Hyung Ho Lee^{a,b*} and Sang Taek Jung^a

^aDepartment of Bio and Nano Chemistry, Kookmin University, Seoul 136-702, Republic of Korea, and ^bDepartment of Integrative Biomedical Science and Engineering, Kookmin University, Seoul 136-702, Republic of Korea

Correspondence e-mail: hhlee@kookmin.ac.kr

Received 5 October 2012

Accepted 29 November 2012

Overexpression, crystallization and preliminary X-ray crystallographic analysis of β -N-acetylglucosaminidase from *Thermotoga maritima* encoded by the *Tm0809* gene

β -N-acetylglucosaminidase (NagA) protein has a chitin-degrading activity and chitin is one of the most abundant polymers in nature. NagA contains a family 3 glycoside (GH3)-type N-terminal domain and a unique C-terminal domain. The structurally uncharacterized C-terminal domain of NagA may be involved in substrate specificity. To provide a structural basis for the substrate specificity of NagA, structural analysis of NagA from *Thermotoga maritima* encoded by the *Tm0809* gene was initiated. NagA from *T. maritima* has been overexpressed in *Escherichia coli* and crystallized at 296 K using ammonium sulfate as a precipitant. Crystals of *T. maritima* NagA diffracted to 3.80 Å resolution and belonged to the monoclinic space group C2, with unit-cell parameters $a = 231.15$, $b = 133.62$, $c = 140.88$ Å, $\beta = 89.97^\circ$. The crystallization of selenomethionyl-substituted protein is in progress to solve the crystal structure of *T. maritima* NagA.

1. Introduction

Chitin, a long chain of *N*-acetylglucosamine, is the second most abundant biomass present in nature, behind only cellulose (Li & Roseman, 2004; Shaikh & Deshpande, 1993). For the utilization of this widely distributed biopolymer, organisms such as plants, insects, fungi and bacteria use chitinolytic enzymes (Bhattacharya *et al.*, 2007; Shaikh & Deshpande, 1993). β -N-Acetylglucosaminidase is one of the enzymes used for the conversion of chitins into monomeric sugars for glycolysis (Choi *et al.*, 2009; Li & Roseman, 2004). Following the depolymerization by chitinases, which degrades long-chain *N*-acetylglucosamines to short-chain chito-oligosaccharides, the resulting oligosaccharides are hydrolysed to monomeric *N*-acetylglucosamines by β -N-acetylglucosaminidase (Keyhani & Roseman, 1999).

Thermotoga maritima is a Gram-negative bacterium isolated from geothermally heated marine floor (Huber *et al.*, 1986). This extreme thermophilic bacterium exhibited optimal growth at approximately 353 K and has been reported to thrive at up to 363 K (Huber *et al.*, 1986). This bacterium, which is one of the most slowly evolving bacteria phylogenetically (Achenbach-Richter *et al.*, 1987), has a morphologically unique sheath-like structure called the 'toga' (Huber *et al.*, 1986). *T. maritima* β -N-acetylglucosaminidase composed of 467 amino acids is encoded by the *nagA* gene (gene locus tag *Tm0809*; GenBank ID 898477; Nelson *et al.*, 1999). The *T. maritima* NagA protein hydrolyses chito-oligosaccharides and is a glycoside hydrolase family 3 (GH3) enzyme according to similarities in amino-acid sequence and three-dimensional folding (Cantarel *et al.*, 2009). Although the function of the *T. maritima* NagA protein has not been thoroughly investigated, it has been reported that the catalytic efficiency (k_{cat}/K_m) for this enzyme is similar to that of the β -N-acetylglucosaminidase isolated from *Thermotoga neapolitana* towards *p*-nitrophenyl *N*-acetyl β -D-chitobiose (pNPGlcNAc), *p*-nitrophenyl di-*N*-acetyl β -D-chitobiose (pNPGlcNAc₂) and *p*-nitrophenyl tri-*N*-acetyl β -D-chitobiose (pNPGlcNAc₃) as substrates (Choi *et al.*, 2009). In sharp contrast to other GH3 family enzymes, the homodimeric β -N-acetylglucosaminidase from *T. maritima* accommodates chitobiose substrates such as di-*N*-acetyl β -D-chitobiose (GlcNAc₂) and tri-*N*-acetyl β -D-chitobiose (GlcNAc₃) preferentially for conversion into monomeric GlcNAc, which is further converted into fructose 6-phosphate, a metabolizable sugar for glycolysis (Cheng *et al.*, 2000;

Chitlaru & Roseman, 1996; