Crystallization and preliminary crystallographic analysis of the putative sugar-binding protein Msmeg_0515 (AgaE) from *Mycobacterium smegmatis*

Feras M. Almourfi, H. Fiona Rodgers, Svetlana E. Sedelnikova and Patrick J. Baker*

Received 10 November 2014

Accepted 1 January 2015

Department of Molecular Biology and Biotechnology, University of Sheffield, Firth Court, Western Bank, Sheffield S10 2TN, England. *Correspondence e-mail: p.baker@sheffield.ac.uk

Keywords: Msmeg_0515; AgaE; *Mycobacterium smegmatis*.

Msmeg_0515, a gene from *Mycobacterium smegmatis* strain 155 encoding the ligand-binding domain, AgaE, of a putative ABC sugar transporter system, has been cloned into a pET-28a vector system, overexpressed in *Escherichia coli* and purified. The truncated protein lacking the first 27 residues, which correspond to a N-terminal signal sequence, was crystallized using the sitting-drop vapour-diffusion technique. The crystals of this protein diffracted to 1.48 Å resolution and belonged to space group $P2_12_12_1$, with unit-cell parameters $a = 64.06$, $b = 69.26$, $c = 100.74$ Å, $\alpha = \beta = \gamma = 90^\circ$ and with one molecule in the asymmetric unit.

1. Introduction

Tuberculosis is the second major global infectious disease after HIV and threatens the lives of approximately one third of the world's population (Dye *et al.*, 1999). The latest World Health Organization estimates concluded that in 2012 tuberculosis caused approximately 1.3 million deaths, with 8.6 million new cases (Eurosurveillance Editorial Team, 2013). Although the introduction of antibacterial drug therapies has helped to decrease the rate of mortality by about 45% since 1990, the mortality rate remains high, partly owing to the emergence of drug-resistant forms of the disease (Eurosurveillance Editorial Team, 2013). There is thus a requirement to develop new antibacterial agents to combat this disease.

The carbon metabolism and the nutrients required for the growth and survival of *Mycobacterium tuberculosis* have been investigated extensively since the discovery of the organism. However, the pathways by which *M. tuberculosis* utilizes essential nutrients inside the human host are still unknown (Sassetti & Rubin, 2003). For example, mycobacterial species use carbohydrates as the carbon source during the initial growth stages (McKinney *et al.*, 2000), but appear to switch to the use of lipids during infection (McKinney *et al.*, 2000; Muñoz-Elías & McKinney, 2005), perhaps as an adaptation to the adaptive immune response (Titgemeyer *et al.*, 2007).

One of the main ways that nutritional molecules such as carbohydrates and lipids enter the *M. tuberculosis* cell is via the ATP-binding cassette (ABC) transporter system (Titgemeyer *et al.*, 2007). This system is a very large superfamily of proteins that are expressed in all kingdoms of life, including bacterial and eukaryotic species (Higgins, 1992). They function as transporters of different macromolecules, such as carbohydrates, amino acids, ions and antibiotics. ABC transporters are divided into two subfamilies based on the direction of substrate translocation as either importers or exporters

(Saurin *et al.*, 1999). The ABC complex is constructed from a number of subunits, with an invariable part that contains two membrane-spanning domains (MSDs), which provide a pathway for specific molecules through the cell membrane, and two cytoplasmic nucleotide-binding domains (NBDs) that provide the energy for the translocation of the substrate molecule by ATP hydrolysis (Licht & Schneider, 2011). In prokaryotic ABC uptake systems, an additional periplasmic binding domain (PBD) is present, which has a highly specific substrate-recognition function, to pass the correct substrate into the transporter (Ames *et al.*, 1990; Gilson *et al.*, 1988). These periplasmic binding domains also play a critical role in cell differentiation, cell communication, cell integrity and pathogenicity (Eitinger *et al.*, 2011).

