59)	
Phasing power‡, anomalous	2.03 (1.55)	1.12 (0.83)	
$R_{\rm Cullis}$§, isomorphous	0.503 (0.574)	0.461 (0.524)	
$R_{\rm Cullis}$§, anomalous	0.739 (0.758)	0.796 (0.850)	

† Z = number of molecules in the unit cell. ‡ Phasing power = ⟨r.m.s. heavy-atom structure factor⟩/⟨r.m.s. lack of closure⟩. § $R_{\rm Cullis} = \langle\varepsilon\rangle/\langle|F_{\rm PH} - F_{\rm H}|\rangle$.

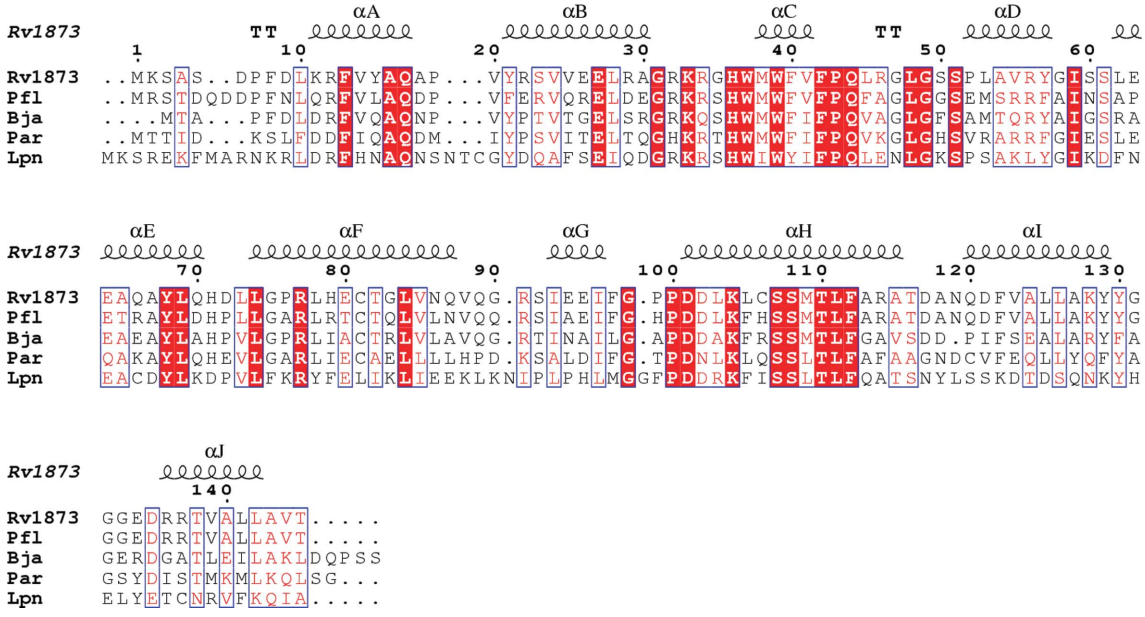


Figure 1
Sequence comparison of selected members of COG5579. Abbreviations used, along with accession numbers for the sequences, are as follows: Rv1873, *M. tuberculosis* strain H37Rv (gi:15609010); Pfl, *Pseudomonas fluorescens* strain PfO-1 (gi:77383958); Bja, *Bradyrhizobium japonicum* strain USDA 110 (gi:27352459); Par, *Psychrobacter arcticum* strain 273-4 (gi:71038919); Lpn, *Legionella pneumophila* subspecies *pneumophila* strain Philadelphia 1 (gi:53753436). Note that only the first 151 from a total of 324 residues are shown for Lpn for brevity. Gaps are represented by '.', while identical residues are white on solid red background and conserved residues are red on white. Each of the sequences is separately numbered and dots serve as a landmark for every ten residues of Rv1873. Secondary-structure elements for Rv1873 are shown above as coils. This figure was produced using *ESPript* v.2.2 (Gouet *et al.*, 2003).

