Jaslyn E. M. M. Wong and
Mickael Blaise*

Centre for Carbohydrate Recognition and
Signalling, Department of Molecular Biology
and Genetics, Aarhus University, Gustav Wieds
Vej 10c, 8000 Aarhus, Denmark

Correspondence e-mail: mick@mb.au.dk

Received 22 August 2013

Accepted 2 October 2013

Cloning, expression, purification, crystallization and preliminary crystallographic analysis of the putative NlpC/P60 endopeptidase, TTHA0266, from *Thermus thermophilus* HB8

Autolysins belong to a protein family involved in peptidoglycan degradation and remodelling. Within this family, NlpC/P60 endopeptidases are involved in the hydrolysis of the peptide arm of peptidoglycan. In this work, the putative NlpC/P60 endopeptidase TTHA0266 from *Thermus thermophilus* HB8 was over-expressed, purified and crystallized. The crystals diffracted to 2.4 Å resolution and belonged to the hexagonal space group $P6_1$, with unit-cell parameters $a = b = 71.19$, $c = 198.68$ Å, $\gamma = 120^\circ$. Selenomethionine-substituted protein was crystallized and the structure was solved by single-wavelength anomalous dispersion.

1. Introduction

Bacterial NlpC/P60 endopeptidases constitute a large protein family involved in peptidoglycan (PGN) hydrolysis (Anantharaman & Aravind, 2003). Most NlpC/P60 endopeptidases cleave the peptide part of PGN between the second and third amino acids. Since a peptide bond between a D-amino acid and an L-amino acid is hydrolyzed, they are named D,L-endopeptidases. Their role is dedicated to PGN hydrolysis during bacterial cell division; more precisely, in cell elongation and daughter-cell separation (Vollmer *et al.*, 2008).

Several bacterial NlpC/P60 endopeptidase structures have been solved by NMR or X-ray crystallography (Aramini *et al.*, 2008; Xu *et al.*, 2009, 2010, 2011; Ruggiero *et al.*, 2010; Böth *et al.*, 2011). These studies showed that the NlpC/P60 catalytic domain consists of an antiparallel β -sheet surrounded by three to six α -helices. The highly conserved catalytic site is composed of a catalytic Cys/His/polar residue triad, in which the polar residue is very often a Glu or His.

The modular organization of the NlpC/P60 is, however, quite diverse. Indeed, in bacteria the domain is often associated with PGN targeting domains: SH3b, CBD (choline-binding domain) or LysM (lysine motif) (Anantharaman & Aravind, 2003; Xu *et al.*, 2010). Hitherto, only crystal structures of the NlpC/P60 domain alone or associated with the SH3b domain have been solved (Xu *et al.*, 2010). Since no structure has been solved with a PGN fragment containing the carbohydrate and peptide components, it is not yet known whether the SH3b, CBD or LysM domains act as a pedestal to anchor the enzyme to the cell wall and/or if they contribute to the precise positioning of the NlpC/P60 catalytic domain for catalysis.

Therefore, determining NlpC/P60 structures with these domains associated with their cognate ligands, *i.e.* PGN fragments, would increase our understanding of the role of these extra modules. Here, we examine the protein encoded by the *ttha0266* gene from *Thermus thermophilus* HB8 that possesses two LysM domains in the N-terminus and a putative NlpC/P60 domain at the C-terminus (Fig. 1). We report the cloning, expression, crystallization and preliminary crystallographic analysis of the *T. thermophilus* TTHA0266 protein.

2. Materials and methods

2.1. Cloning, expression and purification

The *ttha0266* gene was amplified by PCR with the following primers (forward, GACGACGACAAGATGGAGAATCTGTACT-