The *M. tuberculosis* genome possesses 38 putative ABC transporters. Sequence-analysis studies have shown that five of these transporters are most likely to be involved in carbohydrate transport (Titgemeyer *et al.*, 2007). In contrast, within the *M. smegmatis* genome (a nonpathogenic, fast-growing model organism for *M. tuberculosis*), 19 carbohydrate-dependent ABC transporters have been identified, none of which have been studied experimentally (Titgemeyer *et al.*, 2007). One of these putative carbohydrate transporters of *M. smegmatis* is encoded by a proposed operon that may be responsible for the uptake of α -galactosides. This gene cluster comprises *agaA*, encoding an α -galactosidase, *agaF* and *agaK*, which encode two proteins containing membrane-spanning domains, *agaG*, which encodes the nucleotide-binding domain, and *agaE* as the gene that encodes the substrate-binding domain protein (Titgemeyer *et al.*, 2007). This operon is possibly controlled by the transcriptional regulator *agaR*, which belongs to the DeoR/GlpR family of transcriptional regulators (Mortensen *et al.*, 1989; Zeng *et al.*, 1996). *AgaR* is also present in *Escherichia coli*, where it functions as a regulator of genes involved in the metabolism of *N*-acetyl-galactosamine and galactosamine (Ray & Larson, 2004).

The *AgaE* protein (425 amino acids, molecular weight 45.5 kDa) is predicted to be an extracellular membrane-anchored lipoprotein owing to the presence of a lipobox motif (LTAC) in its N-terminal signal sequence, where Cys21 is the residue to which the lipids are attached (Sutcliffe & Harrington, 2004). Furthermore, *AgaE* is suggested to be translocated through the membrane using the twin-arginine translocation (Tat) system as there are two positive residues (Arg3 and Arg4) at the start of the signal sequence (McDonough *et al.*, 2008).

Sequence analysis has revealed that *AgaE* is also present in a number of other mycobacterial species such as *M. mageritense* and *M. cosmeticum* (85 and 84% sequence identity, respectively). In addition, the closest similarity between *AgaE* and a *M. tuberculosis* protein is to the glycerol 3-phosphate-dependent periplasmic binding domain of the ABC transporter UgpB (21% sequence identity), which has been shown to be essential for the survival of *M. tuberculosis in vitro* and is therefore considered to be a vaccine candidate (Sasseti *et al.*, 2001; Zvi *et al.*, 2008). The structure of UgpB has recently been determined, showing that it binds glycerophosphocholine and

Table 1
Macromolecule-production information.

Source organism	<i>M. smegmatis</i>
DNA source	<i>M. smegmatis</i> genomic DNA
Forward primer	5'-CCCCATATGTCGTCATCGGTC-3'
Reverse primer	5'-GGGAAGCTTTTTCGCCGCCAGGG-3'
Cloning vector	pET-28a
Expression vector	pET-28a
Expression host	<i>E. coli</i> BL21 (DE3)
Complete amino-acid sequence of the construct produced (28–425)	SGPVEIAVWHGYQDTEGEAFKGLIDQYNKEHP-DVHVTDLYSSNDLVLQKVLTAVRGGSAPDV-AYMFGSWSPNIAKIPQVVDMSDVVSQSDWN-WDDFYPAEREAATVGDKIVGIPALVDNLAI-VYNKKLFADAGIAPPTADWTWDDFRAAAAK-LTDPAGKQYQWLIIPADGSEDTVWHYVPLW-EAGGDILTPDNEKAAPNSEAGVTALTMLQD-MAVTDKSLYLDTTNENGPKLMSGKVGMLI-TGPWDLSQLSDIDYGVQVMPFAGSSGAHQ-TISGPDNWWVFDNGDKRKQASIDFVKWLTA-PQVQKAFSLQTDLPTRSSVGDDQAVRDQL-DQKLPGSSVFVENLNNAKKARPAVEQYPAT-SEALGQAIIVAVMLGKEQPAALNSAAEAAD-SALAGK

has the commonly observed maltose-binding protein fold, as observed for many periplasmic sugar-binding proteins (Jiang *et al.*, 2014), and it is likely that *AgaE* will also have this fold.

In an attempt to further understand the process of carbohydrate transport in mycobacteria, we have instigated a structural study of *M. smegmatis* *AgaE*; here, we describe the cloning, expression, purification and crystallization of this protein.

2. Materials and methods

2.1. Construct design

Genomic DNA was purified from the nonpathogenic strain *M. smegmatis* strain 155 using a KeyPrep bacterial genomic kit (Anachem). The manufacturer's protocol was followed to give a final concentration of 20–70 μ g of genomic DNA in 100 μ l. The gene encoding the truncated Msmeg_0515 protein (*AgaE*) lacking the signal sequence (residues 28–425) was PCR-amplified using the primers detailed in Table 1, which were designed on the basis of the *M. smegmatis* sequence obtained from the KEGG database (Kanehisa *et al.*, 2014) to include NdeI and HindIII restriction sites, using the BioMix Red (Bioline) PCR reaction mixture according to the manufacturer's instructions for 30 cycles with an annealing temperature of 331 K. DNA fragments were purified using a QIAquick Gel Extraction Kit (Qiagen), restricted for 1 h at 310 K and ligated into pET-28a vector using a T4 ligase (New England Biolabs). Positive transformants were identified by analytical digests.