Table 2
Refinement stastics and model quality.

Refinement resolution	35–1.38
No. of reflections used for refinement	20717
No. of reflections used for R_{free}	1091
R_{work}†/R_{free}‡	0.167/0.200
No. of non-H atoms	1287
Protein	1109
Water	161
Sulfate ion	5
Glycerol	12
Mean *B* factor (Å^2)	15.2
Mean *B* factor, protein atoms (Å^2)	13.0
Mean *B* factor, water molecules (Å^2)	28.9
R.m.s.d. bond lengths (Å)	0.008
R.m.s.d. bond angles (°)	1.17
Residues in most favoured regions of φ–ψ plot (%)	92.4
Residues in additionally allowed regions of φ–ψ plot (%)	7.6

† $R_{work} = \sum \big||F_o| - |F_c|\big| / \sum |F_o|$ for the 95% of the reflection data used in refinement. ‡ $R_{free} = \sum \big||F_o| - |F_c|\big| / \sum |F_o|$ for the remaining 5% of the reflection data excluded from refinement.

(GE Healthcare) charged with nickel was attached in tandem to remove the N-terminally His_6-tagged TEV protease. Fractions containing the flowthrough were retained and then subjected to size-exclusion chromatography for final purification and buffer exchange to 10 m*M* Tris–HCl pH 8.0, 50 m*M* NaCl and 1 m*M* DTT. Purified Rv1873 was concentrated using an Amicon Ultra 5 kDa cutoff (Millipore) and the final protein concentration was determined to be 10 mg ml^{-1} using a Bradford Protein Assay Kit (BioRad) with BSA (Pierce) as a standard. SDS–PAGE analysis of the concentrated protein revealed a single band of approximately 16 kDa as expected for Rv1873. This band was further confirmed to be Rv1873 by employing MALDI–TOF MS analysis of the peptides generated by trypsin digestion (Institute of Biomolecular Design, University of Alberta).

Robotic crystal screening using the Hydra-Plus-One system (Matrix Discoveries) with Intelliplates (Hampton Research) and sparse-matrix crystal screening techniques (Jancarik & Kim, 1991) in 96-well format identified native crystallization conditions at room temperature. The robot was programmed to dispense a 100 µl reservoir and to set up 1 µl sitting drops with a 1:1 ratio of Rv1873 to reservoir solution. Native monoclinic crystals belonging to space group $P2_1$ (Table 1) were grown in 0.1 *M* MES pH 6.5, 38% ammo-

