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

Crystallization and preliminary X-ray analysis of a novel esterase Rv0045c from *Mycobacterium tuberculosis*

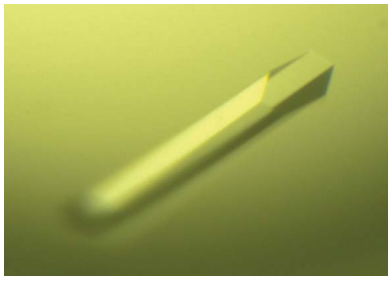
Lipeng Xu,[a,b]‡ Jiubiao Guo,[a]‡ Xiangdong Zheng,[a] Tingyi Wen,[c] Fei Sun,[d] Siguo Liu[b]* and Hai Pang[a]*

[a]School of Medicine, Tsinghua University, Beijing 100084, People's Republic of China, [b]Harbin Veterinary Research Institute, Chinese Academy of Agriculture, Harbin 150001, People's Republic of China, [c]Department of Industrial Microbiology and Biotechnology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, People's Republic of China, and [d]Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, People's Republic of China

‡ These authors contributed equally to this work.

Correspondence e-mail: siguo_liu@yahoo.com.cn, pangh@xtal.tsinghua.edu.cn

Received 15 April 2010
Accepted 1 September 2010



The Rv0045c protein is predicted to be an esterase that is involved in lipid metabolism in *Mycobacterium tuberculosis*. The protein was overproduced in *Escherichia coli*, purified and crystallized using the hanging-drop vapour-diffusion method. The Rv0045c protein crystals diffracted to a resolution of 2.7 Å using a synchrotron-radiation source and belonged to space group $P3_1$ or $P3_2$, with unit-cell parameters $a = b = 73.465$, $c = 48.064$ Å, $\alpha = \beta = 90$, $\gamma = 120^\circ$. Purified SeMet-labelled Rv0045c protein was also crystallized and formed crystals that diffracted to a resolution of 3.0 Å using an in-house X-ray radiation source.

1. Introduction

In 1998, the whole genome of *Mycobacterium tuberculosis* strain H37Rv was determined by the Sanger Center and the Pasteur Institute, leading to the conjecture that at least 250 enzymes are involved in lipid metabolism, compared with about 50 enzymes in *Escherichia coli* (Canaan *et al.*, 2004; Cole *et al.*, 1998; Camus *et al.*, 2002).

M. tuberculosis Rv0045c, a putative esterase, is involved in lipid metabolism. Rv0045c can be classified as a member of the α/β-hydrolase family, which is found in a very wide range of enzymes with distinct phylogenetic origins and catalytic functions (Schrag & Cygler, 1997). The core of the α/β-hydrolase fold consists of a β-sheet core of five to eight strands connected by α-helices to form an $\alpha/\beta/\alpha$ sandwich. In most family members the β-strands are parallel, but some show an inversion in the order of the first strand, resulting in an antiparallel orientation. All of the members of the α/β-hydrolase family have a nucleophile–histidine–acid catalytic triad that has evolved to efficiently operate on substrates of different chemical compositions or physicochemical properties. The catalytic triad is usually composed of a nucleophilic serine in a G-X_1-S-X_2-G pentapeptide motif (where X is any residue) and an acidic residue (aspartate or glutamate) that is hydrogen bonded to a histidine residue. Additionally, in various biological contexts elements of the catalytic triad are borne on loops which are the best-conserved structural features in the fold (Holmquist, 2000; Hotelier *et al.*, 2004). Studies of acetylcholinesterase (AChE), carboxypeptidase II (CPW), dienelactone hydrolase (DLH), haloalkane dehalogenase (HAL) and lipase (GLP) have shown that the strand–nucleophile–helix feature is the most conserved structure in the α/β-hydrolase-fold family, with the nucleophile and acid loops accommodating more than one type of amino acid (Ollis *et al.*, 1992).

The sequence analysis of Rv0045c shows that the protein possesses a nucleophilic G-X_1-S-X_2-G sequence. Moreover, the Rv0045c protein shares 23% identity with YbfF, a new esterase from *E. coli* (PDB codes 3bf7 and 3bf8; Park *et al.*, 2007, 2008), and 21% identity with the complex between the Pp2a-specific methylesterase Pme-1 and the Pp2a core enzyme (PDB code 3c5w; Xing *et al.*, 2008). Recently, the high-resolution structure of the esterase YbfF has been solved. The main catalytic sites of YbfF are Ser89, Asp113, Ser206 and His234. All of these amino-acid residues are conserved in Rv0045c. However, the residues around the active site in YbfF are quite divergent from those in Rv0045c, indicating that Rv0045c has

distinct substrate specificity and catalytic conditions. At present, there is no detailed quaternary structure of the Rv0045c protein. In order to illustrate the molecular mechanism of Rv0045c in hydrolyzing substrates, structural study of the protein is necessary and is in progress.

