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

Expression, purification and preliminary X-ray analysis of the C-terminal domain of an arginine repressor protein from *Mycobacterium tuberculosis*

George J. Lu, Craig R. Garen, Maia M. Cherney, Leonid T. Cherney, Cecilia Lee and Michael N. G. James*

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

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

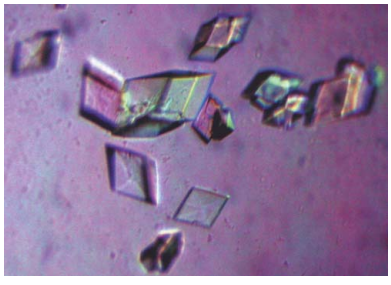
Received 14 August 2007
Accepted 20 September 2007

The gene product of an open reading frame Rv1657 from *Mycobacterium tuberculosis* is a putative arginine repressor protein (ArgR), a transcriptional factor that regulates the expression of arginine-biosynthetic enzymes. Rv1657 was expressed and purified and a C-terminal domain was crystallized using the hanging-drop vapour-diffusion method. Diffraction data were collected and processed to a resolution of 2.15 Å. The crystals belong to space group *P*1 and the Matthews coefficient suggests that the crystals contain six C-terminal domain molecules per unit cell. Previous structural and biochemical studies on the arginine repressor proteins from other organisms have likewise shown the presence of six molecules per unit cell.

1. Introduction

Tuberculosis (TB) is a well known chronic respiratory disease caused by the pathogenic bacterium *Mycobacterium tuberculosis* (*Mtb*). Currently, *Mtb* infects one-third of the world's population and approximately two million deaths occur annually (World Health Organization, 2003). Despite the existence of multiple anti-mycobacterial chemotherapies, drug-resistant strains often appear owing to noncompliance with the recommended regimens. Under these circumstances, the *Mtb* structural genomic consortium was formed in 2000 (http://www.doe-mbi.ucla.edu/TB) with the goal of providing a structural basis for the development of novel effective therapeutics for tuberculosis (Goulding *et al.*, 2002; Smith & Sacchettini, 2003; Terwilliger *et al.*, 2003). The complete genome of the best-characterized strain of *Mtb*, H37Rv, has been sequenced and many of the open reading frames for its proteins have been annotated (Cole *et al.*, 1997; Camus *et al.*, 2002).

As members of the *Mtb* structural genomics consortium, we have conducted molecular cloning and preliminary X-ray analysis on the arginine repressor protein (Rv1657). Presumably, this protein acts a molecular sensor of the intracellular arginine concentration. In response to the binding of arginine, Rv1657 binds to the corresponding DNA operator site and inhibits the transcription of the operon containing most of the arginine-biosynthetic enzymes. Thus, Rv1657 regulates arginine biosynthesis in *Mtb* *via* a feedback-inhibition mechanism. The C-terminal domain of Rv1657 is responsible for the binding of arginine and for the oligomerization of the protein.



2. Experimental methods

2.1. Cloning, expression and purification

The entire genome of the H37Rv strain was cloned into a bacterial artificial chromosome (BAC) library at L'Institut Pasteur (Brosch *et al.*, 1998; Gordon *et al.*, 1999). This library was used as the template to amplify the Rv1657 protein by PCR. Primers were designed for directional cloning of inserts into the Gateway cloning system (Invitrogen). The primer sequences were Rv1657F (5′-GGGACAAGTTTGTACAAAAAAGCAGGCTCC*GAAAACCTGTATTTTCAGGG*ATGAGCCGCGCCAAGGCCGCGC-3′) and Rv1657R (5′-GGGACCACTTTGTACAAGAAAGCTGGGTCTTACCGAAGG-

TTCTCGAACATGCCGGCC-3′). The open reading frame encoding residues 1–170 of Rv1657 was cloned into an expression plasmid with an amino-terminal glutathione-*S*-transferase (GST) fusion partner (pDEST-15, Invitrogen). The expression plasmid was confirmed by DNA-sequence analysis (DNA core facility, Department of Biochemistry, University of Alberta, Canada).

