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

Cloning, expression, purification, crystallization and preliminary X-ray studies of the mannose-binding lectin domain of MSMEG_3662 from *Mycobacterium smegmatis*

Dhabaleswar Patra, Alok Sharma, Divya Chandran and Mamannamana Vijayan*

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Correspondence e-mail: mv@mbu.iisc.ernet.in

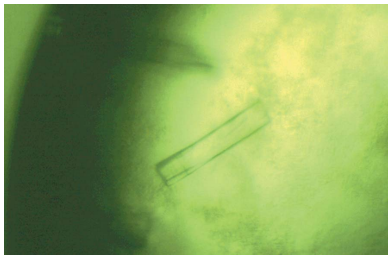
Received 19 February 2011
Accepted 13 March 2011

The mannose-binding lectin domain of MSMEG_3662 from *Mycobacterium smegmatis* has been cloned, expressed, purified and crystallized and the crystals have been characterized using X-ray diffraction. The Matthews coefficient suggests the possibility of two lectin domains in the triclinic cell. The amino-acid sequence of the domain indicates structural similarity to well characterized β-prism II fold lectins.

1. Introduction

During the past couple of decades, we have been pursuing a programme of structural, primarily X-ray crystallographic, studies of plant lectins (Banerjee *et al.*, 1994; Sankaranarayanan *et al.*, 1996; Vijayan & Chandra, 1999; Ramachandraiah *et al.*, 2003; Jeyaprakash *et al.*, 2004; Singh *et al.*, 2005; Kulkarni *et al.*, 2007; Natchiar *et al.*, 2007; Sharma *et al.*, 2009; Chandran, Sharma *et al.*, 2010; Sharma & Vijayan, 2011). Lectins are well known as multivalent proteins of non-immune origin that specifically bind different carbohydrates. This specificity forms the basis of their biological role in symbiosis, cell–cell interactions, innate immunity, mitogenesis *etc.* (Bohlool & Schmidt, 1974; Neeser *et al.*, 1989; Feizi, 2000; Dam & Brewer, 2010). Plant and animal lectins have been extensively characterized (Hardman & Ainsworth, 1972; Becker *et al.*, 1975; Lobsanov *et al.*, 1993; Sharon & Lis, 2004; Taylor & Drickamer, 2007). Some outstanding work has been reported on lectins such as influenza virus haemagglutinin (Wiley *et al.*, 1981; Skehel & Wiley, 2000) and cholera toxin (Zhang *et al.*, 1995; Merritt *et al.*, 1998). However, microbial lectins constitute an underexplored area, although their role in adhesion to the host, including host–pathogen interactions, has been well recognized. This is particularly true of those from mycobacteria, which include several pathogenic species. Therefore, our ongoing structural studies on mycobacterial proteins (Vijayan, 2005; Krishna *et al.*, 2007; Selvaraj *et al.*, 2007; Kaushal *et al.*, 2008; Roy *et al.*, 2008; Prabu *et al.*, 2009; Chandran, Prabu *et al.*, 2010; Chetnani *et al.*, 2010) were extended to lectins and lectin-like molecules from mycobacteria. As a first step in this direction, 11 such molecules or domains were identified in *Mycobacterium tuberculosis* through a bioinformatics analysis (Singh *et al.*, 2007). Preliminary crystallographic studies on one of these, which represent the first structural investigation of a mycobacterial lectin, have recently been reported from this laboratory (Patra *et al.*, 2010). Here, we report the crystallization and preliminary X-ray studies of a lectin domain from *M. smegmatis*.

A recent sequence-database search for β-prism fold lectins in different organisms (Sharma *et al.*, 2007) led to the identification of ABK70862.1 (MSMEG_3662) in *M. smegmatis* as a protein containing a β-prism II fold lectin domain. This domain has no homologue in *M. tuberculosis*. Sequence analysis indicated that the protein contains the lectin domain at the N-terminal region



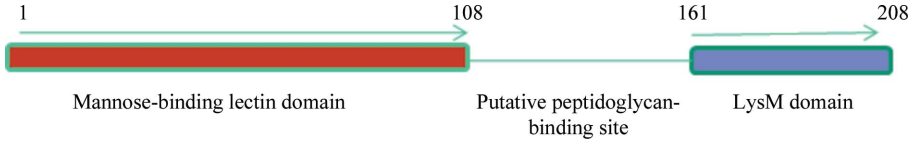


Figure 1
Domain organization of MSMEG_3662.

connected by a long polypeptide stretch to a lysine motif (LysM) domain (Buist *et al.*, 2008) at the C-terminus. The first half of the 208-residue protein constitutes the lectin domain, which forms part of carbohydrate-binding module 50 (CBM50; http://www.cazy.org) as characterized by Boraston *et al.* (2004). This domain has been cloned, expressed and purified. The purified protein has been crystallized and the crystals have been characterized using X-ray diffraction.

