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

Expression, purification, crystallization and preliminary X-ray analysis of the N-terminal domain of GNBP3 from *Drosophila melanogaster*

Yumiko Mishima, Franck Coste, Vanessa Bobezeau, Nadège Hervouet, Christine Kellenberger and Alain Roussel*

Centre de Biophysique Moléculaire, UPR 4301 CNRS, Orléans, France

Correspondence e-mail: roussel@cnrs-orleans.fr

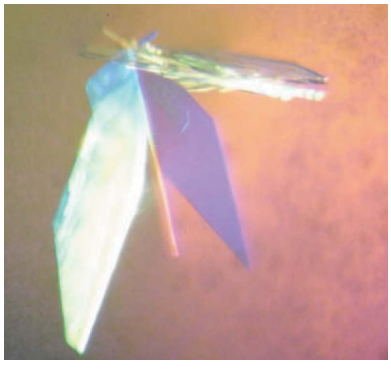
Received 3 March 2009
Accepted 22 April 2009

Gram-negative bacteria-binding protein 3 (GNBP3) is a pattern-recognition receptor which contributes to the defensive response against fungal infection in *Drosophila*. The protein consists of an N-terminal domain, which is considered to recognize β-glucans from the fungal cell wall, and a C-terminal domain, which is homologous to bacterial glucanases but devoid of activity. The N-terminal domain of GNBP3 (GNBP3-Nter) was successfully purified after expression in *Drosophila* S2 cells. Diffraction-quality crystals were produced by the hanging-drop vapour-diffusion method using PEG 2000 and PEG 8000 as precipitants. Preliminary X-ray diffraction analysis revealed that the GNBP3-Nter crystals belonged to the monoclinic space group $C2$, with unit-cell parameters $a = 134.79$, $b = 30.55$, $c = 51.73$ Å, $\beta = 107.4^\circ$, and diffracted to 1.7 Å using synchrotron radiation. The asymmetric unit is expected to contain two copies of GNBP3-Nter. Heavy-atom derivative data were collected and a samarium derivative showed one high-occupancy site per molecule.

1. Introduction

A critical step in the immune response is the identification of an invading organism as foreign. Microbial invasion is detected by the recognition of a limited but highly conserved set of microbe-associated molecular patterns (MAMPs), such as lipopolysaccharides (LPS), peptidoglycans and β-1,3 glucans, which are absent in the host but present on the surface of microbes (Medzhitov & Janeway, 2002; Janeway, 1989). This crucial task is fulfilled by specialized proteins called pattern-recognition receptors (PRRs), which interact with the MAMPs. A multitude of innate immune responses are then evoked. They are either of cellular type, such as phagocytosis, encapsulation, clotting and melanization, or of humoral type, such as the secretion of antimicrobial peptides. In *Drosophila*, the transcription of the genes encoding these peptides is under the control of the Toll and Imd signalling pathways (Hoffmann, 2003).

The PRRs for these pathways are peptidoglycan-recognition proteins (PGRPs) and Gram-negative binding proteins (GNBPs). In contrast to *Drosophila* PGRPs, for which several crystal structures have been reported (Chang *et al.*, 2004; Leone *et al.*, 2008), little is known regarding GNBPs at the molecular and structural level. The name GNBP comes from the discovery of a protein with strong affinity to the cell wall of Gram-negative bacteria in the haemolymph of the silkworm *Bombyx mori* (Lee *et al.*, 1996). This name is misleading, as members of this family display several other functions. The *Drosophila* genome contains three full-length GNBP genes that encode proteins of about 430 residues. To date, no functional role has been attributed to GNBP2. GNBP1 (Gobert *et al.*, 2003) contributes, in association with PGRP-SA and PGRP-SD, to the recognition of Gram-positive bacteria, while GNBP3 is responsible for fungal infection sensing upstream of the Toll pathway (Gottar *et al.*, 2006). Gottar and coworkers showed that the GNBP3^{hades} mutant is sensitive to *Candida albicans* infection and showed by pull-down experi-



ments that recombinant GNBP3 binds to both curdlan and killed *C. albicans*. Thus, GNBP3 can be classified into the β-glucan recognition protein (βGRP) family.