In this paper, the expression, purification, crystallization and preliminary X-ray diffraction analysis of Rv0045c are described. Since molecular replacement failed to solve the phase problem, an SeMet-labelled Rv0045c derivative was expressed and crystallized with the aim of performing a MAD experiment. The SeMet-labelled protein derivative crystallized in the same form as the native form and diffracted to a resolution of 3.0 Å. The crystals reported here will form the basis for the determination of the structure of Rv0045c and the identification of the active site, thereby providing information regarding its catalytic mechanism.

2. Materials and methods

2.1. Protein expression and purification

The *Rv0045c* gene (Gene ID 887029) was subcloned into pET-28a vector (Novagen) *via* *Bam*HI/*Xho*I restriction sites. The fragment was amplified using the standard PCR procedure with the primers P1 (5′-CGC**GGATCC**CTATCTGACGACGAACTGACC-3′; *Bam*HI site in bold) and P2 (5′-TCCG**CTCGAG**TCAGCGTGTGTCGAGCACCCC-3′; *Xho*I site in bold). *E. coli* strain BL21 (DE3) (Novagen) was used for the expression of recombinant Rv0045c protein with a 6×His tag at the N-terminus. Cells carrying the *Rv0045c* gene were grown in LB medium containing 50 µg ml^{-1} kanamycin at 310 K until the OD_{600} reached 0.6–0.8 and were induced with 0.3 m*M* IPTG at 289 K for 20 h.

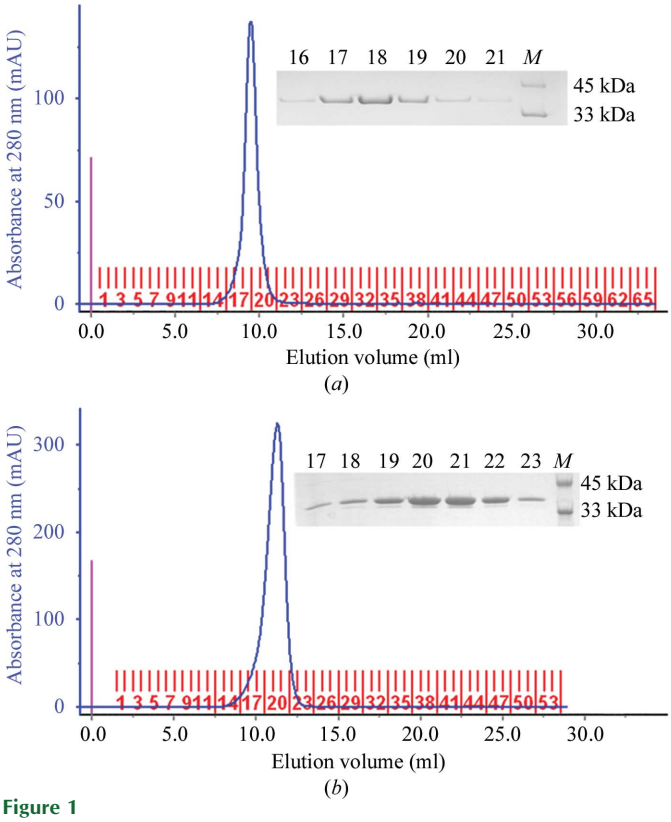


Figure 1
Gel-filtration chromatography of (*a*) native Rv0045c and (*b*) SeMet-labelled Rv0045c. The purity of the target fractions was checked by SDS–PAGE (insets).