2. Materials and methods

2.1. Cloning, expression and purification

The domain organization of the protein encoded by *M. smegmatis* gene *MSMEG_3662* (Fig. 1) was predicted by the Conserved Domain Database (CDD), which indicated two domains: an N-terminal lectin domain of 11.77 kDa (the first 108 residues) and a C-terminal LysM domain of 5.15 kDa (residues 161–208). The DNA sequence coding for the lectin domain was PCR-amplified using the forward primer 5′-AGCA**GCTAGC**ATGGGCGACACTTTGAC-3′ and the reverse primer 5′-GA**CTCGAG**GGGGATCGTCAGGACC-3′ (the *Nhe*I and *Xho*I restriction sites in the forward and reverse primers, respectively, are shown in bold). The last three residues of the lectin domain were omitted to minimize the region that was predicted to be unstructured (Prilusky *et al.*, 2005). The PCR involved 30 cycles, each consisting of initial denaturation at 368 K for 45 s, annealing at 323 K for 1 min and extension at 345 K for 1 min. The PCR products were digested with the restriction enzymes *Nhe*I and *Xho*I and ligated to a pET21b vector (Novagen). The clone was confirmed by DNA sequencing. The construct thus obtained with LEHHHHHH at the C-terminus was used for overproduction of the lectin domain.

The recombinant construct was transformed into *Escherichia coli* strain BL21 (DE) cells for overproduction of the lectin domain. The cells were grown in 2 l LB broth containing 100 µg ml^{-1} ampicillin at 310 K in an incubator–shaker until the OD of the culture at 595 nm reached a value of ∼0.6. The culture was then induced with 1 m*M* IPTG and incubated at 289 K for a further 16 h in an orbital shaker. The cells were harvested by centrifugation at 6000*g* for 15 min, resuspended in a buffer consisting of 30 m*M* Tris–HCl pH 7.4, 300 m*M* NaCl, 5 m*M* imidazole, 10 m*M* β-mercaptoethanol and 10%(*v*/*v*) glycerol and lysed by sonication. The cell lysate was spun at 15 000*g* for 30 min and the supernatant was loaded onto an Ni–NTA column which had been equilibrated with the same buffer. After adequate washing with the abovementioned buffer, the lectin domain was eluted with an imidazole gradient from 5 to 300 m*M*. Fractions containing the pure protein were pooled and dialyzed against Tris–HCl buffer (50 m*M* Tris–HCl pH 7.4, 150 m*M* NaCl). The purity and the mass of the protein were confirmed by SDS–PAGE (Fig. 2*a*) and mass spectrometry (MALDI–TOF; Fig. 2*b*), respectively. The mannose-binding activity of the protein was confirmed through binding of the protein to a mannose Sepharose column. Concentration of protein for crystallization experiments resulted in precipitation, which was prevented by the inclusion of 100 m*M* imidazole in the dialysis buffer. For crystallization experiments, the protein was concentrated to 8 mg ml^{-1} using an Amicon Ultra centrifugal filter device of 3 kDa cutoff.

2.2. Crystallization

Crystallization experiments were carried out using the hanging-drop vapour-diffusion method with 2 µl protein solution and 2 µl precipitant solution in each drop. The well in each case contained 400 µl precipitant solution. The commercially available Crystal Screen and Crystal Screen 2 from Hampton Research were used in the experiments. After 10 d, rectangular plate-shaped crystals appeared in Crystal Screen condition No. 38 consisting of 1.4 *M* trisodium citrate and 0.1 *M* Na HEPES buffer pH 7.5, but they diffracted

