Purification, crystallization and preliminary X-ray crystallographic studies of Rv3899c from *Mycobacterium tuberculosis*Yingjia Song,^{a‡} Jianghui Liu,^{b‡}
De-Feng Li,^c Honglin Li,^a Shihua
Wang,^b Da-Cheng Wang,^c
Jie Zhou^d and Lijun Bi^{a,c,*}^aShanghai Key Laboratory of New Drug Design,
School of Pharmacy, East China University of
Science and Technology, 130 Meilong Road,
Shanghai 200237, People's Republic of China,^bThe Ministry of Education Key Laboratory of
Biopesticide and Chemical Biology and the
School of Life Sciences, Fujian Agriculture and
Forestry University, Fuzhou 350002, People's
Republic of China, ^cNational Laboratory of
Biomacromolecules and Laboratory of
Noncoding RNA, Institute of Biophysics,
Chinese Academy of Sciences, Beijing 100101,
People's Republic of China, and ^dThe Fourth
People's Hospital Foshan, Foshan 528000,
People's Republic of China‡ These authors contributed equally to this
work.

Correspondence e-mail: blj@sun5.ibp.ac.cn

Received 19 September 2014

Accepted 11 December 2014

Rv3899c, a hypothetical protein from *Mycobacterium tuberculosis* that is conserved within the mycobacteria, is predicted to be secreted and has been found in culture filtrates. Here, Rv3899c has been cloned, expressed in *Escherichia coli* and purified using standard chromatographic techniques. The hanging-drop vapour-diffusion method with PEG 3350 as a precipitant was used to crystallize the protein. N-terminal sequencing results showed that the amino-acid sequence of the crystallized protein began with GATAG, indicating that it is a fragment containing residues 184–410 of Rv3899c. Rv3899c^{184–410} crystals exhibited the symmetry of space group $P2_12_12_1$, with unit-cell parameters $a = 49.88$, $b = 54.72$, $c = 75.52$ Å, $\alpha = \beta = \gamma = 90^\circ$, and diffracted to a resolution of 1.90 Å.

1. Introduction

Tuberculosis (TB) is a chronic infectious disease caused by infection with *Mycobacterium tuberculosis*. It is estimated that one-third of the world's population is infected with *M. tuberculosis*, and two million die from the disease every year (World Health Organization, 2013). These data highlight the importance of effective TB prevention. The Bacillus Calmette–Guérin (BCG) vaccine, the only licensed TB vaccine currently available, provides protection for young children but is less effective in adults (Colditz *et al.*, 1995). Consequently, the development of new vaccines that are more effective than the BCG vaccine is both extremely urgent and necessary. In the past decade, impressive strides have been made in the development of new TB vaccines, and clinical trials of several TB vaccine candidates are currently under way (Gröschel *et al.*, 2014).

The function of Rv3899c, a *M. tuberculosis* protein comprised of 410 amino acids with a predicted molecular weight of 40.8 kDa, is unknown; however, Rv3899c has been identified as a potential immunogen in a recent study (Deng *et al.*, 2014). Rv3899c has previously been identified as a culture-filtrate protein using two-dimensional PAGE combined with liquid chromatography-coupled MS/MS (Målen *et al.*, 2007). Rv3899c only has clear homologues in the genus *Mycobacterium*, and annotations of its homologues indicate that this protein has a twin-arginine transport (TAT) signal, possibly explaining its presence in the culture filtrate. The isoelectric point of Rv3899c is 5.67 as predicted by ExPASy (Bjellqvist *et al.*, 1993), and no transmembrane areas were identified by TMHMM (Sonnhammer *et al.*, 1998). Analysis of the secondary structure of Rv3899c using Phyre (Kelley & Sternberg, 2009) suggests that 76% of the sequence is disordered, 24% is α -helices and 1% is β -strands. In this paper, we report the cloning, expression, purification, crystallization and preliminary X-ray studies of Rv3899c.