TCCAGGGACAGGCCACCTACACCGTGGCCCCGGGGACA-CGCTCTAC; reverse, GAGGAGAAGCCCGGTTATGGCGT-TGCGGAAGGCGCGGGGGGAGGTTCC). The part coding for amino acids 1–15, predicted by the *SignalP4.1* server (Petersen *et al.*, 2011) to be a putative peptide signal, was removed. The gene was inserted, in frame with thioredoxin, His and S tags, into the pET-32 Ek/LIC vector. The forward primer also contains the sequence encoding the *Tobacco etch virus* (TEV) protease recognition site, thus allowing the removal of the tags during protein purification. The protein was expressed in *Escherichia coli* BL21 Rosetta2 cells. Cells were grown in six flasks containing 2 l of Luria Broth medium supplemented with kanamycin and chloramphenicol at 310 K until the optical density at 600 nm reached 0.8. The culture was supplemented with 1 mM isopropyl β -D-1-thiogalactopyranoside to induce protein expression and was incubated overnight at 293 K. Cells were harvested by centrifugation at 12 000g for 10 min and the pellet was then resuspended in 200 ml buffer A consisting of 50 mM Tris–HCl pH 8, 0.2 M NaCl, 20 mM imidazole, 5 mM β -mercaptoethanol (β -Me), 1 mM benzamidine. Cells were lysed by sonication and the cell debris was removed by a centrifugation step at 30 000g for 45 min. The supernatant was then incubated for 1 h with 10 ml Ni-NTA beads (Qiagen) at 277 K and the beads were then washed with a buffer composed of 50 mM Tris pH 8, 1 M NaCl, 50 mM imidazole, 5 mM β -Me, 1 mM benzamidine and the protein was eluted with 50 mM Tris pH 8, 0.2 M NaCl, 500 mM imidazole, 5 mM β -Me, 1 mM benzamidine. The protein tags were cleaved overnight at 277 K with TEV protease, which was purified as described previously (Alsarraf *et al.*, 2011) and added in a 1:50(w:w) ratio, and dialyzed against buffer A. After a centrifugation step at 3900g for 20 min, the protein was run over Ni-NTA beads pre-equilibrated with buffer A. The protein in the flowthrough which was free from tags was collected and concentrated to 5 mg ml^{−1} before being loaded onto a Superdex 75

Figure 1 Diagram of (a) the TTHA0266 protein and (b) the putative activity. (a) N-t and C-t indicate the N-terminus and C-terminus, respectively, and SP indicates signal peptide; the amino-acid numbering is indicated below each domain. (b) Simplified view of a peptidoglycan fragment from *T. thermophilus* based on the study of Quintela *et al.* (1995). The dashed line indicates the putative hydrolytic activity of the NlpC/P60 domain between the amino acids D-glutamine and L-ornithine.

Table 1 Data-collection statistics of the native and selenomethionine-substituted crystals. Values in parentheses are for the last resolution shell.

Data collection	Native	Se peak
Beamline	I911-3, MAX-lab	I911-3, MAX-lab
Wavelength (Å)	1.0	0.979
Space group	<i>P</i> 6 ₁	<i>P</i> 6 ₁
Unit-cell parameters (Å)	<i>a</i> = <i>b</i> = 71.19, <i>c</i> = 198.68	<i>a</i> = <i>b</i> = 71.6, <i>c</i> = 197.79
Resolution (Å)	30–2.4 (2.46–2.40)	30–2.6 (2.65–2.60)
<i>R</i> _{meas} † (%)	6.9 (86.6)	6.9 (75.7)
Mosaicity (°)	0.12	0.24
<i>I</i> / <i>σ</i> (<i>I</i>)	20.6 (2.88)	15.3 (2.11)
Completeness (%)	99.9 (99.9)	99.8 (99.8)
Multiplicity	9.7 (9.6)	5.9 (5.9)

† *R*_{meas} is defined according to Diederichs & Karplus (1997).

(GE Healthcare) size-exclusion column and eluted in buffer consisting of 50 mM Tris–HCl pH 8, 0.2 M NaCl, 5 mM β -Me. The selenomethionine-substituted protein was prepared following a procedure described previously (Alsarraf *et al.*, 2011) and purified as for the non-labelled protein.

2.2. Crystallization experiments

The protein was concentrated to 25 mg ml^{−1} and crystallization screens were performed in 96-well Swissci MRC crystallization plates. A Mosquito robot (TTP LabTech) was used to set up sitting drops consisting of 200 nl protein solution and 200 nl reservoir solution. The plates were incubated at 292 K. Seven commercial screens were tested: the Index, PEGRx and SALTRx screens from Hampton Research and the Structure Screen, MIDAS, JCSG-*plus*, MacroSol and FootPrint screens from Molecular Dimensions.