Ochiai & Ashida (1988) reported the first βGRP from the haemolymph of *B. mori*. βGRPs and GNBPs are extracellular proteins composed of a small N-terminal domain of about 100 residues and a longer C-terminal domain which displays homology to bacterial glucanases but lacks the two active glutamic acid residues and therefore the glucanase activity. In 2000, the same authors showed that a recombinant protein consisting of the 102 N-terminal residues of the silkworm βGRP binds to β-1,3 glucan but could not trigger an immune response (Ochiai & Ashida, 2000). Since then, several βGRPs have been isolated and characterized from other insects such as *Manduca sexta* (Ma & Kanost, 2000; Jiang *et al.*, 2004) and *Plodia interpunctella* (Fabrick *et al.*, 2004). A study on the pyralid moth *P. interpunctella* confirmed the importance of the N-terminal part of the βGRP, which binds to β-1,3 glucan and activates the prophenoloxidase cascade (Fabrick *et al.*, 2004). However, the detailed molecular mechanism of carbohydrate recognition and immune-response triggering by the βGRPs remains completely unknown because of the lack of any structural information.

Here, we present the expression, purification, crystallization and preliminary X-ray analysis of the N-terminal domain of *Drosophila* GNBP3.

2. Experimental procedures

2.1. Cloning and expression in *Drosophila* S2 cells

The cDNA encoding the N-terminal domain of GNBP3 (residues 1–128, UniProtKB/Swiss-Prot Q9NHA8) from *D. melanogaster* was cloned in the pMT-V5-HisA vector (Invitrogen) by introducing *Eco*RI and *Xho*I sites using the primers 5′-GGGAAGAATTCGGCATGGCGGATGCATTGCGCTTTGT-3′ (forward) and 5′-TTTTTCTCGAGATTATTACCACTGTACCCATTGACCAC-3′ (reverse). The resulting vector encodes a secreted protein with V5-epitope and hexahistidine tags (Fig. 1*a*). The recombinant plasmid encoding GNBP3-Nter was then co-transfected with the pAc5C-pac vector (an actin5C-driven expression vector for puromycin acetyltransferase; Dimarcq *et al.*, 1997) into *Drosophila* S2 cells according to the protocol from Invitrogen. Stable clones were obtained using puromycin selection. Cells were grown in suspension at 296 K at a cell density of 3–4 $\times$ 10^6 cells ml^{-1} and kept under selection in Schneider's medium (Sigma) containing 0.5 µg ml^{-1} puromycin (Invivogen), 50 µg ml^{-1} streptomycin (Gibco), 50 U ml^{-1} penicillin (Gibco), 2 m*M* Glutamax (Gibco) and 10% foetal bovine serum (Gibco) previously heat-inactivated at 333 K for 30 min. Expression of the secreted protein was induced by the addition of 0.5 m*M* $CuSO_4$. After 5 d, cells were aseptically centrifuged, resuspended in fresh medium and induced again for 5 d. Up to ten inductions could be performed using the same cells.

2.2. Protein purification and tag removal

The collected medium (1 l) was centrifuged and filtered for clarification and the protein was purified by affinity chromatography using Chelating Sepharose Fast Flow (Amersham Biosciences) resin. After loading the protein, the resin was washed with a solution containing 10 m*M* imidazole along with 20 m*M* phosphate buffer pH 7.4, 50 m*M* NaCl. Purification was achieved using stepwise elution with solutions containing an increasing concentration of imidazole (30, 50 and 250 m*M*). The recombinant protein was found in the fraction containing 250 m*M* imidazole (Fig. 1*b*). The V5-His tag was cleaved by overnight incubation with trypsin (in a 1:100 ratio), which was then removed using a HiTrap Benzamidine FF column (1 ml; GE Healthcare). Cleavage occurred after the first arginine residue of the V5-His tag. Four residues of the tag remained after cleavage (Fig. 1*a*). The protein was further purified by size-exclusion chromatography on a Superdex-75 column (HiLoad 26/60 Superdex 75 pg, Amersham Biosciences) in 20 m*M* HEPES pH 7.4, 150 m*M* NaCl. The overall yield was estimated to be about 5 mg per litre of culture. The purity of the protein was assessed by SDS–PAGE and it was concentrated to 10 mg ml^{-1}.

The identity of GNBP3-Nter was confirmed by matrix-assisted laser-desorption ionization–time of flight (MALDI-TOF) mass spectrometry, giving an experimental molecular weight of 12 269.4 Da. The N-terminal sequence (Tyr-Glu-Val-Pro) was determined by Edman degradation. The sequence resulting from these data is shown in Fig. 1.

