Purification and X-ray crystallographic analysis of 7-keto-8-aminopelargonic acid (KAPA) synthase from *Mycobacterium smegmatis*

Shanghua Fan,^a Defeng Li,^b
Joy Fleming,^b Yuan Hong,^b
Tao Chen,^c Lin Zhou,^c Lijun Bi,^b
Dacheng Wang,^b Xianen
Zhang^{b*} and Guanjun Chen^{a*}

^aState Key Laboratory of Microbial Technology, School of Life Science, Shandong University, 27 Shanda Nan Road, Jinan, Shandong 250100, People's Republic of China, ^bNational Laboratory of Biomacromolecules and Laboratory of Noncoding RNA, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, People's Republic of China, and ^cCenter for Tuberculosis Control of Guangdong Province, Guangzhou, People's Republic of China

Correspondence e-mail:
zhangxe@sun5.ibp.ac.cn, guanjun@sdu.edu.cn

Received 17 July 2014
Accepted 11 August 2014

7-Keto-8-aminopelargonic acid synthase (KAPA synthase; BioF) is an essential enzyme for mycobacterial growth that catalyses the first committed step in the biotin-synthesis pathway. It is a pyridoxal 5'-phosphate (PLP)-dependent enzyme and is a potential drug target. Here, the cloning, expression, purification and crystallization of KAPA synthase from *Mycobacterium smegmatis* (MsBioF) and the characterization of MsBioF crystals using X-ray diffraction are described. The crystals diffracted to 2.3 Å resolution and belonged to the monoclinic space group $P2_1$, with unit-cell parameters $a = 70.88$, $b = 91.68$, $c = 109.84$ Å, $\beta = 97.8^\circ$. According to the molecular weight of MsBioF, the unit-cell parameters and the self-rotation function map, four molecules are present in each asymmetric unit with a V_M value of $2.06 \text{ Å}^3 \text{ Da}^{-1}$ and a solvent content of 40.20%.

1. Introduction

Mycobacterium tuberculosis (*Mtb*) is one of the most important human pathogens. In recent years, the emergence and spread of MDR (multidrug-resistant) and XDR (extensively drug-resistant) *Mtb* and co-infection of *Mtb* with HIV has resulted in a worldwide epidemic of tuberculosis (TB), making the treatment and control of TB problematic (Almeida Da Silva & Palomino, 2011). The development of new anti-TB drugs has thus become particularly important and urgent. Owing to the importance of biotin in cell-wall synthesis in mycobacteria (Oh *et al.*, 2006; Kurth *et al.*, 2009) and the absence of the biotin-synthesis pathway in mammals, which must obtain biotin through food intake and the action of gut bacteria (Said, 2009), the biotin-biosynthesis pathway is an attractive target for the design of antituberculosis inhibitors.

The biotin-synthesis pathway encompasses the synthesis of the precursor molecule pimeloyl-CoA (or pimeloyl-ACP; Lin *et al.*, 2010) and four subsequent committed steps which are conserved in all biotin-synthesizing organisms. KAPA synthase (BioF) plays a key role as the 'gatekeeper' of the pathway, catalyzing the first committed step: decarboxylative condensation of L-alanine with pimeloyl-CoA to form 7-keto-8-aminopelargonic acid (AON). Pyridoxal 5'-phosphate (PLP) is an essential cofactor in this reaction (Alexeev *et al.*, 1998; Ploux & Marquet, 1996; Ploux *et al.*, 1992). The structure and function of 7-keto-8-aminopelargonic acid synthase (EcaONS) in *Escherichia coli* has been well studied (Alexeev *et al.*, 1998; Webster *et al.*, 2000).

It has been reported that there are significant differences in the enzymes involved in the biotin-synthesis pathway between *Mtb* and other organisms (Dey *et al.*, 2010). Although it has been reported that *bioF* (Rv1569) is required for the survival of *Mtb* *in vitro* (Dey *et al.*, 2010) and for infection in mice (Sasseti & Rubin, 2003), the structure and catalytic mechanism of BioF in mycobacteria have yet to be resolved. Structural studies of mycobacterial KAPA synthase are of great biochemical interest for establishing the mechanism of the enzyme as well as for the possible development of new antibiotics against mycobacteria. This communication reports the cloning, purification, crystallization, data collection and initial diffraction data

Table 1
Macromolecule-production information.

