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

Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of macrophage growth locus A (MglA) protein from *Francisella tularensis*

Priadarsini Subburaman, Brian P. Austin, Gary X. Shaw, David S. Waugh and Xinhua Ji*

Macromolecular Crystallography Laboratory, National Cancer Institute, 1050 Boyles Street, Frederick, MD 21702, USA

Correspondence e-mail: jix@mail.nih.gov

Received 27 January 2010
Accepted 15 March 2010

Francisella tularensis, a potential bioweapon, causes a rare infectious disease called tularemia in humans and animals. The macrophage growth locus A (MglA) protein from *F. tularensis* associates with RNA polymerase to positively regulate the expression of multiple virulence factors that are required for its survival and replication within macrophages. The MglA protein was overproduced in *Escherichia coli*, purified and crystallized. The crystals diffracted to 7.5 Å resolution at the Advanced Photon Source, Argonne National Laboratory and belonged to the hexagonal space group $P6_1$ or $P6_5$, with unit-cell parameters $a = b = 125$, $c = 54$ Å.

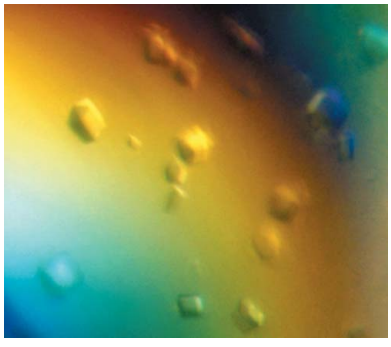
1. Introduction

Francisella tularensis is a pathogenic Gram-negative bacterium that causes a serious infectious disease called tularemia in humans and animals. Tularemia can be acquired by inhalation, from contaminated food or water, by handling infected animal carcasses or from being bitten by infected ticks, deer-biting flies or mosquitoes (Cross & Penn, 2000). *F. tularensis* can persist in the environment for weeks in contaminated animal carcasses, soil, water and vegetation. It can cause several different disease syndromes depending on the route of entry. Infection of *F. tularensis* through the skin or mucous membranes leads to the ulceroglandular form of tularemia, which is the most common form of the disease in humans. Pneumonic and typhoidal forms of tularemia are rare but can be fatal if treatment is not initiated promptly (Ellis *et al.*, 2002).

F. tularensis has long been considered to be a potential biological weapon and was the subject of military research in the United States, the former Soviet Union and Japan. The Centers for Disease Control and Prevention considers *F. tularensis* to be a potential biological warfare agent because it is easy to aerosolize, highly incapacitating (as few as ten bacteria can cause disease) and easy to disseminate (Dennis *et al.*, 2001).

During infection of mammalian hosts, *F. tularensis* survives and replicates inside macrophages. The bacteria rapidly escape the phagosome and replicate within the cytoplasm of host cells (Lauriano *et al.*, 2004). The genome of *F. tularensis* contains the *Francisella* pathogenicity island (FPI), a region that encodes a cluster of 16–19 genes that contribute to the virulence activity and intramacrophage growth of the bacteria (Gray *et al.*, 2002). Studies show that macrophage growth locus A (MglA; Swiss-Prot entry Q5NFG1) regulates more than 100 genes, including all of the FPI genes (Lauriano *et al.*, 2004).

The MglA protein is the only known regulator of virulence-gene expression and is homologous to stringent starvation protein A (SspA; Swiss-Prot entry Q7CL96), with a sequence identity of 18% and a sequence similarity of 50%. The 70 N-terminal amino-acid residues of MglA and SspA are similar to the N-terminal domain of gluthathione *S*-transferase (GST; Swiss-Prot entry P15214), with a sequence identity of 14% and a sequence similarity of 34% for the three N-terminal regions. However, both MglA and SspA lack GST



function as the active-site cysteine residue is replaced by a tyrosine in MglA and a phenylalanine in SspA (Guina *et al.*, 2007; Hansen *et al.*, 2005).