2.3. Crystallization

Crystallization trials were carried out by the hanging-drop vapour-diffusion method at room temperature using Crystal Screens 1 and 2 (Hampton Research) and JCSG+ (Molecular Dimensions Ltd). Drops were prepared by mixing equal volumes (1 µl) of protein solution (5 or 10 mg ml^{-1} protein, 20 m*M* HEPES pH 7.4, 150 m*M* NaCl) and precipitant solution and were equilibrated against 0.3 ml reservoir volume. Crystallization hits with the cleaved form occurred in the presence of polyethylene glycols (PEGs) as a precipitant, especially in the conditions (i) Crystal Screen 1 condition No. 45 (0.2 *M* zinc acetate dehydrate, 18% PEG 8000, 0.1 *M* sodium cacodylate trihydrate pH 6.5) and (ii) JCSG+ condition No. 82 (0.15 *M* potassium bromide, 30% PEG 2000 MME). The resulting optimized conditions were (i) 0.2 *M* zinc acetate dehydrate, 18% PEG 8000 and sodium acetate pH 4.6 at a protein concentration of 4 mg ml^{-1} (form I) and (ii) 40% PEG 200 MME, 0.1 *M* sodium acetate pH 4.6 at a

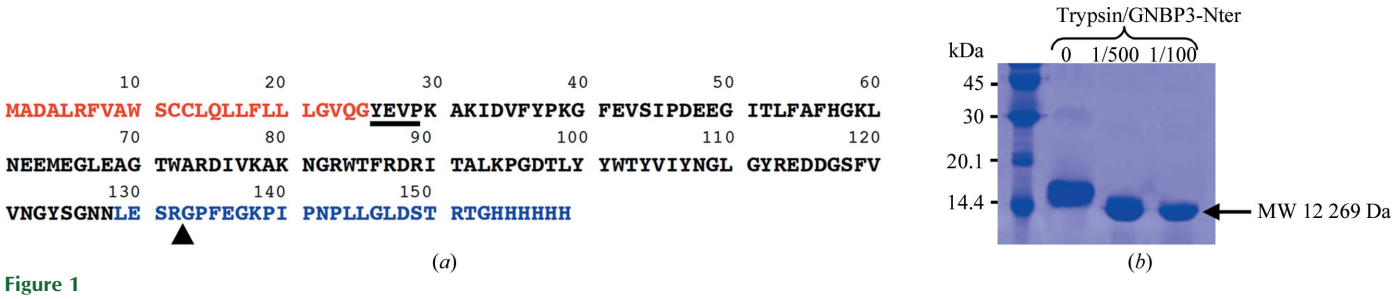


Figure 1
(*a*) Amino-acid sequence of the produced fusion protein. The signal peptide residues are shown in red. The C-terminal fusion residues are shown in blue. The N-terminal residues of the secreted protein sequenced by Edman degradation are underlined. A black triangle shows the trypsin cleavage site. (*b*) SDS–PAGE analysis of the purified protein before (lane 0) and after (lanes 1/500 and 1/100) trypsinolysis.

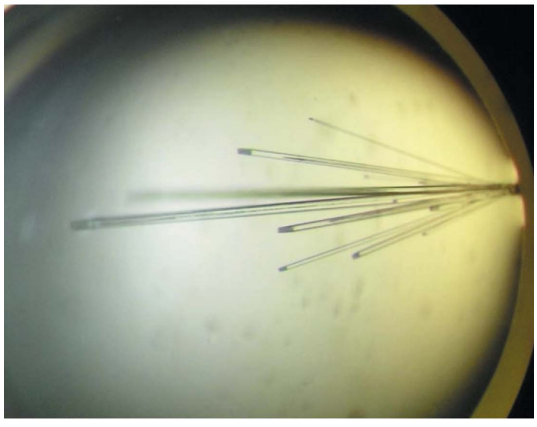
protein concentration of 8 mg ml^{-1} (form II). The crystals obtained under these conditions had different morphologies (Fig. 2).

A heavy-atom derivative search was achieved by soaking crystals in a solution containing 1–10 m*M* of various heavy atoms and the hits were optimized by cocrystallization. Samarium derivatives were obtained by adding 2 m*M* $SmCl_3.6H_2O$ to the optimized crystallization condition of form I before mixing with the protein solution.