Source organism	<i>M. smegmatis</i> MC2 155
DNA source	The genome of <i>M. smegmatis</i>
Forward primer†	GGAATTCATATGGTGACGCGCGCAGTCTTTCAC
Reverse primer‡	CCGGAATTCATGCCGCGCGGTGGCGAG
Cloning vector	pET-28a
Expression vector	pET-28a:MsBioF
Expression host	<i>E. coli</i> BL21 (DE3)
Complete amino-acid sequence of the construct produced	MGSSHHHHHSSGLVPRGSHMMTRAGLSPLAWLADIEQRRRA- EGLRRELVRPPVAAELDLASNDYLGSLQHPDVLGGVEA- LRTWGGGAGGSRVTGNTELHEAFEHQLASFLGAESALVF- SSGYTANLGAVALSGPGLIVSDALSHASLVDACRLSRA- RVVVSPhRDVDAVDAALAARTEARVVVTESVSADGDLA- PLRDLHAVCRRHGALLLVDEAHGLGVRGRGQGLLHEVGL- AGAPDIVMTTTLKALGSGGAVLGPEAVRAHLIDTARSF- IFDTGLAPAAVGAASAAARVLDAEPQARAVLDRAAELAT- IAGVTEAPVSAVSVILGDPEIAVGAAACLDGRGVRVGC- RPPTVPAGTSRLRLAARASLTDEEMALARQLVTDVLATARA

† The *Nde*I site is underlined. ‡ The *Eco*RI site is underlined.

analysis of BioF from *M. smegmatis* (MsBioF), which shares 76% sequence identity with its *Mtb* homologue *Rv1569*.

2. Materials and methods

2.1. Macromolecule production

The *MSMEG_3189* (*MsbioF*) gene was amplified from the chromosomal DNA of *M. smegmatis* using the polymerase chain reaction (PCR). The PCR product was then cloned into pET-28a (Invitrogen) with His₆ at the N-terminus using the *Nde*I and *Eco*RI restriction sites and was confirmed by DNA sequencing. The resulting expression vector pET-28a:MsBioF was transformed into *E. coli* strain BL21 (DE3) and the cells were grown in LB medium containing 30 µg ml⁻¹ kanamycin (Table 1). The transformed *E. coli* BL21 (DE3) cells were grown in the presence of 30 µg ml⁻¹ kanamycin at 310 K until the OD₆₀₀ reached 0.6–0.7. Expression of the protein was induced with 0.1 mM IPTG for 12 h at 289 K in order to maximize the yield of soluble protein. Cells were harvested from a 2 l culture by centrifugation before lysis in buffer A (20 mM Tris–HCl pH 7.9, 500 mM

Table 2
Crystallization.

Method	Hanging-drop vapour diffusion
Plate type	16-well hanging-drop plate
Temperature (K)	298
Protein concentration (mg ml ⁻¹)	5
Buffer composition of protein solution	20 mM Tris–HCl pH 7.9, 100 mM NaCl, 2 mM DTT
Composition of reservoir solution	150 mM DL-malic acid pH 7.4, 100 mM NaCl, 25% PEG 3350
Volume and ratio of drop	2 µl (1:1)
Volume of reservoir (µl)	200

NaCl, 5 mM imidazole, 2 mM β-mercaptoethanol) by ultrasonication. After removal of insoluble debris by centrifugation, the supernatant was applied onto a Ni–NTA (Novagen) column at 4°C and the protein was eluted with buffer B (20 mM Tris–HCl pH 7.9, 500 mM NaCl, 200 mM imidazole, 2 mM β-mercaptoethanol). The eluted protein was loaded onto a HiPrep 26/10 Desalting Column (GE Healthcare Biosciences) to decrease the concentration of NaCl, and the protein was eluted using buffer C (20 mM Tris–HCl pH 7.9, 50 mM NaCl, 2 mM DTT). Fractions containing MsBioF were pooled and loaded onto a HiTrap Q Sepharose anion-exchange column (GE Healthcare Biosciences) pre-equilibrated with buffer B. The pure protein was eluted with a linear salt gradient (0–1 M NaCl) over 20 column volumes. The eluted protein was subjected to size-exclusion chromatography using a Superdex 200 10/300 GL column (GE Healthcare Biosciences) equilibrated with buffer D (20 mM Tris–HCl pH 7.9, 100 mM NaCl, 2 mM DTT) on an ÄKTAprime Plus system (GE Healthcare Biosciences). The purified protein was concentrated to 5 mg ml⁻¹. The molecular weight and purity of the protein were checked on a 12% SDS–PAGE gel (Fig. 1).