Native crystals were optimized by setting up hanging drops by mixing 1 μ l protein solution at 24 mg ml^{−1} and 1 μ l reservoir solution consisting of 0.2 M ammonium sulfate, 16% PEG 8000. Crystals were transferred into a solution consisting of 0.2 M ammonium sulfate, 16% PEG 8000, 20% ethylene glycol for a few seconds prior to cryocooling in liquid nitrogen.

Selenomethionine-derivative crystals were optimized by setting up hanging drops by mixing 1 μ l protein solution at 24 mg ml^{−1} and 1 μ l reservoir solution consisting of 0.1 M sodium citrate pH 5.5, 16% PEG 4000, 15% 2-propanol at 292 K. Selenomethionine-derivative crystals were soaked for a few seconds in a solution consisting of 0.1 M sodium citrate pH 5.5, 16% PEG 4000, 20% ethylene glycol prior to cryocooling in liquid nitrogen.

2.3. Data collection and processing

Native and single-wavelength anomalous dispersion (SAD) data sets were collected on the I911-3 beamline at MAX-lab, Lund, Sweden. The native data set consisting of 150 frames was collected at a wavelength of 1 Å with a crystal-to-detector distance of 232 mm and an oscillation range of 1°. After performing a fluorescence scan, a SAD data set of 200 frames was collected at 0.979 Å with a crystal-to-detector distance of 263 mm and an oscillation range of 1°. Both data sets were processed with *XDS* (Kabsch, 2010).

3. Results and discussion

The TTHA0266 protein from *T. thermophilus* was successfully expressed and purified. About 4 mg highly pure protein was obtained per litre of culture (Fig. 2a). Over the seven crystallization screens, about 100 crystallization conditions were obtained. In many

Figure 2

(a) SDS-PAGE of TTHA0266 protein after size-exclusion chromatography. Molecular-mass markers indicated on the left (in kDa). (b) The TTHA0266 protein crystals. The picture was taken under polarized light. (c) Diffraction pattern of the native TTHA0266 protein crystals. Diffraction spots extending to 2.4 Å are indicated in the enlarged box.

conditions, crystals appeared within a few seconds or minutes after mixing the protein and reservoir solutions. These crystals were reproduced in 24-well plates. After optimization, the bipyramidal-shaped crystals had a size of about 150 µm on the longest edge (Fig. 2b).

An initial native data set was collected to 2.4 Å resolution (Fig. 2c). The crystal belonged to space group $P6_1$ with unit-cell parameters $a = b = 71.19$, $c = 198.68$ Å (Table 1). The Matthews coefficient (V_M) of the crystal was $2.75 \text{ Å}^3 \text{ Da}^{-1}$, with 55.34% of the unit cell occupied by solvent, assuming two molecules (molecular mass 26.4 kDa) per asymmetric unit (Matthews, 1968). Molecular replacement failed to provide phase information despite exhaustive testing of NlpC/P60 domains and/or LysM domains available in the Protein Data Bank as search models. The closest models tried were the structure of the NlpC/P60, lipoprotein Spr from *E. coli* (Aramini *et al.*, 2008), which presented 37% identity over 125 residues, and the LysM domain of the CVNH-LysM lectin from *Magnaporthe oryzae*, which presented 44% identity over 48 residues (Koharudin *et al.*, 2011). Polyalanine and homology models were also tested without success. We therefore purified and crystallized a selenomethionine-substituted protein. A SAD data set was collected to 2.6 Å resolution (Table 1) with

Figure 3

Experimental electron-density map after density modification. The 2.6 Å resolution map is contoured at the 1σ level. The positions of SeMet42 and Arg43 located in the N-terminal LysM domain are indicated. The figure was prepared with *PyMOL* (Schrödinger).