GST-Rv1657 was expressed in *Escherichia coli* BL21(DE3) pLysS cells (Novagen). Incubation of the transformed cells at 310 K was continued until the $OD_{600\,nm}$ reached 0.8–1.0. Subsequently, the temperature was shifted to 295 K and the medium was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 m*M*. After overnight incubation, the cells were harvested by centrifugation at 2150*g* for 15 min. Bacterial pellets were resuspended in phosphate-buffered saline (PBS) containing Complete protease inhibitor (Roche) and 10 µg ml^{-1} hen egg-white lysozyme (Sigma). For purification, the cells were lysed by freeze–thaw and then subjected to ultrasonication in the resuspension buffer. The lysate was cleared by centrifugation (30 min, 20 000*g*) and the supernatant was loaded onto a 5 ml GSTrap glutathione Sepharose column (GE Healthcare) pre-equilibrated with PBS. The GST-Rv1657 fusion protein was competitively eluted from the column with 10 m*M* reduced glutathione (Sigma) in 50 m*M* Tris–HCl pH 8.0, 2 m*M* DTT and 0.02% NaN_3. The GST tag and the N-terminal recombination site were removed by proteolytic cleavage using recombinant tobacco etch virus protease (rTEV, Invitrogen). The site recognized by rTEV protease is encoded in the forward primer (italicized in the sequence given) and leaves a single glycine residue in the P1′ position following complete proteolysis. After dialysis against PBS, the cleaved protein mixture was once again loaded onto a GSTrap column and the flowthrough fractions containing Rv1657 were dialyzed against 10 m*M* Tris–HCl pH 7.4 and 100 m*M* NaCl. The resulting solution was concentrated using an Amicon Ultra (5 kDa cutoff; Millipore). The final protein concentration reached a maximum of 12 mg ml^{-1}. The whole process of purification was performed at 277 K and the result of each step was monitored by 16% SDS–PAGE (Fig. 1).

2.2. Crystallization

The initial screening of crystallization conditions for native full-length Rv1657 was performed at 295 K using the sitting-drop vapour-diffusion technique in 96-well Intelli-plates (Hampton Research). Crystal Screens I and II and the Index Screen (Hampton Research) were applied using equal volumes (0.5 µl) of protein and precipitating solutions. Over two months, initial crystals were obtained from a variety of conditions. After optimization of the best screening conditions, X-ray diffraction-quality crystals were grown in hanging drops in 24-well VDX plates (Hampton Research) containing 1 µl protein solution at 10 mg ml^{-1} and 0.5 µl precipitating solution and the drops were equilibrated against 1 ml precipitating solution (20% PEG 10 000, 0.1 *M* HEPES pH 7.5). Crystals grew to dimensions of 100 × 100 × 50 µm over two months.

SDS–PAGE analysis of the crystals suggested that cleavage occurred during the crystallization process (Fig. 2). In order to identify the cleavage site, *ClustalW* (Thompson *et al.*, 1994) was used to perform sequence alignment (Fig. 3) between Rv1657 and three other ArgR proteins of known structure: those from *Bacillus stearothermophilus* (PDB code 1b4a; 31% identity; Ni *et al.*, 1999), *B. subtilis* (PDB code 1f9n; 29% identity; Dennis *et al.*, 2002) and *E. coli* (PDB codes 1xxa and 1aoy; 29 and 25% identity, respectively;

Table 1
Crystal parameters and data-collection statistics for native data sets from Rv1657.

Values in parentheses are for the highest resolution shell.

Space group	*P*1
Unit-cell parameters	
a (Å)	53.22
b (Å)	57.24
c (Å)	57.33
α (°)	66.19
β (°)	62.21
γ (°)	82.00
No. of molecules per ASU	6
Data collection	
Temperature (K)	100
Detector	Quantum Q330
Wavelength (Å)	0.97848
Resolution (Å)	50.00–2.15 (2.23–2.15)
Unique reflections	28194 (2498)
Multiplicity	1.8 (1.7)
$I/\sigma(I)$	11.77 (2.26)
Completeness (%)	95.0 (84.0)
R_{sym}†	0.061 (0.252)

† $R_{sym} = \sum_{\mathbf{h}}\sum_i |I_{\mathbf{h}i} - \langle I_{\mathbf{h}}\rangle| / \sum_{\mathbf{h}}\sum_i \langle I_{\mathbf{h}}\rangle$, where $I_{\mathbf{h}i}$ is the *i*th observation of reflection **h** and $\langle I_{\mathbf{h}}\rangle$ is the weighted average intensity for all observations *i* of reflection **h**.