The *F. tularensis* genome contains an SspA homolog, which is also annotated as SspA (Charity *et al.*, 2007). MglA and SspA physically interact with one another and associate with RNA polymerase (RNAP) as a heterodimer. They can also form homomeric complexes, but these are unlikely to associate with RNAP. MglA and SspA co-regulate the expression of the same set of genes. The ability of MglA to associate with RNAP is dependent on SspA (Charity *et al.*, 2007). Solving the structure of MglA may improve our understanding of how it exerts its effects on virulence-gene expression. The conformational changes involved in its association with RNAP and other structural details may prove very useful in the design of an effective vaccine against *F. tularensis*.

2. Materials and methods

2.1. Cloning and expression of MglA from *F. tularensis*

The open reading frame (ORF) encoding MglA was amplified from genomic DNA by polymerase chain reaction (PCR) using the following oligonucleotide primers: forward primer 5′-GAG AAC CTG TAC TTC CAG TCC CAT CAT CAT CAT CAT CAT ATG CTT TTA TAC ACA AAA AAA GAT G-3′ and reverse primer 5′-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTT ATT AAG CTC CTT TTG CTT TGA TAG-3′. The resulting PCR amplicon was subsequently used as the template for a second PCR with primer PE-277 (Evdokimov *et al.*, 2002) and the reverse primer. The final PCR amplicon was inserted into pDONR201 (Invitrogen, Carlsbad, California, USA) by recombinational cloning and the nucleotide sequence of the insert was confirmed experimentally. Next, the MglA ORF, together with DNA encoding a tobacco etch virus (TEV) protease-cleavage site followed by a hexahistidine (His_6) tag in-frame to its N-terminus, was recombined into pSN1542 (Nallamsetty & Waugh, 2006) to generate the plasmid expression vector. The NusA-TEV-His_6-MglA fusion protein was expressed in *Escherichia coli* BL21 (DE3)-RIL CodonPlus cells (Stratagene, Valencia, California, USA). The cells were grown at 310 K in Luria broth (LB) containing 100 mg ml^{-1} ampicillin and 30 mg ml^{-1} chloramphenicol to mid-log phase (OD_{600} = 0.5), whereupon isopropyl-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 m*M* and the temperature was reduced to 303 K. 4 h after induction, the cells were pelleted by centrifugation and stored at 193 K.

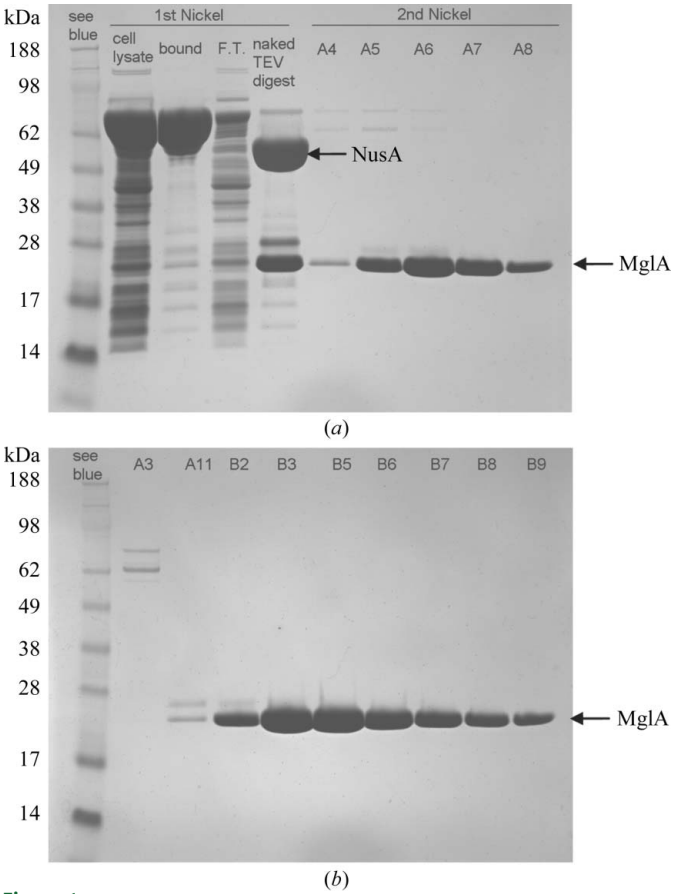


Figure 1
Protein purification. (*a*) First pass over Ni–NTA column followed by overnight TEV cleavage and subsequent second Ni–NTA-bound fractions. (*b*) Gel-filtration chromatography.