Figure 4

Crystal packing and solvent channel. (a) The density-modified map contoured at 1σ and the model are centred on the solvent channel. The two monomers are shown in magenta and blue, while the symmetry-related molecules are displayed in orange. The red dashed circle indicates the putative position of the nonvisible LysM_2 domain in the solvent channel. (b) Crystal packing of the well defined LysM_2 domain from the second monomer.

anomalous signal extending to 3.6 Å according to *phenix.xtriage* from the *PHENIX* package (Adams *et al.*, 2010). The *HySS* and *SOLVE* programs from the *PHENIX Autosol* wizard (Terwilliger *et al.*, 2009) were used to solve the structure. Four of the six selenium sites were found with a mean figure of merit of 0.49 using a resolution cutoff of 3.6 Å. After density modification, clear electron density was visible for the two catalytic domains and the two N-terminal LysM domains, *i.e.* the two LysM_1 (Fig. 1a), allowing initial automatic model building (Fig. 3). After automatic building, most of the residues of the two NlpC/P60 catalytic domains and of two of the four LysM domains were traced. A third LysM domain (LysM_2) was further found by molecular replacement using *Phaser* (McCoy *et al.*, 2007) from the

PHENIX package. The second LysM_2 could not be located by molecular replacement and no electron density was visible for it, even though numerous density-modification trials were performed.

We think that the fourth LysM domain cannot be located because it might adopt several distinct conformations owing to the existence of flexibility induced by the linker between the LysM_1 and LysM_2 domains and by a long nonstructured linker between the LysM_2 domain and the catalytic domain (Fig. 1a). Consequently, the second LysM_2 domain positioned in a large solvent channel might be quite flexible, thus preventing a defined electron density for it (Fig. 4a). The other LysM_2 domain is well defined because it is packed against symmetry-related molecules (Fig. 4b). Additionally, the long linker ranging from residues 109 to 130 was only partly visible owing to missing electron density for residues 10–118.

The model was then manually rebuilt with *Coot* (Emsley *et al.*, 2010) and refined with the *PHENIX* package. The model was finally used to perform molecular replacement with the 2.4 Å resolution native data set; the structure was rebuilt and refined but the fourth LysM domain was however still not visible. The structure of the TTHA0266 protein together with its biochemical characterization will be published elsewhere (Wong *et al.*, in preparation).

This work was supported by the Danish National Research Foundation grant No. DNR79, Danscatt and BiostructX. We thank S. Thirup and J. Stougaard for constant support and critical reading of the manuscript, and also the staff at MAX-lab for help during data collection.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Alsarraf, H. M. A. B., Laroche, F., Spaink, H., Thirup, S. & Blaise, M. (2011). *Acta Cryst.* **F67**, 1253–1256.
- Anantharaman, V. & Aravind, L. (2003). *Genome Biol.* **4**, R11.
- Aramini, J. M., Rossi, P., Huang, Y. J., Zhao, L., Jiang, M., Maglaqui, M., Xiao, R., Locke, J., Nair, R., Rost, B., Acton, T. B., Inouye, M. & Montelione, G. T. (2008). *Biochemistry*, **47**, 9715–9717.
- Böth, D., Schneider, G. & Schnell, R. (2011). *J. Mol. Biol.* **413**, 247–260.
- Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 133–144.
- Koharudin, L. M., Viscomi, A. R., Montanini, B., Kershaw, M. J., Talbot, N. J., Ottonello, S. & Gronenborn, A. M. (2011). *Structure*, **19**, 662–674.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. (2011). *Nature Methods*, **8**, 785–786.
- Quintela, J. C., Pittenauer, E., Allmaier, G., Arán, V. & de Pedro, M. A. (1995). *J. Bacteriol.* **177**, 4947–4962.
- Ruggiero, A., Marasco, D., Squeglia, F., Soldini, S., Pedone, E., Pedone, C. & Berisio, R. (2010). *Structure*, **18**, 1184–1190.
- Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H. & Hung, L.-W. (2009). *Acta Cryst.* **D65**, 582–601.
- Vollmer, W., Doris, B., Charlier, P. & Foster, S. (2008). *FEMS Microbiol. Rev.* **32**, 259–286.
- Xu, Q. *et al.* (2009). *Structure*, **17**, 303–313.
- Xu, Q. *et al.* (2010). *Acta Cryst.* **F66**, 1354–1364.
- Xu, Q., Rawlings, N. D., Chiu, H.-J., Jaroszewski, L., Klock, H. E., Knuth, M. W., Miller, M. D., Elsliger, M.-A., Deacon, A. M., Godzik, A., Lesley, S. A. & Wilson, I. A. (2011). *PLoS One*, **6**, e22013.