Crystal structure of Rv3899c^{184–410}, a hypothetical protein from *Mycobacterium tuberculosis*

Yingying Liu,^{a‡} Yunrong Gao,^{b‡} Defeng Li,^b Joy Fleming,^b Honglin Li^a and Lijun Bi^{a,b*}

^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, and ^bNational Laboratory of Biomacromolecules and Laboratory of Noncoding RNA, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing 100101, People's Republic of China. *Correspondence e-mail: blj@sun5.ibp.ac.cn

Received 18 March 2016

Accepted 6 July 2016

Edited by A. Nakagawa, Osaka University, Japan

‡ These authors contributed equally to this work.

Keywords: *Mycobacterium tuberculosis*; Rv3899c^{184–410}; crystal structure.

PDB reference: Rv3899c^{184–410}, 5imu

Supporting information: this article has supporting information at journals.iucr.org/f

Rv3899c is a hypothetical protein from *Mycobacterium tuberculosis* which is conserved across mycobacteria. It is predicted to be secreted and has been found in culture filtrates. It has been proposed as a potential vaccine candidate; however, its biological function is unknown. Here, the global structure of Rv3899c^{184–410}, a fragment of Rv3899c, is reported. The structure resembles the shell of a sea snail, and its N- and C-termini form two relatively independent compact domains: an $\alpha/\beta/\alpha$ sandwich folding domain and an α -helix bundle domain. There are no reported protein structures for any Rv3899c homologues; this structure provides the first structural glimpse of a new protein family consisting of Rv3899c and its homologues.

1. Introduction

Tuberculosis (TB), caused by infection with the pathogen *Mycobacterium tuberculosis*, remains a threat to global public health. It is estimated that there were 9.6 million cases of tuberculosis worldwide and 1.5 million deaths from TB in 2014 (World Health Organization, 2015). Effective TB prevention and control is thus extremely important and urgent. The Bacillus Calmette–Guérin (BCG) vaccine is the only licensed TB vaccine, but it provides insufficient protection (Moliva *et al.*, 2015). Improving this vaccine, or developing more effective vaccines, is thus an important strategy for improved TB prevention and control. The secretome of *M. tuberculosis* is likely to be an ideal source of vaccine candidates (Zheng *et al.*, 2013).

Rv3899c, annotated as a conserved hypothetical protein, is composed of 410 amino acids and has a predicted molecular mass of 40.8 kDa (Lew *et al.*, 2011). Its function is unknown and it only has clear homologues in the genus *Mycobacterium*. Little else is known about this protein except that it is secreted; it has a twin-arginine transport (TAT) signal and has been identified in culture filtrates using two-dimensional PAGE combined with liquid chromatography-coupled MS/MS (Målen *et al.*, 2007). In addition, it has been found to be present in *M. tuberculosis* H37Rv-infected guinea pig lungs at 90 d, but not 30 d, post-infection (Kruh *et al.*, 2010). Deng and coworkers identified protein Rv3899c as a potential immunogen that can induce a strong immunological reaction, implying its potential as a vaccine candidate (Deng *et al.*, 2014). Of interest, the Rv3899c gene is located in the vicinity of the ESX-2 operon, suggesting a possible functional relationship between Rv3899c and the ESX-2 system, one of five type VII or ESAT-6 secretion systems (ESX-1 to ESX-5) that

are essential for mycobacterial virulence and/or viability (Houben *et al.*, 2014; Newton-Foot *et al.*, 2016).

As a first step to studying Rv3899c at the molecular level, we have purified Rv3899c and solved the structure of a fragment of this protein (amino-acid residues 184–410). Its structure resembles the shell of a sea snail in shape and consists of an $\alpha/\beta/\alpha$ sandwich folding domain and a α -helix bundle domain.

2. Materials and methods

2.1. Macromolecule production

The expression, purification and crystallization of Rv3899c have been reported previously (Song *et al.*, 2015). In brief, Rv3899c protein was expressed in *Escherichia coli* BL21(DE3) cells and purified first with an Ni^{2+} -chelating column (GE Healthcare) followed by a Superdex G200 size-exclusion chromatography column (GE Healthcare).

2.2. Crystallization

The protein was crystallized using a solution consisting of 0.2 M ammonium acetate, 0.1 M bis-tris pH 5.5, 25% PEG 3350. The crystallized protein was previously shown to be a fragment of Rv3899c comprising residues 184–410 (Song *et al.*, 2015).

2.3. Data collection and processing

Diffraction data for native Rv3899c^{184–410} were collected on a Rigaku R-Axis IV⁺⁺ image plate using Cu $K\alpha$ radiation at

Table 1

Data collection and processing.