2. Materials and methods

2.1. Macromolecule production

The gene encoding Rv3899c (UniProt accession No. O05446) was amplified by the polymerase chain reaction (PCR) using *M. tuberculosis* H37Rv genomic DNA as a template (Brosch *et al.*, 1998). The presence of 10% dimethyl sulfoxide (DMSO), a co-solvent that improves the denaturation of GC-rich DNA, proved to be essential

Table 1

Macromolecule-production information.

Source organism	<i>M. tuberculosis</i> H37Rv
DNA source	<i>M. tuberculosis</i> genome
Forward primer†	AAAACATATGCTGACGGGGCAACCGG
Reverse primer‡	AAAAGAATTCTCAGGCAGCGAACCTCGT
Cloning vector	pET-28a
Expression vector	pET-28a:Rv3899c
Expression host	<i>E. coli</i> BL21(DE3)
Complete amino-acid sequence of the construct produced§	MGSSHHHHHSSGLVPRGSHMMVTGQPAAGAHSLSEGAMTA-MQSGSVPPQATPPITTPPVVSAPTMAAGIEATHGPVDTP-ANTSGAPPASTGTTGPVAPTPTVAGPVAAPVVGGSAPVAGPLPAYGSDLRPPVVAAPVPSVPTAPVSGAPVAPSAS-SAPSAGGALVSPVERAASKAVAGQAGSSTMAGASALSA-TAGATAGAVSARAAEQRLQRIQVDAVARQEPRISSWAAGLR-DDGTTTLTLDLAGGWIPPHVRLPANVTLEPTARRRDAD-VIDLLGAVVAVAHESNTYVAEPGDPALTGDRSARSAT-PKVDFEGPTLVEAVRRRDSLPRIAQAIALPAVRKTVGLENEAELLHGCTAVKESVLKAYPSHETAVGDWMLLAATEAL-IDEQDYLANVHLAWYAVTTRRGSGRFAA

† The NdeI site is underlined. ‡ The EcoRI site is underlined. § The underlined amino-acid sequence is the crystallized protein fragment.

for PCR amplification. This requirement for DMSO is consistent with the 65.6% GC content of *M. tuberculosis* genomic DNA (Cole *et al.*, 1998). The synthetic oligonucleotide primers used are listed in Table 1 and contained NdeI and EcoRI restriction sites. 30 cycles of PCR comprising denaturation at 367 K (30 s), annealing at 332 K (30 s) and extension at 341 K (80 s) were performed. The PCR product was cloned into a pET-28a expression vector (Novagen) between the NdeI and EcoRI restriction sites, and the recombinant plasmid was sequenced to confirm the identity of the cloned DNA fragment and to ensure that no mutations had been introduced by the PCR amplification step. The recombinant plasmid was subsequently transformed into *Escherichia coli* BL21(DE3) competent cells for expression.

E. coli BL21(DE3) recombinant cells were grown in LB medium containing 30 µg ml⁻¹ kanamycin at 310 K. When the OD₆₀₀ reached 0.6–0.8, IPTG was added to the culture medium to a final concentration of 0.4 mM to induce protein expression. After 16 h, the cells were harvested by centrifugation at 4000g for 15 min at 277 K. The cells were resuspended in lysis buffer (20 mM Tris–HCl pH 8.0, 500 mM NaCl, 10 mM imidazole, 2 mM PMSF, 10% glycerol) and

Figure 1

SDS–PAGE of purified Rv3899c. Lane 1, molecular-weight markers (labelled in kDa); lane 2, Rv3899c after gel-filtration column chromatography; lane 3, dissolved Rv3899c.

Table 2

Crystallization.