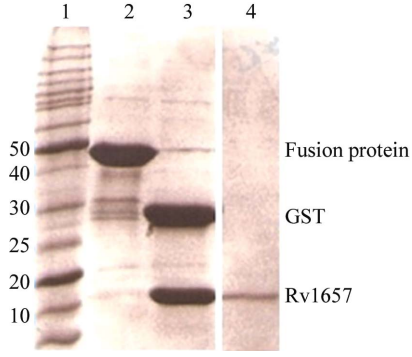


Figure 1
16% SDS–PAGE analysis of the purification stages of full-length Rv1657 (17 kDa). Lane 1, protein molecular-weight standards (kDa); lane 2, GST-Rv1657 fusion protein; lane 3, fusion protein cleaved with TEV protease; lane 4, purified Rv1657 fraction from the flowthrough fraction from the GSTrap column.

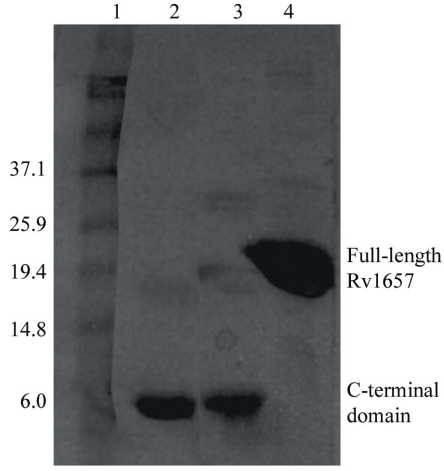


Figure 2
20% SDS–PAGE analysis of the cleavage during crystallization. Lane 1, protein molecular-weight standards (kDa); lane 2, dissolved crystals that had been grown at 295 K for two months; lane 3, protein solution corresponding to that in crystallization drops stored at 277 K for two months; lane 4, a second Rv1657 preparation containing full-length protein also stored at 277 K for two months. The sample in lane 4 is overloaded and appears to be greater than 19.4 kDa; however, the molecular weight of this sample was shown to be 17 456.7 Da by mass spectrometry.

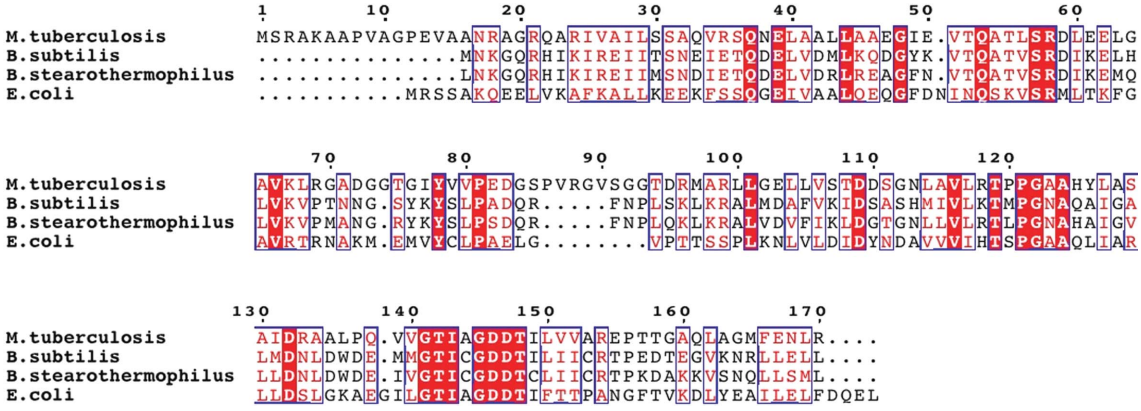


Figure 3
Sequence alignment of arginine repressor proteins from four different organisms. The numbering above the alignment corresponds to the sequence of Rv1657 from *M. tuberculosis*. Sequence alignment was performed using the program *ClustalW* (Thompson *et al.*, 1994) and the figure was generated using the program *ESPript* (Gouet *et al.*, 1999).

Sunnerhagen *et al.*, 1997; Van Duyne *et al.*, 1996). By comparing the sequences and known structures, we predicted Rv1657 to potentially be separated into two domains by virtue of a long linker containing residues 81–91. To confirm the content of the crystals, they were analyzed by MALDI–TOF mass spectrometry (Institute of Biomolecular Design, University of Alberta, Canada) and found to be composed of a single 8129.8 Da molecule. The expected molecular weight for Gly92–Arg170, 8130.24 Da, corresponds well to the measured value, which suggests that it is this domain of the protein that is most likely to be present in our crystals.