2.2. Crystallization

Initial crystallization conditions for the purified protein were screened using Crystal Screen and Crystal Screen 2 (Hampton Research, USA; Table 2). Crystals were obtained after one week under three conditions: (i) 0.1 M Tris pH 8.5, 25% (w/v) polyethylene glycol 3350, (ii) 0.2 M ammonium sulfate, 0.1 M HEPES pH 7.5, 25% (w/v) polyethylene glycol 3350 and (iii) 0.15 M DL-malic acid pH 7.0, 20% (w/v) polyethylene glycol 3350. Diffraction-quality crystals were obtained on further optimization of condition (iii) (Fig. 2). These crystals all appeared in 7 d and diffraction-quality crystals reached maximum dimensions (0.3 mm) in 14 d.

Figure 1
SDS–PAGE of purified MsBioF protein, confirming its molecular weight of 42 kDa. Lane M, molecular-weight markers (labelled in kDa); lane 1, MsBioF protein purified on a Ni–NTA (Novagen) column; lane 2, crude MsBioF protein sample after concentration; lane 3, MsBioF protein successively purified by HiTrap Q Sepharose anion-exchange chromatography and gel-filtration chromatography.

Figure 2
MsBioF crystals grown using the hanging-drop method in 150 mM DL-malic acid pH 7.4, 100 mM NaCl, 25% PEG 3350.

2.3. Data collection and processing

Prior to the collection of diffraction data, crystals were quickly transferred to corresponding cryoprotectant solutions containing 5% (v/v) glycerol and were then flash-cooled in liquid nitrogen. Diffraction data were collected from a single crystal at a wavelength of 0.9795 Å (Fig. 3). The data from 360 images were processed using *MOSFLM* (v.7.0.4; Leslie, 2006) and *SCALA* (v.6.0) from the *CCP4* program suite (Winn *et al.*, 2011). Final data-collection and processing statistics are given in Table 3.

3. Results and discussion

The biotin-synthesis pathway is a potential source of antituberculosis drug targets. The structure and catalytic mechanism of mycobacterial BioF (*Rv1569*), which catalyses the first committed step in the biotin-synthesis pathway, has not been studied in detail. Here, MsBioF, the *M. smegmatis* homologue of *Rv1569*, was successfully cloned, expressed, purified and crystallized. We first cloned the *MsbioF* gene into a pET-28a plasmid with a 6×His tag at its N-terminus and transformed the expression vector pET-28a:*MsbioF* into *E. coli* BL21 (DE3) cells. The integrity of the gene and its transformation were validated by PCR and sequencing, and the expression of MsBioF was confirmed by Western blotting of the synthesized gene product using an anti-His tag monoclonal antibody. The expressed protein was then successively purified using Ni-NTA column chromatography, a desalting column and anion-exchange and gel-filtration chromatography. Protein purity was assessed on a 12% SDS-PAGE gel. The molecular weight of MsBioF was 42 kDa, the same as its predicted size, and the purity of the protein was greater than 95% (Fig. 1). Diffraction data were collected from a single crystal which was crystallized using 150 mM DL-malic acid pH 7.4, 100 mM NaCl, 25% PEG 3350 (Fig. 2). The diffraction resolution was 2.3 Å (Fig. 3). Data processing revealed that the crystal belonged to the monoclinic space group *P*₂₁, with unit-cell parameters *a* = 70.88, *b* = 91.68, *c* = 109.84 Å,

Table 3

MsBioF diffraction data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	Beamline BL17U, SSRF
Wavelength (Å)	0.9795
Temperature (K)	100
Detector	ADSC Quantum 315 CCD
Crystal-to-detector distance (mm)	350
Rotation range per image (°)	1
Total rotation range (°)	360
Exposure time per image (s)	1
Space group	<i>P</i> ₂ ₁
<i>a</i> , <i>b</i> , <i>c</i> (Å)	70.88, 91.68, 109.84
α , β , γ (°)	90.0, 97.8, 90.0
Mosaicity (°)	0.75
Resolution range (Å)	63.03–2.30 (2.42–2.30)
Total No. of reflections	428697 (58947)
No. of unique reflections	61286 (8828)
Completeness (%)	98.9 (98.1)
Multiplicity	7.0 (6.7)
$\langle I/\sigma(I) \rangle$	10.6 (4.9)
$R_{\text{r.i.m.}}$ (%)	11.9 (37.6)
Overall <i>B</i> factor from Wilson plot (Å ²)	36.6