	Rv3899c ^{184–410}	Hg ²⁺ derivative
Diffraction source	Rigaku MicroMax-007 HF	Rigaku MicroMax-007 HF
Wavelength (Å)	1.5418	1.5418
Temperature (K)	100	100
Detector	R-Axis IV ⁺⁺	Saturn CCD
Crystal-to-detector distance (mm)	150	60
Rotation range per image (°)	1	1
Total rotation range (°)	180	360
Exposure time per image (s)	240	180
Space group	$P2_12_12_1$	$P2_12_12_1$
<i>a</i> , <i>b</i> , <i>c</i> (Å)	50.15, 54.97, 75.80	47.54, 54.01, 74.41
Mosaicity (°)	0.87	0.92
Resolution range (Å)	44.51–1.90 (2.00–1.90)	45.54–2.08 (2.14–2.08)
Total No. of reflections	109835 (10690)	132756 (2190)
No. of unique reflections	16928 (2179)	19928 (757)
Completeness (%)	100 (93.0)	90.0 (46.3)
Multiplicity	6.6 (4.8)	6.6 (2.9)
$\langle I/\sigma(I) \rangle$	13.1 (3.5)	15.8 (2.0)
$R_{\text{r.i.m.}}$ (%)	10.0 (42.5)	10.8 (81.7)
Overall <i>B</i> factor from Wilson plot (Å ²)	17.4	12.2

100 K, processed with *iMosflm* (Battye *et al.*, 2011) and scaled with *SCALA* (Evans, 2006) in the CCP4 program suite (Winn *et al.*, 2011), as reported previously (Song *et al.*, 2015). An Hg²⁺-derivative crystal was obtained by soaking the native crystal with 1 mM HgCl₂ for 24 h. Data for the Hg²⁺-derivative crystal were collected on a Rigaku Saturn 944 HG CCD using Cu $K\alpha$ radiation at 100 K with an oscillation step of 1°, 180 s

Figure 1

Overall structure of Rv3899c^{184–410}. Cartoon representation (*a*), topological diagram of the secondary structure (*b*) and the electrostatic potential surface (*c*) of Rv3899c^{184–410}. β -Strands and α -helices are coloured green and cyan, respectively. The proposed ligand-binding pocket is highlighted by a black ellipse.

exposure time per image and a crystal-to-detector distance of 60 mm. The data were then processed with *XDS* (Kabsch, 2010) and scaled with *XSCALE* (Kabsch, 2010). Diffraction data-collection and processing statistics are summarized in Table 1.

2.4. Structure solution and refinement

The structure was solved by the single isomorphous replacement with anomalous scattering (SIRAS) method with *AutoSol* (Adams *et al.*, 2010) in the *PHENIX* program suite

(Adams *et al.*, 2010) using both native and Hg^{2+} -derivative data. Most of the model was built using *AutoBuild* in the *PHENIX* program suite and it was completed by manual modelling using *Coot* (Emsley *et al.*, 2010). The model was refined with *PHENIX* (Adams *et al.*, 2010), with the value of the parameter 'X-ray data.remove outliers' being TRUE (the default value) to exclude reflections with abnormal values that tend to reduce the performance of the refinement engine. Data-collection and model-refinement statistics are summarized in Table 2. The coordinate and structure-factor files have been deposited in the PDB under accession number 5imu.

Figure 2

Sequence alignment and secondary-structure distribution of Rv3899c^{184–410}. Homologous proteins to Rv3899c from different species of *Mycobacterium* were identified by *BLAST* and seven were chosen for sequence-alignment analysis: *M. avium* (ABK67464), *M. rhodesiae* JS60 (EHB55683), *M. kansasii* 824 (ETZ97236), *M. kyorinense* (WP_045378944), *M. marinum* (WP_020731216), *M. arupense* (WP_046190023) and *M. chimaera* (WP_054585524). The multiple sequence alignment was performed with *ClustalW* (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) and was displayed with secondary structures using *ESPrpt* (Robert & Gouet, 2014).

Table 2
Structure refinement.

Resolution range (Å)	44.51–1.90
Completeness (%)	99.1
σ Cutoff	0
No. of reflections, working set	16078
No. of reflections, test set	850
Final R_{cryst}	0.160
Final R_{free}	0.190
No. of non-H atoms	
Protein	1646
Ion	1
Water	348
Total	1995
R.m.s. deviations	
Bonds (Å)	0.003
Angles (°)	0.616
Average B factors (Å ²)	
Protein	17.4
Ion	38.3
Water	28.2
Total	19.3
Ramachandran plot	
Favoured regions (%)	98.60
Additionally allowed (%)	1.40

Structure figures were prepared using *PyMOL* v.1.8 (<http://www.pymol.com>).

3. Results and discussion

The Rv3899c protein is comprised of 410 amino acids and has a predicted molecular mass of 40.8 kDa. However, the stable crystallized protein studied here only comprised residues 184–410 and is referred to as Rv3899c^{184–410} (Song *et al.*, 2015). The crystal structure of Rv3899c^{184–410} was determined here to 1.9 Å resolution, with R_{cryst} and R_{free} values of 0.160 and 0.190, respectively. The electron-density map was of good quality and residues 187–403, one potassium ion and 348 water molecules were built in the final model.