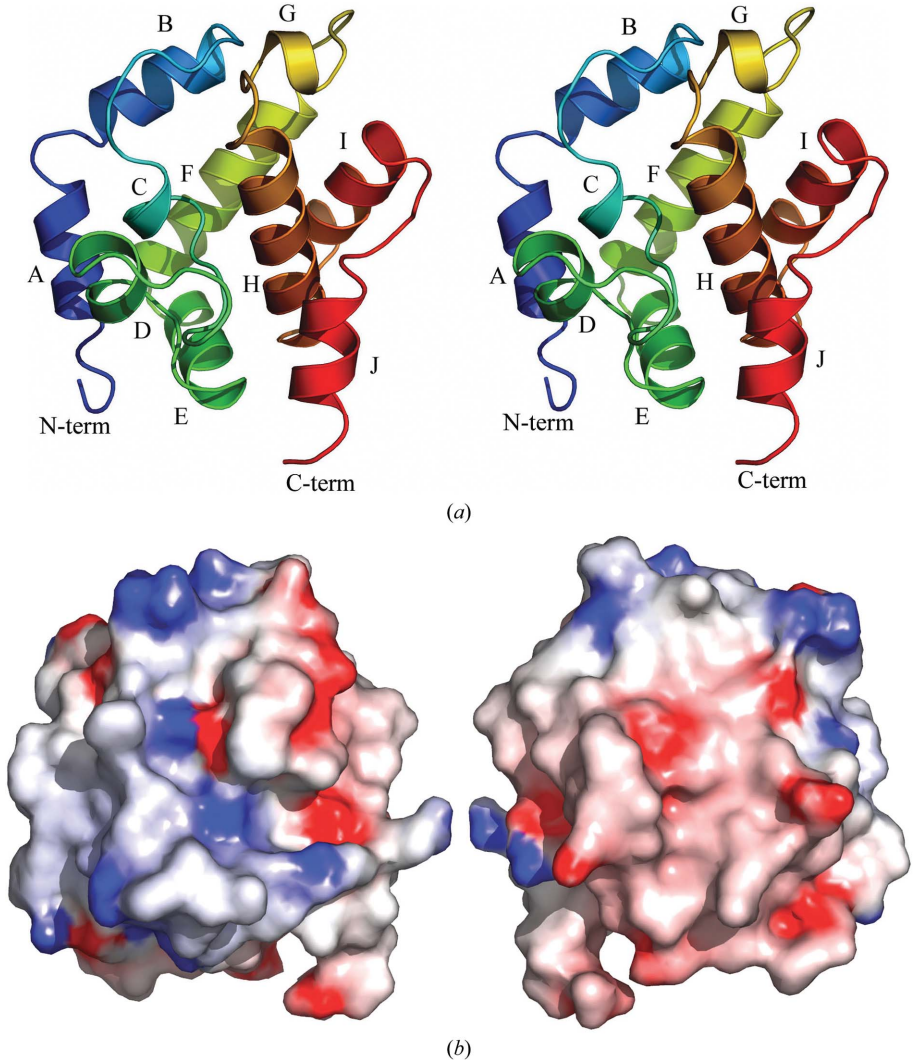


Figure 2
The overall structure and surface of Rv1873. (*a*) Stereo representation of the Rv1873 molecule present in the crystallographic asymmetric unit. The colouring is blue at the N-terminus to red at the C-terminus. The ten α-helices comprising the secondary structure are labelled *A*–*J*. (*b*) The electrostatic surface potential for Rv1873 drawn with the program *GRASP* (Nicholls, 1991). Positive and negative surface regions are blue and red, respectively, for each half of the protein (molecule rotated about the *y* axis). The leftmost of these is the same depiction as the stereo image shown in (*a*).

nium sulfate and 5%(*v*/*v*) ethylene glycol at 298 K. Heavy-atom soaking with 40 m*M* trimethyllead acetate for 30 min yielded the crystal from which heavy-atom data were collected.

Crystals for X-ray data collection were first rinsed in cryoprotectant [25%(*v*/*v*) glycerol in mother liquor] and then flash-frozen in liquid nitrogen. Native and trimethyllead acetate (TMLA) data sets were collected at beamline 8.3.1 at the Advanced Light Source at Lawrence Berkeley National Laboratory. Integrated intensity data were reduced, merged and scaled using the *HKL*-2000 program suite (Otwinowski & Minor, 1997). Crystal parameters, data-collection statistics and data processing are summarized in Table 1.

The program *SOLVE* (Terwilliger & Berendzen, 1999) was used to locate two lead sites using data collected at peak and low remote wavelengths. The heavy-atom parameters were refined and phase calculations were carried out using *SHARP* (de La Fortelle & Bricogne, 1997). The initial phases were improved by solvent flattening with *SOLOMON* (Abrahams & Leslie, 1996) implemented in *SHARP*, followed by initial model building into the resulting electron-density map with *ARP*/*wARP* (Perrakis *et al.*, 1999). Manual fitting of side chains into the electron density was accomplished using the program *XFIT* from the *XtalView* package (McRee, 1999) coupled with refinement using *REFMAC* 5.2.0005 (Murshudov *et al.*, 1999) from the *CCP*4 suite of programs (Collaborative Computational Project, Number 4, 1994). Atomic *xyz* coordinates and the isotropic *B* factors were likewise refined using *REFMAC* 5.2.005 and subsequent refinement using native data produced a final model of 1.38 Å resolution. Refinement statistics and model-quality parameters are summarized in Table 2. All figures were prepared using *PyMOL* (DeLano, 2003) except where noted otherwise.