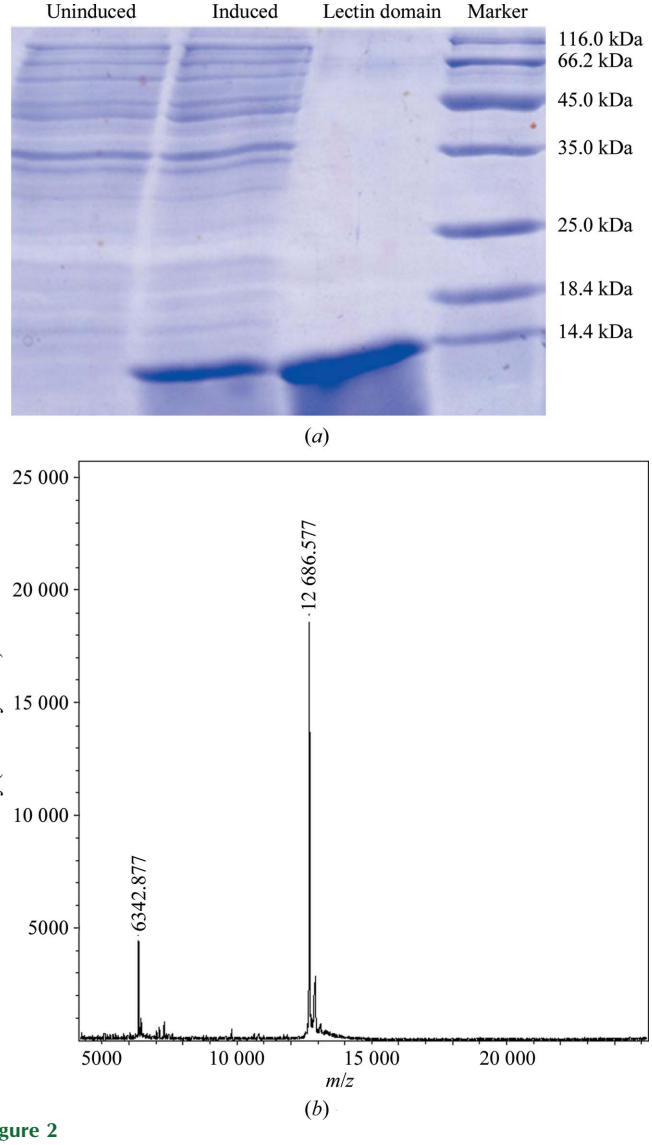


Figure 2
(*a*) Electrophoretic profile of the lectin domain and (*b*) mass-spectrometric analysis of the purified protein.

Table 1
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

Wavelength (Å)	1.5418
Crystal-to-detector distance (mm)	200
Rotation range per image (°)	1
Total rotation range (°)	362
Exposure time per image (min)	10
Resolution range (Å)	45.00–2.40 (2.53–2.40)
Space group	*P*1
Unit-cell parameters (Å, °)	a = 22.98, b = 42.56, c = 57.56, α = 84.59, β = 85.94, γ = 84.34
Mosaicity (°)	0.92
Total No. of measured intensities	30338 (4376)
Unique reflections	7980 (1154)
Multiplicity	3.8 (3.8)
Mean $I/\sigma(I)$	9.9 (4.4)
Completeness (%)	94.5 (91.6)
$R_{\rm merge}$ (%)†	10.3 (28.3)
$R_{\rm meas}$ (%)‡	12.0 (32.9)
Overall B factor from Wilson plot (Å^2)	35.0

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{\rm meas} = \sum_{hkl}[N/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$.

poorly. Reducing the concentration of trisodium citrate to 1.2 *M* and the addition of 6%(*v*/*v*) glycerol resulted in better quality crystals (Fig. 3) which could be used for diffraction studies.

2.3. X-ray data collection and processing

The diffraction data were collected at 100 K using a MAR 345 detector mounted on a Bruker Microstar Ultra II Cu $K\alpha$ rotating-anode X-ray generator. A single crystal with dimensions of around 0.48 × 0.12 × 0.05 mm was used for data collection. No separate cryoprotectant other than the 6%(*v*/*v*) glycerol in the crystallization buffer was required for data collection at 100 K. Diffraction data were processed using *MOSFLM* (Leslie, 1992) and scaled using *SCALA* (Winn *et al.*, 2011). The crystal parameters, data-collection parameters and processing statistics are summarized in Table 1.

3. Results and discussion

The lectin domain with a C-terminal His tag has been purified to homogeneity (Fig. 2*a*). It has a molecular weight of about 12.6 kDa (Fig. 2*b*). The CD spectrum (not shown) indicates that the protein primarily involves β-structure. The protein has been crystallized (Fig. 3) and the crystals were characterized using X-ray diffraction. A value of 2.22 Å^3 Da^{-1} for the Matthews coefficient suggests the possibility of two lectin domains in the triclinic cell (Matthews, 1968).