The harvested cells were resuspended in buffer *A* (20 m*M* Tris–HCl, 150 m*M* NaCl, 10 m*M* imidazole pH 7.5) and disrupted by sonication. After centrifugation, the supernatant was loaded onto Ni Sepharose 6 Fast Flow resin (GE Healthcare) pre-equilibrated with buffer *A*. The resin was washed with buffer *B* (20 m*M* Tris–HCl, 150 m*M* NaCl, 20 m*M* imidazole pH 7.5) and the protein was eluted with buffer *C* (20 m*M* Tris–HCl, 150 m*M* NaCl, 200 m*M* imidazole pH 7.5). The target protein was dialyzed against buffer *D* (20 m*M* Tris pH 7.5) at 277 K to remove imidazole and salt, and was then concentrated using a 10 kDa Centricon concentrator (Millipore). The concentrated protein was applied onto a 1 ml Resource Q column (GE Healthcare) which was equilibrated with buffer *D* and eluted with an NaCl gradient (0–2 *M*). Finally, the target fractions were refined by loading them onto a Superdex 75 10/300 GL column (GE Healthcare) in buffer *D* and the eluted fractions were checked by SDS–PAGE (Fig. 1*a*). Selenomethionine-labelled Rv0045c was overproduced in *E. coli* strain B834 (DE3) (Novagen). It was prepared by the same procedure as described above except for the use of LeMaster medium (LeMaster & Richards, 1985) instead of LB medium. The purity of the SeMet-labelled Rv0045c eluted from the gel-filtration chromatography was checked by SDS–PAGE (Fig. 1*b*). The native and SeMet-labelled Rv0045c proteins were both concentrated to 5 mg ml^{-1} prior to crystallization.

2.2. Crystallization

Initial crystallization trials were performed in 16-well Linbro plates (Tianyu, Tianjin, People's Republic of China) using the hanging-drop vapour-diffusion method. The Crystal Screen and Crystal Screen 2 kits from Hampton Research (Aliso Viejo, California, USA) were used to find initial crystallization conditions. Purified Rv0045c protein was crystallized from drops that were prepared by mixing 1 µl protein solution and 1 µl precipitant solution. Small needle-shaped crystals of Rv0045c protein (Fig. 2) were grown from Crystal Screen condition No. 6 consisting of 0.2 *M* $MgCl_2$, 100 m*M* Tris–HCl pH 8.5, 30%(*w*/*v*) polyethylene glycol (PEG) 4000. Optimization of the crystallization conditions was performed by altering the pH and the salt, PEG and protein concentrations. The final optimized crystals of Rv0045c protein (Fig. 3) used for data collection were obtained within one week from buffer consisting of 0.2 *M* $MgCl_2$, 100 m*M* imidazole pH 7.0, 19%(*w*/*v*) PEG 4000 at 289 K. Crystallization drops consisting of

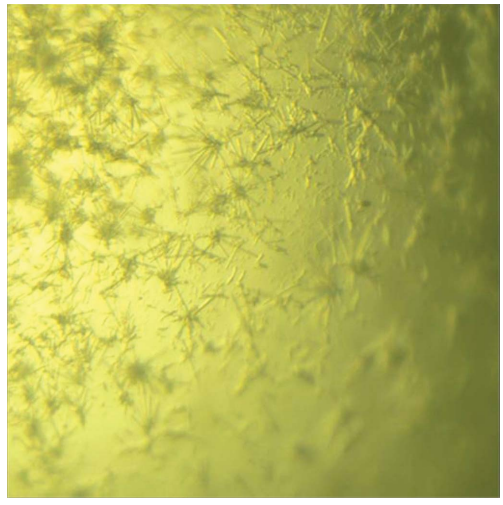


Figure 2
Small needle-shaped crystals of native Rv0045c protein obtained by initial screening.

1.5 µl protein solution and 1.5 µl precipitant solution were equilibrated against 200 µl reservoir solution. Crystallization of SeMet-labelled Rv0045c was performed under similar conditions and the crystals (Fig. 4) used for data collection were obtained using the same conditions.

2.3. Data collection

Prior to data collection, crystals were washed in a solution consisting of 0.2 *M* $MgCl_2$, 100 m*M* imidazole pH 7.0, 19%(*w*/*v*) PEG 4000, 20%(*v*/*v*) glycerol and were then transferred to a cryostream and cooled to 100 K in a nitrogen-gas stream. A complete data set for native Rv0045c crystals was collected on synchrotron beamline BL5A at KEK Photon Factory in Japan. The crystal-to-detector distance was 250 mm and the oscillation range was set to 2.0°.