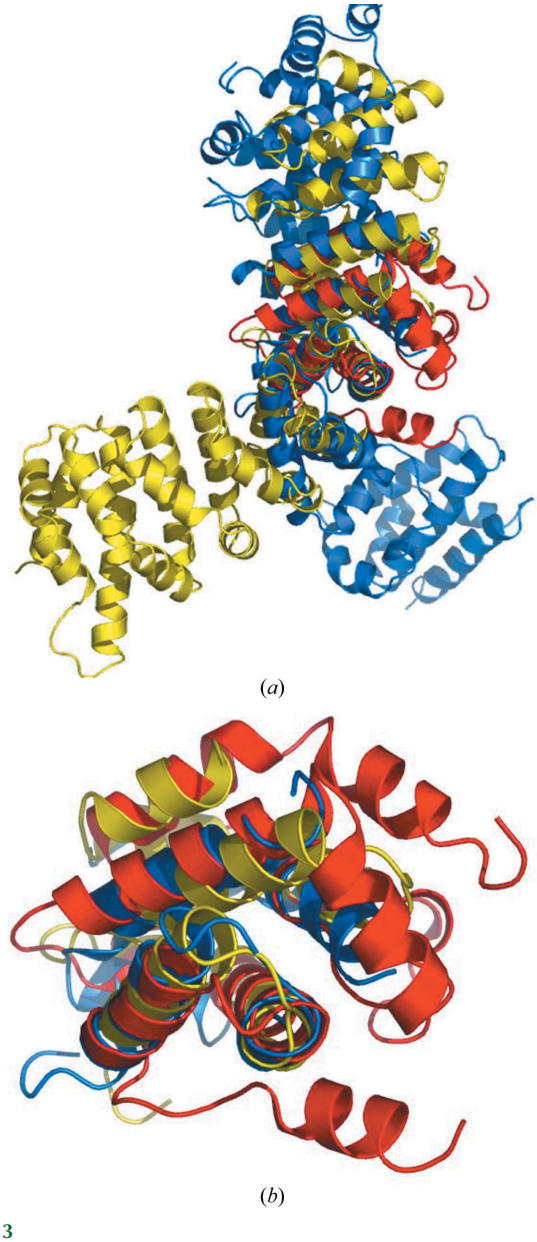


Figure 3
(*a*) Cartoon representation of the global molecular superposition of Rv1873 (PDB code 2d2y; red), molecule *A* of *S. cerevisiae* α-karyopherin (PDB code 1bk5; blue) and *S. cerevisiae* vacuolar ATP synthase (PDB code 1ho8; yellow). (*b*) A close-up representation depicting the alignment of helices in the three molecules. Molecular colouring and orientation are the same as in (*a*).

3. Results and discussion

The final model has been refined at 1.38 Å resolution to a crystallographic *R* factor and $R_{\rm free}$ of 0.167 and 0.201, respectively. A total of 139 of the 145 expected residues are present; there is no electron density for residues Gly−1 to Ser5 at the N-terminus. The Pro99-Pro100 peptide bond adopts a *cis* conformation. There are 160 ordered water molecules that were added during the refinement. The monoclinic crystals have a single molecule per asymmetric unit and a Matthews coefficient $V_{\rm M}$ of 1.77 Å^3 Da^{-1} (30.69% solvent content). The root-mean-square deviation from standard geometry is 0.008 Å and 1.17° for bond lengths and angles, respectively. Main-chain torsion angles conform to standard values for nonglycine/nonproline residues, with 92.4% falling within the most favoured regions of the φ–ψ Ramachandaran plot and 7.6% in the additionally allowed regions as determined by *PROCHECK* (Laskowski *et al.*, 1993).