Method	Hanging-drop vapour diffusion
Plate type	16-well hanging-drop plate
Temperature (K)	289
Protein concentration (mg ml ⁻¹)	10
Buffer composition of protein solution	20 mM Tris–HCl pH 8.0, 100 mM NaCl, 5% glycerol
Composition of reservoir solution	200 mM ammonium acetate, 100 mM bis-tris pH 5.5, 25% PEG 3350
Volume and ratio of drop	2 µl (1:1)
Volume of reservoir (µl)	200

Table 3

Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	Rigaku MicroMax-007 HF, IBP
Wavelength (Å)	1.5418
Temperature (K)	100
Detector	R-Axis IV ⁺⁺
Crystal-to-detector distance (mm)	150
Rotation range per image (°)	1
Total rotation range (°)	180
Exposure time per image (s)	240
Space group	P2 ₁ 2 ₁ 2 ₁
Unit-cell parameters (Å, °)	<i>a</i> = 49.88, <i>b</i> = 54.72, <i>c</i> = 75.52, $\alpha = \beta = \gamma = 90$
Mosaicity (°)	0.87
Resolution range (Å)	44.31–1.90 (2.00–1.90)
Total No. of reflections	109835 (10690)
No. of unique reflections	16664 (2179)
Completeness† (%)	100 (93)
Multiplicity	6.6 (4.8)
$\langle I/\sigma(I) \rangle$	13.1 (3.5)
<i>R</i> _{int} ‡ (%)	10.0 (42.5)
Overall <i>B</i> factor from Wilson plot (Å ²)	17.4

† The completeness starts to drop significantly at 1.8 Å resolution. ‡ $R_{int} = \sum_{hkl} \{ [N(hkl)] / [N(hkl) - 1] \}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl*, $\sum_i$ is the sum over the individual measurements of a reflection and $\sum_{hkl}$ is the sum over all reflections.

lysed by sonication. All subsequent steps were performed on ice or at 277 K. The cell lysate was centrifuged at 12 000g for 40 min at 277 K. The supernatant was applied onto a Ni–NTA affinity column containing 7 ml Ni Sepharose (GE Healthcare) pre-equilibrated with lysis buffer for 1.5 h. Bound proteins were eluted from the column with 30 ml elution buffer consisting of 20 mM Tris–HCl pH 8.0, 500 mM NaCl, 250 mM imidazole, 10% glycerol. Fractions containing the target protein were pooled, concentrated using Amicon filters with a 10 kDa cutoff and further purified by gel-filtration column chromatography using a Superdex 200 column (GE Healthcare) pre-equilibrated with a buffer consisting of 20 mM Tris–HCl pH 8.0, 100 mM NaCl, 5% glycerol.

2.2. Crystallization

Protein samples at a concentration of 20 mg ml⁻¹ were used in crystallization studies. The protein concentration was determined by the method of Bradford (1976) using a protein-assay kit with bovine serum albumin as the standard (Bio-Rad). Initial crystallization trials were carried out in 16-well plates at 289 K using the hanging-drop vapour-diffusion method (1 + 1 µl drops, 200 µl reservoir solution) with screening kits from Hampton Research (Crystal Screen, Crystal Screen 2 and Index). Crystals appeared in several conditions after 4 d. The best condition was 0.2 M ammonium acetate, 0.1 M bis-tris pH 5.5, 25% PEG 3350 (Index condition No. 78) based on the diffraction of the crystal. Refinement of crystallization conditions was achieved by altering the concentration of the protein sample to 10 mg ml⁻¹. Crystallization information is summarized in Table 2.

Figure 2
Crystals of Rv3899c^{184–410}.

2.3. Data collection and processing

Before X-ray data collection, crystals were flash-cooled in liquid nitrogen after soaking in cryoprotectant solution consisting of 200 mM ammonium acetate, 100 mM bis-tris pH 5.5, 25% PEG 3350, 20% glycerol. Diffraction data were collected on a MicroMax-007 HF-IP X-ray diffraction data-collection system (Rigaku, Japan) at the Institute of Biophysics, Chinese Academy of Sciences (IBP). A total of 180 images were collected with an oscillation step of 1° and 4 min exposure per image. The crystal-to-detector distance was maintained at 150 mm. Diffraction data were processed and scaled using *HKL-2000* (Otwinowski & Minor, 1997). A summary of the X-ray diffraction data-collection and processing statistics is given in Table 3.