3. Results and discussion

The *Rv0045c* gene was successfully amplified by PCR and the protein produced contained a 6×His tag at the N-terminus. The molecular weight of the recombined Rv0045c protein purified from *E. coli* was about 35.5 kDa. The initial crystals of native Rv0045c (Fig. 2) did not provide good-quality diffraction data. After optimization of the crystallization conditions, larger crystals of native protein (Fig. 3) that diffracted to a resolution of 2.7 Å were grown. Various cryoprotectants were tested, but only the use of 20%(*v*/*v*) glycerol improved the diffraction pattern. It is important to note that dehydration of the Rv0045c crystals was not useful in improving the diffraction quality. Diffraction data for the crystals were obtained in the resolution range 50–2.70 Å (data not shown) and were processed using the *HKL*-2000 program package (Otwinowski & Minor, 1997). Data-collection and processing statistics are shown in Table 1. Fig. 5 shows the diffraction pattern of native Rv0045c crystals obtained using an in-house (Institute of Biophysics, Chinese Academy of Sciences) rotating-anode FR-E X-ray generator and an R-AXIS IV^{++} image-plate detector (Rigaku, Japan). The Matthews coefficient was 2.1 Å^3 Da^{-1} (Matthews, 1997). Autoindexing in *HKL*-2000 suggested that the Laue group was *P*3, with unit-cell parameters $a = b = 73.465$, $c = 48.064$ Å, $\alpha = \beta = 90$, $\gamma = 120°$. As the Matthews coefficient for a single molecule in the asymmetric unit of a *P*3 cell was 2.1 Å^3 Da^{-1}, giving a solvent content of 42%, and there was an $l = 3n$ condition on the 00*l* reflections, the space group of the crystals was either $P3_1$ or $P3_2$. 250 mm was chosen as the crystal-to-detector distance because it

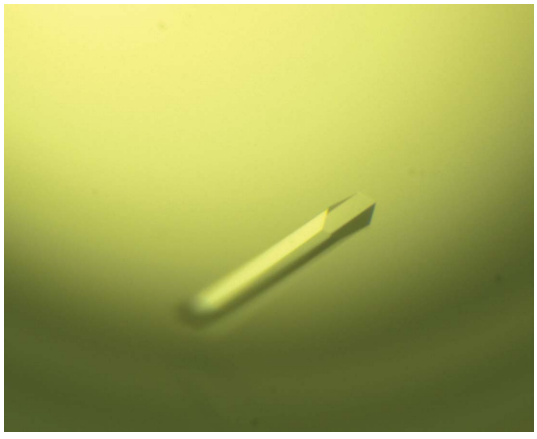


Figure 3
Single crystal of native Rv0045c protein obtained after optimization of the crystallization conditions.

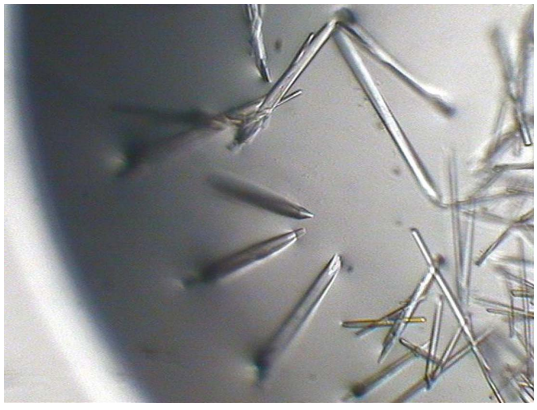


Figure 4
Crystals of SeMet-labelled Rv0045c.

Table 1
Statistical analysis of X-ray diffraction data for native Rv0045c crystals.

Values in parentheses are for the highest resolution shell.

Space group	$P3_1$ or $P3_2$
Unit-cell parameters (Å, °)	$a = b = 73.465$, $c = 48.064$, $\alpha = \beta = 90$, $\gamma = 120$
No. of protomers per asymmetric unit	1
Temperature (K)	100
Wavelength (Å)	1.0719
Oscillation range (°)	2.0
Crystal-to-detector distance (mm)	250
No. of frames	120
Resolution range (Å)	50–2.70 (2.75–2.70)
Completeness (%)	100.0 (99.7)
R_{merge}† (%)	8.2 (41.6)
Mean $I/\sigma(I)$	24.1 (2.3)
Redundancy	6.5 (4.5)

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean intensity of symmetry-equivalent reflections.