2.4. Data collection and X-ray crystallographic analysis

Prior to data collection, the crystals of native GNBP3-Nter were soaked in cryoprotectant solution (form I, 0.2 *M* zinc acetate dehydrate, 18% PEG 8000, sodium acetate pH 4.6 and 20% ethylene glycol; form II, 40% PEG 200 MME, 40 m*M* sodium acetate pH 4.6 and 18% ethylene glycol) and flash-frozen in liquid nitrogen. Diffraction experiments were conducted at 100 K using an X-ray wavelength of 0.934 Å on beamline ID14-1 of the European Synchrotron Radiation Facility (ESRF, Grenoble, France). Data were processed with *MOSFLM* (Leslie, 2006) and scaled with programs from the *CCP*4 suite v.6.0.2 (Evans, 2006). Crystal parameters and diffraction statistics are shown in Table 1. Both forms of the crystals belonged to space group *C*2 and have very similar unit-cell parameters.

(*a*)



(*b*)

Figure 2
Crystals of N-terminus domain of GNBP3. (*a*) Crystal form I and (*b*) crystal form II.

Table 1
Data-collection statistics.

Values in parentheses are for the highest resolution shell.

	Native (form I)	Native (form II)	Sm derivative (form I)
Wavelength (Å)	0.934	0.934	1.5418
Space group	*C*2	*C*2	*C*2
Unit-cell parameters			
a (Å)	134.79	135.03	134.53
b (Å)	30.55	30.56	30.62
c (Å)	51.73	51.74	51.43
β (°)	107.4	107.3	107.3
Resolution range (Å)	26.28–1.70 (1.79–1.70)	34.54–1.69 (1.79–1.69)	26.28–2.20 (2.30–2.20)
Measured reflections	74900 (9680)	83501 (11779)	144614 (20538)
Unique reflections	21491 (2992)	22450 (3194)	10472 (1502)
Completeness (%)	95.6 (92.3)	98.2 (96.7)	99.9 (100.0)
Multiplicity	3.5 (3.2)	3.7 (3.7)	13.8 (13.7)
$R_{\rm merge}$† (%)	7.1 (35.6)	6.0 (33.5)	3.7 (10.3)
Mean $I/\sigma(I)$	8.1 (2.1)	9.9 (2.3)	16.6 (7.0)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean intensity of the symmetry-equivalent reflections.

3. Results and discussion

GNBP3-Nter was cloned in pMT-V5-His vector for expression in *Drosophila* S2 cells. Owing to the presence of a signal peptide, the protein was secreted into the medium. The sequence of the mature protein was determined by Edman sequencing to be ^{26}YEVP. The expression yield was estimated to be more than 15 mg per litre of culture. Despite two-step purification by affinity and size-exclusion chromatography, crystallization trials were unsuccessful. The V5-His tag along with the linker represents 32 residues and accounts for nearly a quarter of the secreted protein. Crystallization trials gave positive results after removal of the tag by limited proteolysis. Two conditions were found and optimized, leading to crystals with different morphologies (Fig. 2). Diffraction data were collected to high resolution from both crystal forms.

Although no structure has yet been determined for a member of the GNBP family, an effort was made to utilize the molecular-replacement method using partial models. A *BLAST* search gives several proteins as hits; for example, the Pdz domain of human RIM2B (PDB code 1wfg; 54.8% similarity over 71 residues) and