3. Results and discussion

Rv3899c was successfully overexpressed in *E. coli* BL21(DE3) cells, and a crystallization-grade protein sample which migrated with an apparent molecular weight of approximately 47 kDa on SDS-PAGE (Fig. 1) was obtained after two purification steps comprising of nickel-affinity chromatography and size-exclusion chromatography. The yield of the pure protein was ~10 mg per litre of culture. Initial crystals of Rv3899c appeared after about 4 d under several conditions. Various optimization procedures such as altering the protein concentration and the temperature of crystal growth were attempted in order to obtain diffraction-quality crystals. Crystals were finally obtained at 289 K using the hanging-drop vapour-diffusion method with a reservoir solution consisting of 0.2 M ammonium acetate, 0.1 M bis-tris pH 5.5, 25% PEG 3350 (Fig. 2). The dissolved protein showed an apparent molecular weight of approximately 25 kDa on SDS-PAGE (Fig. 1), indicating that it was likely to be a fragment of Rv3899c. N-terminal sequencing revealed that the protein fragment began with GATAG, indicating that the crystallized protein comprised residues 184–410 of Rv3899c, hereafter referred to as Rv3899c^{184–410}. Analysis of the secondary structure of Rv3899c^{184–410} using *Phyre* predicts that 27% of the sequence is disordered, 49% is α -helices and 4% is β -strands. The diffraction data were processed to 1.90 Å resolution (Fig. 3). Systematic absences clearly indicated the space group to be *P*₂₁₂₁₂₁. The unit-cell parameters of the native crystal are $a = 49.88$, $b = 54.72$, $c = 75.52$ Å, $\alpha = \beta = \gamma = 90^\circ$. According to the unit-cell parameters and the molecular weight of the Rv3899c^{184–410} crystals, solvent-content analysis suggested the

Figure 3
X-ray diffraction pattern of Rv3899c^{184–410}.

presence of one molecule per asymmetric unit, with a V_M value (Matthews, 1968) of $2.06 \text{ Å}^3 \text{ Da}^{-1}$ and a solvent content of 40.38%. Other statistics are given in Table 3. These crystals had excellent reproducibility and are being used for structure determination. Heavy-atom screening is currently under way in order to solve the three-dimensional structure by multiple isomorphous replacement. The work described here will pave the way for structure determination of Rv3899c^{184–410}.

We would like to gratefully acknowledge the help of the staff of the Protein Science Core Facility Center at the Institute of Biophysics, Chinese Academy of Sciences for crystal diffraction data collection. Financial support for this project was provided by the Chinese Ministry of Science and Technology 973 program (Grant No. 2011CB910302).

References

- Bjellqvist, B., Hughes, G. J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J., Frutiger, S. & Hochstrasser, D. (1993). *Electrophoresis*, **14**, 1023–1031.
- Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
- 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.
- Colditz, G. A., Berkey, C. S., Mosteller, F., Brewer, T. F., Wilson, M. E., Burdick, E. & Fineberg, H. V. (1995). *Pediatrics*, **96**, 29–35.
- Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.
- Deng, J. *et al.* (2014). *Cell Rep.*, doi:10.1016/j.celrep.2014.11.023.
- Gröschel, M. I., Prabowo, S. A., Cardona, P. J., Stanford, J. L. & van der Werf, T. S. (2014). *Vaccine*, **32**, 3162–3168.
- Kelley, L. A. & Sternberg, M. J. (2009). *Nature Protoc.* **4**, 363–371.
- Målen, H., Berven, F. S., Fladmark, K. E. & Wiker, H. G. (2007). *Proteomics*, **7**, 1702–1718.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Sonnhammer, E. L., von Heijne, G. & Krogh, A. (1998). *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 175–182.
- World Health Organization (2013). *Global Tuberculosis Report 2013*. Geneva: World Health Organization. http://www.who.int/tb/publications/global_report/.