The structure resembles the shell of a sea snail (Fig. 1). It consists of eight α -helices and three β -strands arranged in the order $\alpha 1$ – $\beta 1$ – $\beta 2$ – $\alpha 2$ – $\beta 3$ – $\alpha 3$ – $\alpha 4$ – $\alpha 5$ – $\alpha 6$ – $\alpha 7$ – $\alpha 8$ in the primary structure (Fig. 2). It can be divided into two relatively independent domains (Fig. 1a): (i) an $\alpha/\beta/\alpha$ sandwich folding domain located in the bottom half of the shell which consists of an antiparallel β -sheet formed by β -strands $\beta 1$, $\beta 2$ and $\beta 3$, together with the peripheral helices $\alpha 1$, $\alpha 2$ and $\alpha 3$, and (ii) a compact helix bundle located in the upper half of the shell formed from helices $\alpha 4$ – $\alpha 8$. In addition, a pocket, surrounded by helices $\alpha 3$, $\alpha 7$ and $\alpha 8$ and the loop connecting strands $\beta 2$ and $\alpha 2$, was found at the interface of the two domains. The pocket has a surface area of 342 Å² and a volume of 359 Å³ as calculated using *CASP* (Dundas *et al.*, 2006). Negatively and positively charged regions are distributed at the top and bottom of this pocket, respectively. We propose that this pocket is a potential ligand-binding site and, based on the electrostatic potential surface surrounding the pocket, the ligand may be a protein with a complementary electrostatic potential surface and a ‘finger’ that inserts into the pocket.

A search for homologous sequences in the Protein Data Bank using *BLASTp* (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>)

did not reveal any proteins with structures similar to Rv3899c^{184–410} or to its N-terminal $\alpha/\beta/\alpha$ sandwich lobe or C-terminal helix bundle. A subsequent search of the database of nonredundant protein sequences (nr) showed that Rv3899c (including its N-terminal $\alpha/\beta/\alpha$ sandwich domain and C-terminal helix bundle) only has clear homologues in the genus *Mycobacterium*. Sequence alignment revealed that residues in the $\alpha/\beta/\alpha$ sandwich fold are more conserved than those in the helix bundle, implying that the sandwich domain may play a more important functional role than the helix-bundle domain (Fig. 2).

In conclusion, we have solved the structure of the Rv3899c^{184–410} protein, a secreted protein that has previously been identified as an antigen in culture filtrates. The structure reported here gives a first structural glimpse of a new protein family consisting of Rv3899c and its homologues. Functional studies are required to investigate its biochemical characteristics, its roles in host and bacterial cells, and its potential use as a vaccine candidate.

Acknowledgements

This work was supported by the Key Program of the Chinese Academy of Sciences (KJZD-EW-L02 to LB and LF), the National Natural Science Foundation of China (31270792 to LF and 31170132 to LB) and the Key Project Specialized for Infectious Diseases of the Chinese Ministry of Health (2013ZX10003006 to LB).

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Deng, J. *et al.* (2014). *Cell Rep.* **9**, 2317–2329.
- Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. & Liang, J. (2006). *Nucleic Acids Res.* **34**, W116–W118.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Houben, E. N., Korotkov, K. V. & Bitter, W. (2014). *Biochim. Biophys. Acta*, **1843**, 1707–1716.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kruh, N. A., Troudt, J., Izzo, A., Prenni, J. & Dobos, K. M. (2010). *PLoS One*, **5**, e13938.
- Lew, J. M., Kapopoulou, A., Jones, L. M. & Cole, S. T. (2011). *Tuberculosis*, **91**, 1–7.
- Målen, H., Berven, F. S., Fladmark, K. E. & Wiker, H. G. (2007). *Proteomics*, **7**, 1702–1718.
- Molina, J. I., Turner, J. & Torrelles, J. B. (2015). *Vaccine*, **33**, 5035–5041.
- Newton-Foot, M., Warren, R. M., Sampson, S. L., van Helden, P. D. & Gey van Pittius, N. C. (2016). *BMC Evol. Biol.* **16**, 62.
- Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.
- Song, Y., Liu, J., Li, D.-F., Li, H., Wang, S., Wang, D.-C., Zhou, J. & Bi, L. (2015). *Acta Cryst.* **F71**, 107–109.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- World Health Organization (2015). *Global Tuberculosis Report 2015*. Geneva: World Health Organization. http://www.who.int/tb/publications/global_report/.
- Zheng, J., Ren, X., Wei, C., Yang, J., Hu, Y., Liu, L., Xu, X., Wang, J. & Jin, Q. (2013). *Mol. Cell. Proteomics*, **12**, 2081–2095.