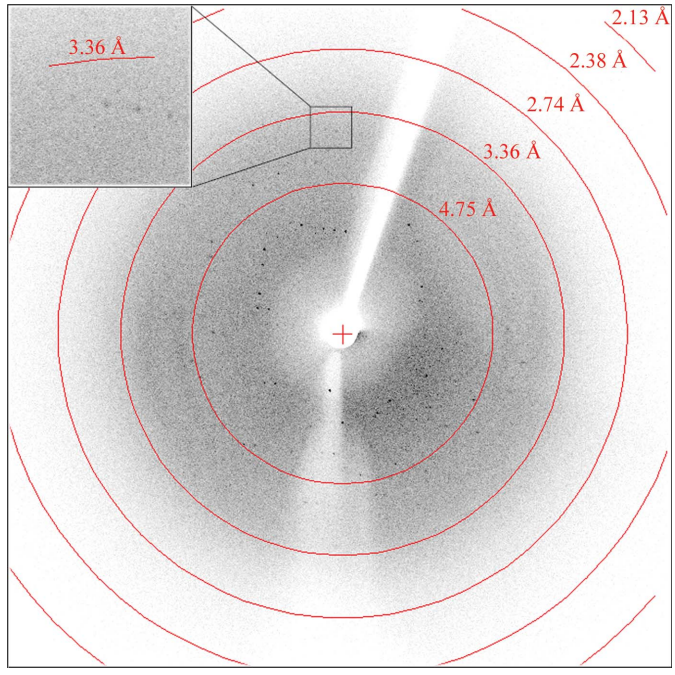


Figure 5
Diffraction pattern of native Rv0045c crystals obtained using the in-house source. An enlarged view of the marked region is shown. The red circles indicate resolutions of 4.75, 3.36, 2.74, 2.38 and 2.13 Å.

was long enough for a fine data set to be collected from the crystals; because of a shortage of beamtime the oscillation range was set to 2.0°.

We first attempted to solve the structure of the Rv0045c protein by molecular replacement (MR) using the coordinates of YbfF, a new esterase from *E. coli* (PDB codes 3bf7 and 3bf8; Park *et al.*, 2008), and of the complex between the Pp2a-specific methylesterase Pme-1 and the Pp2a core enzyme (PDB code 3c5w; Xing *et al.*, 2008) as search models. However, no significant peaks were found in the rotation and translation functions despite exhaustive attempts and the calculated phases from the MR solutions did not provide any interpretable electron density. Therefore, SeMet-labelled Rv0045c was prepared and crystallized for MAD structure determination. SeMet-labelled Rv0045c crystals grew under the same conditions as used for the native crystals and diffracted to a resolution of 3.0 Å using a Rigaku FR-E rotating-anode X-ray generator equipped with a Rigaku R-AXIS IV++ imaging-plate area detector. Higher resolution diffraction data collection from the SeMet-labelled Rv0045c crystals is planned.

We thank Q. J. Zhou and K. Zhang of the Institute of Biophysics, Chinese Academy of Sciences for their assistance during X-ray diffraction data collection. We acknowledge the three anonymous referees for the enhancement of the manuscript during the review process. This work was supported by the Foundation of the State Key Laboratory of Veterinary Biotechnology (SKLVBF2009), the Harbin Veterinary Research Institute (HVRI), the National Drug Discovery Program (grant 2008ZX09401-05) and the 973 Program of the Ministry of Science and Technology of China (grant 2006CB806506).

References

Camus, J.-C., Pryor, M. J., Médigue, C. & Cole, S. T. (2002). *Microbiology*, **148**, 2967–2973.
Canaan, S., Maurin, D., Chahinian, H., Pouilly, B., Durousseau, C., Frassinetti, F., Scappuccini-Calvo, L., Cambillau, C. & Bourne, Y. (2004). *Eur. J. Biochem.* **271**, 3953–3961.
Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.
Holmquist, M. (2000). *Curr. Protein Pept. Sci.* **1**, 209–235.
Hotelier, T., Renault, L., Cousin, X., Negre, V., Marchot, P. & Chatonnet, A. (2004). *Nucleic Acids Res.* **32**, D145–D147.
LeMaster, D. M. & Richards, F. M. (1985). *Biochemistry*, **24**, 7263–7268.
Matthews, B. W. (1997). *The Proteins*, Vol. 3, edited by H. Neurath & R. L. Hill, pp. 468–477. New York: Academic Press.
Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., Sussman, J. L., Verschueren, K. H. G. & Goldman, A. (1992). *Protein Eng.* **5**, 197–211.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Park, S.-Y., Kim, J.-T., Kang, S. G., Woo, J.-H., Lee, J.-H., Choi, H.-T. & Kim, S.-J. (2007). *Appl. Microbiol. Biotechnol.* **77**, 107–115.
Park, S.-Y., Lee, S.-H., Lee, J., Nishi, K., Kim, Y.-S., Jung, C.-H. & Kim, J.-S. (2008). *J. Mol. Biol.* **376**, 1426–1437.
Schrag, J. D. & Cygler, M. (1997). *Methods Enzymol.* **284**, 85–107.
Xing, Y., Li, Z., Chen, Y., Stock, J. B., Jeffrey, P. D. & Shi, Y. (2008). *Cell*, **133**, 154–163.