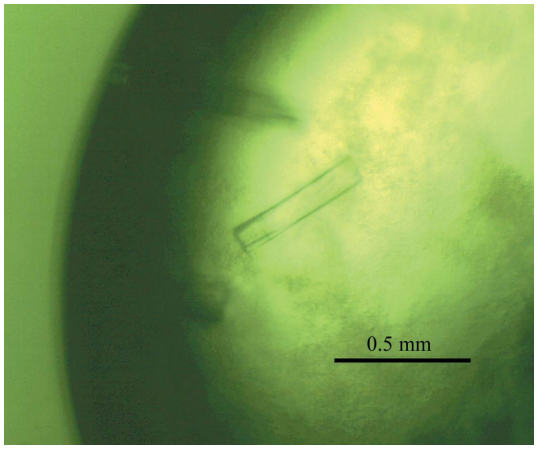


Figure 3
A crystal obtained from a solution consisting of 1.2 *M* trisodium citrate, 0.1 *M* Na HEPES pH 7.5, 6%(*v*/*v*) glycerol.

Sequence comparison indicates that the protein is similar to monocot mannose-binding β-prism II fold lectins of known structure [PDB entries 1msa (Hester *et al.*, 1995), 1kj1 (Chandra *et al.*, 1999; Ramachandraiah *et al.*, 2002), 1npl (Sauerborn *et al.*, 1999), 1b2p (Wood *et al.*, 1999), 1dlp (Wright *et al.*, 2000) and 1xd5 (Liu *et al.*, 2005)], with sequence identity ranging from 27.0 to 37.7%. The sequence identity of the *M. smegmatis* lectin domain to the lectins from *Allium sativum* (garlic; 37.7%) and *Gastrodia elata* (36.8%) is higher than 35% and these structures will be used for structure solution using molecular replacement.

The data used in the present work were collected at the X-ray Facility for Structural Biology at the Institute, supported by the Department of Science and Technology (DST). MV is a DAE Homi Bhabha Professor. DP is a CSIR Research Fellow. Financial support from the DST is acknowledged.

References

Banerjee, R., Mande, S. C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S. K., Suguna, K., Surolia, A. & Vijayan, M. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 227–231.
Becker, J. W., Reeke, G. N. Jr, Wang, J. L., Cunningham, B. A. & Edelman, G. M. (1975). *J. Biol. Chem.* **250**, 1513–1524.
Bohlool, B. B. & Schmidt, E. L. (1974). *Science*, **185**, 269–271.
Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. (2004). *Biochem. J.* **382**, 768–781.
Buist, G., Steen, A., Kok, J. & Kuipers, O. P. (2008). *Mol. Microbiol.* **68**, 838–847.
Chandra, N. R., Ramachandraiah, G., Bachhawat, K., Dam, T. K., Surolia, A. & Vijayan, M. (1999). *J. Mol. Biol.* **285**, 1157–1168.
Chandran, A. V., Prabu, J. R., Manjunath, G. P., Patil, K. N., Muniyappa, K. & Vijayan, M. (2010). *Acta Cryst.* F**66**, 1093–1095.
Chandran, T., Sharma, A. & Vijayan, M. (2010). *Acta Cryst.* F**66**, 1037–1040.
Chetnani, B., Kumar, P., Surolia, A. & Vijayan, M. (2010). *J. Mol. Biol.* **400**, 171–185.
Dam, T. K. & Brewer, C. F. (2010). *Glycobiology*, **20**, 270–279.
Feizi, T. (2000). *Immunol. Rev.* **173**, 79–88.
Hardman, K. D. & Ainsworth, C. F. (1972). *Biochemistry*, **11**, 4910–4919.
Hester, G., Kaku, H., Goldstein, I. J. & Wright, C. S. (1995). *Nature Struct. Biol.* **2**, 472–479.
Jeyaprakash, A. A., Srivastav, A., Surolia, A. & Vijayan, M. (2004). *J. Mol. Biol.* **338**, 757–770.
Kaushal, P. S., Talawar, R. K., Krishna, P. D. V., Varshney, U. & Vijayan, M. (2008). *Acta Cryst.* D**64**, 551–560.
Krishna, R., Prabu, J. R., Manjunath, G. P., Datta, S., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2007). *J. Mol. Biol.* **367**, 1130–1144.
Kulkarni, K. A., Katiyar, S., Surolia, A., Vijayan, M. & Suguna, K. (2007). *Proteins*, **68**, 762–769.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Liu, W., Yang, N., Ding, J., Huang, R., Hu, Z. & Wang, D.-C. (2005). *J. Biol. Chem.* **280**, 14865–14876.
Lobsanov, Y. D., Gitt, M. A., Leffler, H., Barondes, S. H. & Rini, J. M. (1993). *J. Biol. Chem.* **268**, 27034–27038.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Merritt, E. A., Kuhn, P., Sarfaty, S., Erbe, J. L., Holmes, R. K. & Hol, W. G. (1998). *J. Mol. Biol.* **282**, 1043–1059.
Natchiar, S. K., Suguna, K., Surolia, A. & Vijayan, M. (2007). *Crystallogr. Rev.* **13**, 3–28.
Neeser, J. R., Chambaz, A., Golliard, M., Link-Amster, H., Fryder, V. & Kolodziejczyk, E. (1989). *Infect. Immun.* **57**, 3727–3734.
Patra, D., Srikalaivani, R., Misra, A., Singh, D. D., Selvaraj, M. & Vijayan, M. (2010). *Acta Cryst.* F**66**, 1662–1665.
Prabu, J. R., Thamotharan, S., Khanduja, J. S., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2009). *Biochim. Biophys. Acta*, **1794**, 1001–1009.

Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I. & Sussman, J. L. (2005). *Bioinformatics*, **21**, 3435–3438.
Ramachandraiah, G., Chandra, N. R., Surolia, A. & Vijayan, M. (2002). *Acta Cryst.* D**58**, 414–420.
Ramachandraiah, G., Chandra, N. R., Surolia, A. & Vijayan, M. (2003). *Glycobiology*, **13**, 765–775.
Roy, S., Saraswathi, R., Chatterji, D. & Vijayan, M. (2008). *J. Mol. Biol.* **375**, 948–959.
Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A. & Vijayan, M. (1996). *Nature Struct. Biol.* **3**, 596–603.
Sauerborn, M. K., Wright, L. M., Reynolds, C. D., Grossmann, J. G. & Rizkallah, P. J. (1999). *J. Mol. Biol.* **290**, 185–199.
Selvaraj, M., Roy, S., Singh, N. S., Sangeetha, R., Varshney, U. & Vijayan, M. (2007). *J. Mol. Biol.* **372**, 186–193.
Sharma, A., Chandran, D., Singh, D. D. & Vijayan, M. (2007). *J. Biosci.* **32**, 1089–1110.
Sharma, A., Sekar, K. & Vijayan, M. (2009). *Proteins*, **77**, 760–777.
Sharma, A. & Vijayan, M. (2011). *Glycobiology*, **21**, 23–33.
Sharon, N. & Lis, H. (2004). *Glycobiology*, **14**, R53–R62.
Singh, D. D., Chandran, D., Jeyakani, J. & Chandra, N. (2007). *Protein Pept. Lett.* **14**, 683–691.
Singh, D. D., Saikrishnan, K., Kumar, P., Surolia, A., Sekar, K. & Vijayan, M. (2005). *Glycobiology*, **15**, 1025–1032.
Skehel, J. J. & Wiley, D. C. (2000). *Annu. Rev. Biochem.* **69**, 531–569.
Taylor, M. E. & Drickamer, K. (2007). *Curr. Opin. Struct. Biol.* **19**, 572–577.
Vijayan, M. (2005). *Tuberculosis*, **85**, 357–366.
Vijayan, M. & Chandra, N. (1999). *Curr. Opin. Struct. Biol.* **9**, 707–714.
Wiley, D. C., Wilson, I. A. & Skehel, J. J. (1981). *Nature (London)*, **289**, 373–378.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Wood, S. D., Wright, L. M., Reynolds, C. D., Rizkallah, P. J., Allen, A. K., Peumans, W. J. & Van Damme, E. J. M. (1999). *Acta Cryst.* D**55**, 1264–1272.
Wright, L. M., Reynolds, C. D., Rizkallah, P. J., Allen, A. K., Van Damme, E. J., Donovan, M. J. & Peumans, W. J. (2000). *FEBS Lett.* **468**, 19–22.
Zhang, R.-G., Westbrook, M. L., Westbrook, E. M., Scott, D. L., Otwinowski, Z., Maulik, P. R., Reed, R. A. & Shipley, G. G. (1995). *J. Mol. Biol.* **251**, 550–562.