2.2. Overexpression

The plasmid carrying the *agaE* gene was transformed into *E. coli* BL21 (DE3) cells (Novagen) using a heat-shock protocol (Nishimura *et al.*, 1990) and one colony was used to inoculate 50 ml LB medium supplemented with 50 μ g ml⁻¹ kanamycin in a 250 ml flask and grown overnight at 310 K on a

shaker at 250 rev min⁻¹. To obtain sufficient protein, 5 ml of this culture was used to inoculate 500 ml LB medium in a 2 l flask supplemented as above. The secondary culture was grown at 310 K on a shaking tray (250 rev min⁻¹) for approximately 2 h to reach an optical density (OD₆₀₀) of 0.6. The cells were induced by the addition of 1 mM IPTG and incubated at 310 K on a shaking tray (250 rev min⁻¹) for an additional 4 h. The cells were harvested by centrifugation at 18 600g at 277 K (JLA-10.500 rotor, Avanti J-25I, Beckman) and stored at 253 K.

2.3. Purification

As the AgaE construct contained an N-terminal 6×His tag, Ni-NTA affinity chromatography was used as the initial purification step. 1 g of the cell paste was defrosted, resuspended on ice in 10 ml 50 mM Tris buffer pH 8.0 and disrupted by sonication using a Soniprep 150 machine set at 16 µm amplitude over two cycles of 20 s each. Cell debris was removed by centrifugation at 70 000g for 10 min. The soluble protein fractions (~20 ml) were applied onto a 5 ml Ni-HP cartridge column and the AgaE protein was eluted from the column using a 50 ml linear gradient of 0–0.35 M imidazole in 50 mM Tris-HCl pH 8.0, 0.5 M NaCl. AgaE eluted from the column at about 0.15 M imidazole, and the two fractions containing the highest protein concentration were combined for further purification by gel filtration. The sample was reduced to 1 ml using a Vivaspin concentrator (MWCO 30 000) and was injected into a HiLoad 16/60 Superdex 200 gel-filtration column (GE Healthcare) equilibrated with 50 mM Tris pH 8.0, 0.5 M NaCl buffer. Gel filtration was performed at a flow rate of 1.5 ml min⁻¹ and 2 ml fractions were collected. The elution volume for AgaE suggested that the apparent molecular weight of the protein in solution was

Figure 1
SDS-PAGE (NuPAGE 4–12% bis-tris gel, Invitrogen) showing the purification steps of *M. smegmatis* AgaE. Lane 1, Mark12 (Invitrogen; labelled in kDa); lane 2, cell-free extract; lane 3, pooled fractions from a 5 ml Ni-HP cartridge column; lane 4, pooled fractions from a HiLoad 16/60 Superdex 200 gel-filtration column (GE Healthcare).

Table 2
Crystallization.

Method	Sitting-drop vapour diffusion
Plate type	MRC 2 drop 96-well
Temperature (K)	289
Protein concentration (mg ml ⁻¹)	11.5
Buffer composition of protein solution	10 mM Tris pH 8.0
Composition of reservoir solution	0.2 M ammonium chloride pH 6.3, 20% (w/v) PEG 3350
Volume and ratio of drop	0.2 µl + 0.2 µl
Volume of reservoir (µl)	40

about 47.6 kDa, suggesting that AgaE is a monomer. SDS-PAGE analysis showed a strong band at around 45.5 kDa corresponding to AgaE (Fig. 1).

2.4. Crystallization

Prior to crystallization, pooled samples of purified truncated AgaE protein (residues 28–425) were concentrated using a Vivaspin centrifugal concentrator with a 30 kDa molecular-weight cutoff (Sartorius, Germany) and the buffer was exchanged to 10 mM Tris pH 8.0. The final protein concentration was approximately 11.5 mg ml⁻¹ (Bradford, 1976). Preliminary crystallization conditions were screened with NeXtal Suites (The JCSG+, PACT, PEG and Classics Suites) using a Matrix Hydra II Plus One crystallization robot (Thermo Fisher Scientific, USA) as detailed in Table 2.