The crystal structure of the Rv1873 monomer consists of ten α-helical segments labelled *A*–*J* that fold to form a single-domain protein (Fig. 2*a*). A surface representation of the molecule produced by *GRASP* (Nicholls, 1991) shows the electrostatic surface for Rv1873 (Fig. 2*b*). Based upon its helical structure, Rv1873 can be grouped in the all-α class of the Structural Classification Of Proteins database (SCOP v.1.69; Murzin *et al.*, 1995). A *DALI* search (Holm & Sander, 1997) using PDB coordinates for Rv1873 (PDB code 2d2y) revealed 117 structurally similar proteins with a *Z* value greater than 2.0. Only 16 of these have a *Z* score greater than 4.0, the two highest being *Saccharomyces cerevisiae* α-karyopherin (PDB code 1bk5; *Z* score 4.9 for 96 pairs of aligned C$^\alpha$ atoms; r.m.s.d. = 4.1 Å; Conti *et al.*, 1998) and *S. cerevisiae* vacuolar ATP synthase (PDB code 1ho8; *Z* score 4.8 for 96 pairs of aligned C$^\alpha$ atoms; r.m.s.d. = 3.3 Å; Sagermann *et al.*, 2001). These structures were compared by superposition, revealing a similar fold among the three proteins (Fig. 3) and suggesting that the Rv1873 fold class is a right-handed superhelix as are 1bk5 and 1ho8. Three right-handed turns of the Rv1873 superhelix coil around a core of hydrophobic solvent-inaccessible residues roughly circumscribed by α*E*, α*F*, α*H* and α*I*. As seen in Fig. 3, the Rv1873 fold is structurally similar to those of 1bk5 and 1ho8 based upon the alignment of helices, but it is nonetheless only a single copy of the right-handed superhelical motif that is repeated multiple times in the other two molecules. An extended poorly structured loop extending from the C-terminus of α*B* through to the N-terminus of α*D* also exists in Rv1873 that appears to be an insertion when compared with 1bk5 and 1ho8. Interestingly, this loop is rather highly conserved based upon the amino-acid sequences of other COG5579 members. In spite of the limited structural similarity predicted by

Table 3
Small hydrogen-bonded structural elements of Rv1873 (PDB code 2d2y) as determined by *MSDmotif* (Golovin *et al.*, 2004).

Structural motif	Motif sequence and location
α-β motif	LKRFV[14], QAPVY[21], VYRSV[24], MWFVF[42], IEEIF[97], LFARA[115]
Asx motif	DRRTV[139]
β-turn type il	YGGG[133]
β-turn type ir	DPFD[10], WMWF[42]
β-turn type iir	LRGL[48]
Nest type lr	DLK[11], GLG[49]
Nest type rl	FDL[10], AGR[32], YGG[132]
ST motif	SPLAV[55], SLEEA[65]
ST staple	PVYRS[23], LHECT[82], LKLCS[107], CSSMT[110]
Schellmann loop (six-residue)	ELRAGR[32]

DALI, no functional inferences could be made based upon this information.

The PDB coordinates for Rv1873 were used to search the EMBL–EBI macromolecular structure database motif server (*MSDmotif*; Golovin *et al.*, 2004) in an attempt to identify small three-dimensional hydrogen-bonded motifs. The results of this analysis are summarized in Table 3, revealing that 16 out of 23 of these hydrogen-bonded motifs are common features of either the N- or C-termini of helices, as might be expected for a primarily α-helical protein. The six-residue Schellmann loop is found at the C-terminus of α-helix *B*, imparting a sharp turn to the following loop, which was shown to deviate from the alignment in Fig. 3. The Protein Interfaces, Surfaces and Assemblies (*PISA*) server (Krissinel & Henrick, 2005) at the MSD was queried using the PDB coordinates file for Rv1873. The single molecule in the asymmetric unit of the Rv1873 crystal has six potential interfaces with symmetry-related molecules. However, the most extensive interface surface area is only 341.0 Å^2, involving 8% of the total number of residues (11 out of 145). This interface was given a complexation likelihood score of 0.0, suggesting that Rv1873 is monomeric. This analysis is backed up by the observation that Rv1873 elutes from a calibrated size-exclusion column at the molecular weight expected for the monomer.