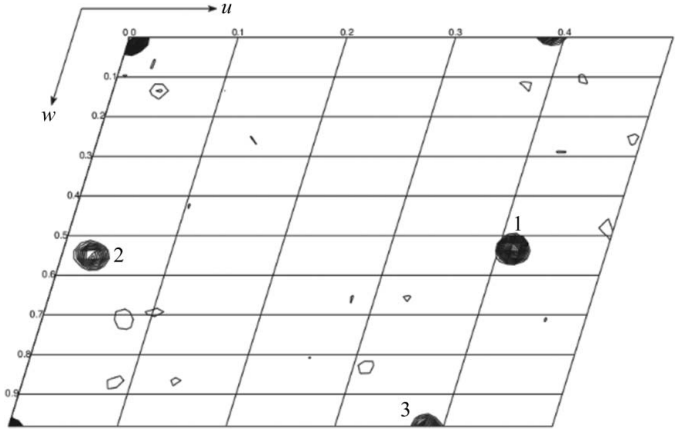


Figure 3
Harker section ($0 < u < 0.5$, $v = 0.5$, $0 < w < 1$) of the anomalous difference Patterson map of the samarium soak of GNBP3-Nter calculated at 3.0 Å resolution using *PATTERSON* from the *CCP*4 suite. The map is drawn with a minimum contour level of 2.0σ with a 1.0σ increment. The three main peaks are labelled. This Patterson map is consistent with three sites of substitution at approximately ($u = 0.41$, $w = 0.53$) for peak 1, ($u = 0.03$, $w = 0.55$) for peak 2 and ($u = 0.38$, $w = 0.98$) for peak 3.

muconate-lactonizing enzyme (PDB code 1bkh; 54.1% similarity over 66 residues). Molecular replacement using these protein fragments as the search model did not give any clear solutions. Therefore, it was necessary to solve the phase problem by means of multiple isomorphous replacement or multiple-wavelength anomalous dispersion methods. About 40 crystals soaked in various heavy-atom derivatives were tested on beamline ID14-1 at the ESRF. Of the nine partial data sets collected, it appeared that samarium gave satisfying results. Cocrystallization experiments were conducted to optimize the signal. A SAD data set was collected to 2.2 Å resolution in-house using Cu $K\alpha$ radiation. Diffraction statistics are shown in Table 1. Samarium sites were identified (Fig. 3), refined and used for phase calculation with the *PHENIX* suite (Adams *et al.*, 2002). Further model building and refinement against the high-resolution native data set of form II are currently in progress.

We would like to thank the mass-spectrometry service of the laboratory and the European Synchrotron Radiation Facility (ESRF) at Grenoble and in particular the beamline ID23-1 staff for their assistance. This work was supported by an ATIP program of the Centre National de la Recherche Scientifique (2005–2008 to AR) from which YM was a recipient of a postdoctoral fellowship, and by a grant from the Agence Nationale de la Recherche (ANR-MIME to AR).

References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.
Chang, C. I., Pili-Floury, S. S., Herve, M., Parquet, C., Chelliah, Y., Lemaitre, B., Mengin-Lecreulx, D. & Deisenhofer, J. (2004). *PLoS Biol.* **2**, e277.
Dimarcq, J. L., Imler, J. L., Lanot, R., Ezekowitz, R. A., Hoffmann, J. A., Janeway, C. A. & Lagueux, M. (1997). *Insect Biochem. Mol. Biol.* **27**, 877–886.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Fabrick, J. A., Baker, J. E. & Kanost, M. R. (2004). *J. Biol. Chem.* **279**, 26605–26611.
Gobert, V., Gottar, M., Matskevich, A. A., Rutschmann, S., Royet, J., Belvin, M., Hoffmann, J. A. & Ferrandon, D. (2003). *Science*, **302**, 2126–2130.
Gottar, M., Gobert, V., Matskevich, A. A., Reichhart, J. M., Wang, C., Butt, T. M., Belvin, M., Hoffmann, J. A. & Ferrandon, D. (2006). *Cell*, **127**, 1425–1437.
Hoffmann, J. A. (2003). *Nature (London)*, **426**, 33–38.
Janeway, C. A. (1989). *Cold Spring Harbor Symp. Quant. Biol.* **54**, 1–13.
Jiang, H., Ma, C., Lu, Z. Q. & Kanost, M. R. (2004). *Insect Biochem. Mol. Biol.* **34**, 89–100.
Lee, W. J., Lee, J. D., Kravchenko, V. V., Ulevitch, R. J. & Brey, P. T. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 7888–7893.
Leone, P., Bischoff, V., Kellenberger, C., Hetru, C., Royet, J. & Roussel, A. (2008). *Mol. Immunol.* **45**, 2521–2530.
Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.
Ma, C. & Kanost, M. R. (2000). *J. Biol. Chem.* **275**, 7505–7514.
Medzhitov, R. & Janeway, C. A. Jr (2002). *Science*, **296**, 298–300.
Ochiai, M. & Ashida, M. (1988). *J. Biol. Chem.* **263**, 12056–12062.
Ochiai, M. & Ashida, M. (2000). *J. Biol. Chem.* **275**, 4995–5002.