2.5. Data collection

Crystals were washed with a cryoprotectant solution (crystallization buffer and 25% ethylene glycol), mounted on the diffractometer using a fibre loop, flash-cooled to 100 K using an Oxford Cryosystems Cryostream 700 and tested using a Rigaku MicroMax-007 copper rotating-anode generator, Rigaku confocal optics and a MAR Research MAR345 image plate. Subsequently, data were collected on beamline I04 at the Diamond Light Source synchrotron.

Figure 2
A photomicrograph of a crystal of *M. smegmatis* AgaE grown from 0.2 M ammonium chloride pH 6.3, 20% (w/v) PEG 3350.

3. Results and discussion

Inspection of the crystallization trials showed that crystals grew after 3 d in two different conditions from The JSCG+ Suite: A9 [0.2 M ammonium chloride pH 6.3, 20% (w/v) PEG 3350] and B9 [0.1 M citric acid pH 5.0, 20% (w/v) PEG 600] (Fig. 2). Initial X-ray diffraction images from crystals grown in either of these conditions showed that the crystals were of similar quality and diffracted X-rays to a similar resolution. A crystal grown from condition A9 (Fig. 3) was transferred to beamline I04 of the Diamond Light Source and data were collected using 12.8 keV X-rays and an ADSC Q315 detector. A total of 230 images of 0.5° rotation and 1.0 s exposure were collected to a resolution of 1.48 Å and processed with the *xia2* pipeline (Winter, 2010) using *XDS/XSCALE* (Kabsch, 2010*a,b*), *POINTLESS* (Evans, 2006), *AIMLESS* (Winter, 2010) and other programs from the *CCP4* suite (Winn *et al.*, 2011). Initial autoindexing showed that the crystal of the putative sugar-binding protein AgaE belonged to the primitive orthorhombic class with point group *P222*, with unit-cell parameters $a = 64.06$, $b = 69.26$, $c = 100.74$ Å, $\alpha = \beta = \gamma = 90^\circ$. Inspection of the data and analysis of the systematic absences showed that a $2n$ condition was present for the $h00$, $0k0$ and $00l$ reflections, indicating that the space group was most likely to be *P2₁2₁2₁*. Data-collection and processing statistics are shown in Table 3.

Calculation of possible values for V_M , assuming that the crystallized truncated protein had a molecular weight of 42.8 kDa, showed that the asymmetric unit contained a single AgaE molecule with a V_M of $2.6 \text{ Å}^3 \text{ Da}^{-1}$ (Matthews, 1976),

Figure 3
A representative 0.5° rotation image of data collected from a crystal of the putative ligand-binding domain AgaE of an ABC transporter system from *M. smegmatis* on beamline I04 at Diamond Light Source, Oxford, England. An enlarged view of the region indicated by the square is shown on the right. The blue circle is drawn at 1.5 Å resolution.

Table 3

Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

Diffraction source	I04, Diamond Light Source
Wavelength (Å)	0.96863
Temperature (K)	100
Detector	ADSC Q315
Crystal-to-detector distance (mm)	183.7
Temperature (K)	100
Rotation range per image (°)	0.5
Total rotation range (°)	115.5
Exposure time per image (s)	1.0
Space group	<i>P2₁2₁2₁</i>
a , b , c (Å)	64.06, 69.26, 100.74
α , β , γ (°)	90, 90, 90
Mosaicity (°)	0.18
Resolution range (Å)	34.6–1.48 (1.52–1.48)
Total No. of reflections	75131 (5493)
$R_{\text{p.i.m.}}$	0.029 (0.35)
R_{merge}	0.047 (0.59)
Completeness (%)	99.7 (99.9)
Multiplicity	4.6 (4.6)
$\langle I/\sigma(I) \rangle$	15.8 (2.3)
Overall B factor from Wilson plot (Å ²)	17.3

corresponding to a solvent content of 53%. The sequence of AgaE suggests that it has the maltose-binding domain fold and thus initial attempts to solve the structure using the molecular-replacement method with *Phaser* (Winn *et al.*, 2011) and the structure of the closest sequence homologue in the PDB, the maltose-binding protein from *Alicyclobacillus acidocaldarius* (PDB entry 1urd, 27% sequence identity; Schäfer *et al.*, 2004), were performed. However, no plausible solutions were found. As members of this protein family are known to undergo domain movements on substrate binding, further searches using the two domains of PDB entry 1urd independently were attempted, but to no avail. Attempts are now being made to determine the structure of AgaE by expressing and crystallizing a selenomethionine derivative in order to exploit the anomalous scattering method of phase determination and therefore to remove any possible bias arising from incorrect domain orientation in a molecular-replacement search model.