The Catalytic Site Atlas server (*CSA*; Porter *et al.*, 2004) was queried using Rv1873 atomic coordinates to search for any correlation with known arrangements of catalytic site side chains gleaned from the PDB. A single possible active-site constellation was identified using the PDB entry for the structure of bovine epithelial nitric oxide synthase (eNOS; Raman *et al.*, 1998; PDB code 3nse) as a template. Cys186, Arg189, Trp358 and Glu363 of 3nse were shown to correlate with an r.m.s.d. of 1.23 Å with Cys81, Arg77, Trp37 and Asp101 of Rv1873. Aside from Cys81, which is present in all but one of the compared amino-acid sequences (Fig. 1), the other members of this potential active site are conserved among all members of COG5579. However, when looking at the position of the side chains in Rv1873, it is clear that this particular assemblage of side chains could never represent an active site or binding site: Trp37 is a buried side chain and Asp101 is on the other side of the molecule compared with Arg77 and Cys81. Careful analysis of this fold to attempt to identify possible molecular function of Rv1873 cannot immediately be inferred based on the structural information presented here.

X-ray diffraction data were collected at beamline 8.3.1 of the Advanced Light Source (ALS) at Lawrence Berkeley Laboratory 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), the Alberta Heritage Foundation for Medical Research (AHFMR) and Western Economic Diversification of the Canadian Government. Many thanks to the members of the James laboratory for helpful discussions and critical reading of the manuscript. MNGJ is supported by the Canadian Institutes of Health Research, the Alberta Heritage Foundation for Medical Research and is a Canada Institutes of Health Research (CIHR) Canada Research Chair.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Biswal, B. K., Garen, G., Cherney, M. M., Garen, C. & James, M. N. G. (2006). *Acta Cryst.* F**62**, 136–138.
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.
Camus, J.-C., Pryor, M. J., Medigue, C. & Cole, S. T. (2002). *Microbiology*, **148**, 2967–2973.
Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Conti, E., Uy, M., Leighton, L., Blobel, G. & Kuriyan, J. (1998). *Cell*, **94**, 193–204.
Corbett, E. L., Watt, C. J., Walker, N., Maher, D., Williams, B. G., Raviglione, M. C. & Dye, C. (2003). *Arch. Int. Med.* **63**, 1009–1021.
DeLano, W. L. (2003). *PyMOL*. DeLano Scientific LLC, San Carlos, CA, USA.
Espinal, M. A. (2003). *Tuberculosis*, **83**, 44–51.
Golovin, A. *et al.* (2004). *Nucleic Acids Res.* **32**, D211–D216.
Gouet, P., Robert, X. & Courcelle, E. (2003). *Nucleic Acids Res.* **31**, 3320–3323.
Holm, L. & Sander, C. (1997). *Nucleic Acids Res.* **25**, 231–234.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Krissinel, E. & Henrick, K. (2005). *CompLife 2005*, edited by M. R. Berthold, pp. 163–174. Berlin: Spinger-Verlag.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Marchler-Bauer, A. & Bryant, S. H. (2004). *Nucleic Acids Res.* **32**, W327–W331.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). *J. Mol. Biol.* **247**, 536–540.
Nicholls, A. (1991). *Proteins*, **11**, 281–296.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. J. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Porter, C., Bartlett, G. & Thornton, J. (2004). *Nucleic Acids Res.* **32**, D129–D133.
Raman, C., Li, H., Martasek, P., Kral, V., Masters, B. & Poulos, T. (1998). *Cell*, **95**, 939–950.
Raviglione, M. C. (2003). *Tuberculosis*, **83**, 4–14.
Sagermann, M., Stevens, T. H. & Matthews, B. W. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 7134–7139.
Studier, F. W. (1991). *J. Mol. Biol.* **219**, 37–44.
Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D. & Koonin, E. V. (2001). *Nucleic Acids Res.* **29**, 22–28.
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.