[†] $R_{\text{r.i.m.}} = \sum_{hkl} \{ [N(hkl)/[N(hkl) - 1]]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl) \}$, 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.

$\alpha = 90.0$, $\beta = 97.8$, $\gamma = 90.0^\circ$. According to the molecular weight of MsBioF and the unit-cell parameters, solvent-content analysis indicated that there were either three or four molecules per asymmetric unit, with V_M values (calculated according to Matthews, 1968) of 2.74 and 2.06 Å³ Da^{−1}, respectively, and solvent contents of 55.15 and 40.20%, respectively. Calculation of the self-rotation function (Fig. 4) indicated the presence of at least one noncrystallographic twofold symmetry axis along the *x* or *z* axis, suggesting that there should be an even number of molecules, *i.e.* two or four molecules, per asymmetric unit. In addition, it has been reported that EcaONS is a dimer (Alexeev *et al.*, 1998), implying that most likely all of the bacterial homologues are obligate dimers. It is thus likely that there are four

Figure 3

Diffraction pattern collected from MsBioF crystals. The oscillation width was 1° and the exposure time was 1 s.

Figure 4

Self-rotation function map at $\chi = 180^\circ$ calculated from the diffraction data, showing at least one NCS twofold axis along the *x* or *z* directions. This figure was generated using *MOLREP* (Vagin & Teplyakov, 2010).

rather than three MsBioF molecules per asymmetric unit. Based on both the Matthews coefficient and the self-rotation function, and the previous report for EcAONS, we propose that there are four molecules per asymmetric unit, with a V_M value of $2.06 \text{ \AA}^3 \text{ Da}^{-1}$ and a solvent content of 40.20%. Further structural refinement is currently under way.

We are very grateful to the staff of the Shanghai Synchrotron Radiation Facility (SSRF) for crystal diffraction data collection. This work was funded by the Chinese Ministry of Science and Technology 973 program (grant 2011CB910302).

References

- Alexeev, D., Alexeeva, M., Baxter, R. L., Campopiano, D. J., Webster, S. P. & Sawyer, L. (1998). *J. Mol. Biol.* **284**, 401–419.
- Almeida Da Silva, P. E. & Palomino, J. C. (2011). *J. Antimicrob. Chemother.* **66**, 1417–1430.
- Dey, S., Lane, J. M., Lee, R. E., Rubin, E. J. & Sacchettini, J. C. (2010). *Biochemistry*, **49**, 6746–6760.
- Kurth, D. G., Gago, G. M., de la Iglesia, A., Bazet Lyonnet, B., Lin, T. W., Morbidoni, H. R., Tsai, S. C. & Gramajo, H. (2009). *Microbiology*, **155**, 2664–2675.
- Leslie, A. G. W. (2006). *Acta Cryst.* **D62**, 48–57.
- Lin, S., Hanson, R. E. & Cronan, J. E. (2010). *Nature Chem. Biol.* **6**, 682–688.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Oh, T.-J., Daniel, J., Kim, H.-J., Sirakova, T. D. & Kolattukudy, P. E. (2006). *J. Biol. Chem.* **281**, 3899–3908.
- Ploux, O. & Marquet, A. (1996). *Eur. J. Biochem.* **236**, 301–308.
- Ploux, O., Soularue, P., Marquet, A., Gloeckler, R. & Lemoine, Y. (1992). *Biochem. J.* **287**, 685–690.
- Said, H. M. (2009). *J. Nutr.* **139**, 158–162.
- Sassetti, C. M. & Rubin, E. J. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 12989–12994.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* **D66**, 22–25.
- Webster, S. P., Alexeev, D., Campopiano, D. J., Watt, R. M., Alexeeva, M., Sawyer, L. & Baxter, R. L. (2000). *Biochemistry*, **39**, 516–528.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.