Acknowledgements

We thank Professor Jeff Green for the kind gift of a *M. smegmatis* cell culture. FMA acknowledges the Ministry of Higher Education, Saudi Arabia for scholarship funding. We acknowledge Diamond Light Source for provision of beamtime and thank Dr Ralf Flaig for assistance with station I04.

References

- Ames, G. F., Mimura, C. S. & Shyamala, V. (1990). *FEMS Microbiol. Rev.* **6**, 429–446.
- Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
- Dye, C., Scheele, S., Dolin, P., Pathania, V. & Ravigliione, M. C. (1999). *JAMA*, **282**, 677–686.
- Eitinger, T., Rodionov, D. A., Grote, M. & Schneider, E. (2011). *FEMS Microbiol. Rev.* **35**, 3–67.
- Eurosurveillance Editorial Team (2013). *Euro Surveill.* **18**(43), pii:20615.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Glison, E., Alloing, G., Schmidt, T., Claverys, J. P., Dudler, R. & Hofnung, M. (1988). *EMBO J.* **7**, 3971–3974.

- Higgins, C. F. (1992). *Annu. Rev. Cell Biol.* **8**, 67–113.
- Jiang, D., Zhang, Q., Zheng, Q., Zhou, H., Jin, J., Zhou, W., Bartlam, M. & Rao, Z. (2014). *FEBS J.* **281**, 331–341.
- Kabsch, W. (2010a). *Acta Cryst.* **D66**, 125–132.
- Kabsch, W. (2010b). *Acta Cryst.* **D66**, 133–144.
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. (2014). *Nucleic Acids Res.* **42**, D199–205.
- Licht, A. & Schneider, E. (2011). *Cent. Eur. J. Biol.* **6**, 785–801.
- Matthews, B. W. (1976). *Annu. Rev. Phys. Chem.* **27**, 493–523.
- McDonough, J. A., McCann, J. R., Tekippe, E. M., Silverman, J. S., Rigel, N. W. & Braunstein, M. (2008). *J. Bacteriol.* **190**, 6428–6438.
- McKinney, J. D., Höner zu Bentrup, K., Muñoz-Elías, E. J., Miczak, A., Chen, B., Chan, W.-T., Swenson, D., Sacchettini, J. C., Jacobs, W. R. Jr & Russell, D. G. (2000). *Nature (London)*, **406**, 735–738.
- Mortensen, L., Dandanell, G. & Hammer, K. (1989). *EMBO J.* **8**, 325–331.
- Muñoz-Elías, E. J. & McKinney, J. D. (2005). *Nature Med.* **11**, 638–644.
- Nishimura, A., Morita, M., Nishimura, Y. & Sugino, Y. (1990). *Nucleic Acids Res.* **18**, 6169.
- Ray, W. K. & Larson, T. J. (2004). *Mol. Microbiol.* **51**, 813–826.
- Sassetti, C. M., Boyd, D. H. & Rubin, E. J. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 12712–12717.
- Sassetti, C. M. & Rubin, E. J. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 12989–12994.
- Saurin, W., Hofnung, M. & Dassa, E. (1999). *J. Mol. Evol.* **48**, 22–41.
- Schäfer, K., Magnusson, U., Scheffel, F., Schiefner, A., Sandgren, M. O. J., Diederichs, K., Welte, W., Hülsmann, A., Schneider, E. & Mowbray, S. L. (2004). *J. Mol. Biol.* **335**, 261–274.
- Sutcliffe, I. C. & Harrington, D. J. (2004). *FEMS Microbiol. Rev.* **28**, 645–659.
- Titgemeyer, F., Amon, J., Parche, S., Mahfoud, M., Bail, J., Schlicht, M., Rehm, N., Hillmann, D., Stephan, J., Walter, B., Burkovski, A. & Niederweis, M. (2007). *J. Bacteriol.* **189**, 5903–5915.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.
- Zeng, G., Ye, S. & Larson, T. J. (1996). *J. Bacteriol.* **178**, 7080–7089.
- Zvi, A., Ariel, N., Fulkerson, J., Sadoff, J. C. & Shafferman, A. (2008). *BMC Med. Genomics*, **1**, 18.