2.2. Purification of MglA

All chromatography steps were carried out at 277 K. *E. coli* cell paste was suspended in ice-cold 50 m*M* phosphate-buffered saline (PBS) pH 7.5, 25 m*M* imidazole, 150 m*M* NaCl (buffer *A*) containing Complete EDTA-free protease-inhibitor cocktail (Roche Molecular Biochemicals, Indianapolis, Indiana, USA). The cells were lysed with an APV-1000 homogenizer (Invensys, Roholmsvej, Denmark) at 69 MPa and centrifuged at 30 000*g* for 30 min. The supernatant was filtered through a 0.45 µm cellulose acetate membrane and loaded onto an Ni–NTA Superflow column (Qiagen, Valencia, California, USA) equilibrated in buffer *A*. The column was washed to baseline with buffer *A*. The protein was eluted with a linear gradient to 250 m*M* imidazole. Fractions containing the fusion protein were pooled, cleaved overnight with untagged TEV protease and then subjected to another round of Ni–NTA chromatography as described above. The elution fractions containing MglA with a Ser residue and the His_6 tag at the N-terminus (Ser-His_6-MglA) were pooled, concentrated to ∼10 ml using an Amicon stirred cell with a YM10 membrane (Millipore, Billerica, Massachusetts, USA) and applied onto a 26/60 Sephacryl-100 preparative size-exclusion column (GE Healthcare, Piscataway, New Jersey, USA) equilibrated in 25 m*M* Tris pH 7.5, 150 m*M* NaCl, 2 m*M* tris(2-carboxyethyl)phosphine hydrochloride (TCEP); the peak fractions corresponding to Ser-His_6-MglA were pooled and concentrated to 32 mg ml^{-1} with an Amicon stirred cell (Millipore, Billerica, Massachusetts, USA). The molecular weight of Ser-His_6-MglA was confirmed by electrospray mass spectrometry.

2.3. Crystallization

Initial crystallization screens for Ser-His_6-MglA (in the final sizing buffer) were carried out by the sitting-drop vapor-diffusion method using screening kits from Hampton Research (Aliso Viejo, California, USA) and Qiagen (Valencia, California, USA). Crystallization screening was performed at room temperature using a Hydra II Plus One robot (Matrix Liquid Handling Solutions, Thermo Fisher Scientific, Waltham, Massachusetts, USA). 1:3, 1:1 and 3:1 ratios of protein:well solution were used for each condition and the size of the sitting drops was 1.2 µl. Index condition D12 [0.2 *M* calcium chloride, 0.1 *M* bis-tris pH 5.5, 45%(*v*/*v*) MPD] from Hampton Research gave crystals, but the crystals did not diffract X-rays at all and therefore glutathione (GSH) was used as an additive. The Ser-His_6-MglA protein and GSH were mixed in a 1:5 molar ratio and incubated for 3 h at room temperature. The precipitate was removed from the mixture by centrifugation and crystallization conditions were

screened as described above. Small crystals were observed from Hampton Research Index condition E9, consisting of 0.05 *M* ammonium sulfate, 0.05 *M* bis-tris pH 6.5 and 30%(*v*/*v*) pentaerythritol ethoxylate (15/4 EO/OH), using a 1:1 ratio of well:protein solution.

2.4. X-ray diffraction analysis

The crystals were soaked in a cryoprotectant solution comprising reservoir solution plus 20%(*v*/*v*) ethylene glycol. Subsequently, the crystals were flash-frozen in liquid nitrogen. The preliminary X-ray diffraction experiment was carried out at the Southeast Regional Collaborative Access Team (SER-CAT) facility at the Advanced Photon Source, Argonne National Laboratory. Other cryoprotectants were not tested owing to difficulty in producing a sufficient number of crystals.