2.3. Data collection

Crystals for synchrotron data collection were first rinsed in cryoprotectant (25% glycerol in mother liquor) and then flash-cooled in liquid nitrogen. Native data sets were collected on beamline 7-1 at the Stanford Synchrotron Radiation Laboratory (SSRL) and on beamline 8.3.1 at the Advanced Light Source (ALS) at the Lawrence Berkeley National Laboratory. Raw data were reduced, integrated and scaled using the *HKL*-2000 program suite (Otwinowski & Minor, 1997).

3. Results

Native data sets were collected from Rv1657 C-terminal domain crystals (Fig. 4). Crystallographic statistics of the native data are summarized in Table 1. According to the measured molecular weight of the Rv1657 C-terminal domain, 8129.8 Da, the Matthews coefficient (Matthews, 1968) is calculated to be 2.89 $Å^3$ Da^{-1} and the solvent content is 57.48%. The result assumes the presence of six molecules of the C-terminal domain of Rv1657 per asymmetric unit, in agreement with existing biological evidence that a hexamer is the active biological unit (Charlier *et al.*, 1992). The hexamer constitutes the asymmetric unit in the arginine repressor structures determined from other organisms (PDB codes 1f9n, 1b4a and 1xxa). Crystallization of the full-length transcriptional factor is currently being pursued, as is a molecular-replacement solution for the structure of the C-terminal domain. Results for both will be published elsewhere in the future.

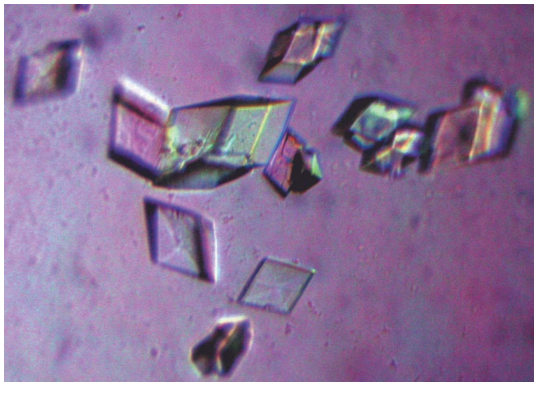


Figure 4
Rv1657 crystals. The approximate dimensions of the diffraction-quality crystals were 100 × 100 × 50 µm and the crystals were grown in 20% PEG 10 000, 0.1 *M* HEPES buffer pH 7.5 over two months.

X-ray diffraction data were collected both on beamline 7-1 at the Stanford Synchrotron Radiation Laboratory (SSRL) and on beamline 8.3.1 at the Advanced Light Source (ALS) at Lawrence Berkeley National Laboratory under agreements with the Alberta Synchrotron Institute (ASI). The ALS is operated by the Department of Energy and supported by the National Institutes of Health. Beamline 8.3.1 was funded by the National Science Foundation, the University of California and Henry Wheeler. The ASI synchrotron-access program is supported by grants from the Alberta Science and Research Authority (ASRA) and the Alberta Heritage Foundation for Medical Research (AHFMR). Research in the laboratory of MNGJ is supported by Alberta Heritage Foundation for Medical Research (AHFMR); MNGJ is the holder of a Canada Research Chair in Protein Structure and Function.

References

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.
Charlier, D., Roovers, M., Van Vliet, F., Boyen, A., Cunin, R., Nakamura, Y., Glansdorff, N. & Pierard, A. (1992). *J. Mol. Biol.* **226**, 367–386.
Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.
Dennis, C. A., Glykos, N. M., Parsons, M. R. & Phillips, S. E. V. (2002). *Acta Cryst.* D**58**, 421–430.
Gordon, S. V., Brosch, R., Billault, A., Garnier, T., Eiglmeier, K. & Cole, S. T. (1999). *Mol. Microbiol.* **32**, 643–655.
Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Goulding, C. W. *et al.* (2002). *Curr. Drug. Targets Infect. Disord.* **2**, 121–141.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Ni, J., Sakanyan, V., Charlier, D., Glansdorff, N. & Van Duyne, G. D. (1999). *Nature Struct. Biol.* **6**, 427–432.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Smith, C. V. & Sacchettini, J. C. (2003). *Curr. Opin. Struct. Biol.* **13**, 658–664.
Sunnerhagen, M., Nilges, M., Otting, G. & Carey, J. (1997). *Nature Struct. Biol.* **4**, 819–826.
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.
Van Duyne, G. D., Ghosh, G., Maas, W. K. & Sigler, P. B. (1996). *J. Mol. Biol.* **256**, 377–391.
World Health Organization (2003). *The World Health Organization Global Tuberculosis Program.* http://www.who.int/gtb/.