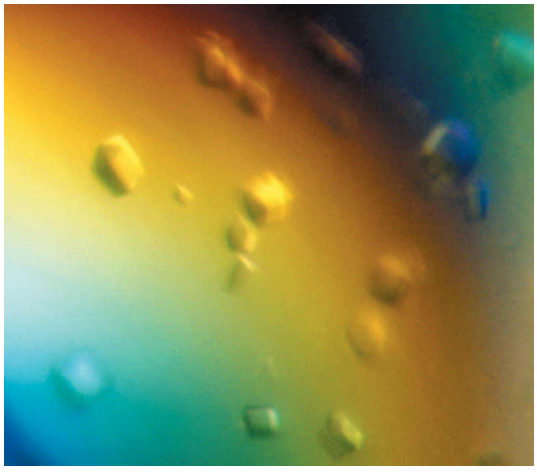


Figure 2
Crystals of Ser-His$_6$-MglA viewed under polarized light. The dimensions of these crystals were approximately 0.05 × 0.05 × 0.03 mm.

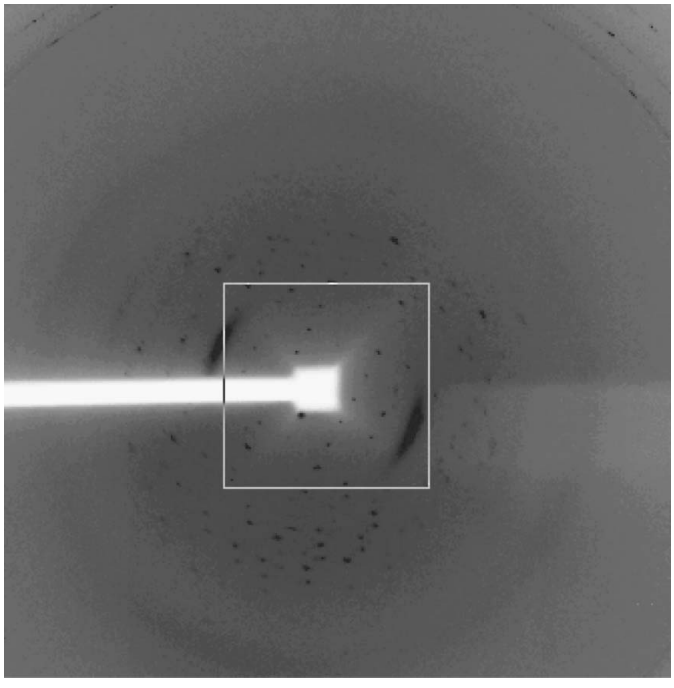


Figure 3
X-ray diffraction image of the Ser-His$_6$-MglA crystals reaching 7.5 Å resolution.

Table 1
Preliminary X-ray diffraction statistics.

Synchrotron-radiation source	22-ID, SER-CAT, APS
Wavelength (Å)	1.0
Space group	$P6_5$ or $P6_1$
Unit-cell parameters (Å)	$a = b = 125$, $c = 54$
Total No. of reflections	1031
No. of unique reflections	588
Resolution (Å)	7.5
Completeness (%)	90.3
$\langle I/\sigma(I)\rangle$	9.0
$R_{\rm merge}$† (%)	14.2

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$.

3. Results and discussion

3.1. Expression and purification

MglA was successfully overexpressed in *E. coli* as a fusion to the C-terminus of NusA. The linker between NusA and MglA contained a TEV protease cleavage site followed by a His$_6$ tag. The fusion protein bound to the Ni–NTA column with about 80% efficiency. The His$_6$ tag was placed distal to the TEV protease cleavage site because earlier work indicated that an MglA fusion protein with four consecutive glycine residues following the TEV site was not cleavable. This may have been a consequence of steric occlusion (*i.e.* the protease-cleavage site is too close to ordered structure in the target protein). The addition of His$_6$ between the TEV site and the N-terminus of MglA gave rise to a fusion protein that could be cleaved with greater than 98% efficiency as estimated by SDS–PAGE (Fig. 1*a*). Elution fractions A4–A8 from the second Ni–NTA column were pooled and loaded onto a gel-filtration column, which separated the impurities from Ser-His$_6$-MglA. Fractions B2–B9 (Fig. 1*b*) were pooled and concentrated to 30 mg ml^{-1} for crystallization trials.

3.2. Crystallization and X-ray diffraction analysis

Since the N-terminal 70 amino-acid residues of MglA are similar to the active domain of GST, we used GSH as an additive to improve the quality of the crystals. Crystals of Ser-His$_6$-MglA were observed from Hampton Research Index condition E9 and the crystals (Fig. 2) diffracted to 7.5 Å resolution (Fig. 3) at the synchrotron-radiation source. Preliminary X-ray diffraction analysis revealed that the crystals belonged to space group $P6_1$ or $P6_5$, with unit-cell parameters $a = b = 125$, $c = 54$ Å, $\alpha = \beta = 90$, $\gamma = 120^\circ$. Preliminary X-ray diffraction statistics are summarized in Table 1.

The Matthews coefficient and solvent content, which we estimated to be 2.5 Å^3 Da^{-1} and 51%, respectively, are consistent with the presence of two monomers in the asymmetric unit (Matthews, 1968).

We tried to optimize the crystals around crystallization condition E9 from Index screen (Hampton Research) by changing the salt concentration and precipitant concentration. We also tried to add additives such as ethylene glycol and glycerol, but could not change the crystal lattice. Microseeding was attempted but was unsuccessful. We also attempted reductive methylation and thermolysin digestion, but no crystals were obtained using either methylated or digested protein. We are in the process of trying surface-entropy reduction and cyclic pentylation in the hope of obtaining better diffracting crystals.

This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. Electrospray mass-spectrometry experiments were conducted on the LC/ES-MS instrument maintained by the Biophysics Resource in the Structural Biophysics Laboratory, Center for Cancer Research,

National Cancer Institute at Frederick. X-ray diffraction data were collected on the SER-CAT 22-ID beamline of the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at http://www.ser-cat.org/members.html. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract No. W-31-109-Eng-38.

References

Charity, J. C., Costante-Hamm, M. M., Balon, E. L., Boyd, D. H., Rubin, E. J. & Dove, S. L. (2007). *PLoS Pathog.* **3**, e84.

Cross, J. T. & Penn, R. L. (2000). *Principles and Practice in Infectious Diseases*, 5th ed., edited by G. Mandell, J. Bennett & R. Dolin, ch. 216. Philadelphia: Churchill Livingstone.

Dennis, D. T. *et al.* (2001). *JAMA*, **285**, 2763–2773.

Ellis, J., Oyston, P. C., Green, M. & Titball, R. W. (2002). *Clin. Microbiol. Rev.* **15**, 631–646.

Evdokimov, A. G., Tropea, J. E., Routzahn, K. M. & Waugh, D. S. (2002). *Acta Cryst.* D**58**, 398–406.

Gray, C. G., Cowley, S. C., Cheung, K. K. & Nano, F. E. (2002). *FEMS Microbiol. Lett.* **215**, 53–56.

Guina, T., Radulovic, D., Bahrami, A. J., Bolton, D. L., Rohmer, L., Jones-Isaac, K. A., Chen, J., Gallagher, L. A., Gallis, B., Ryu, S., Taylor, G. K., Brittnacher, M. J., Manoil, C. & Goodlett, D. R. (2007). *J. Bacteriol.* **189**, 6580–6586.

Hansen, A. M., Gu, Y., Li, M., Andrykovitch, M., Waugh, D. S., Jin, D. J. & Ji, X. (2005). *J. Biol. Chem.* **280**, 17380–17391.

Lauriano, C. M., Barker, J. R., Yoon, S. S., Nano, F. E., Arulanandam, B. P., Hassett, D. J. & Klose, K. E. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 4246–4249.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Nallamsetty, S. & Waugh, D. S. (2006). *Protein Expr. Purif.* **45**, 175–182.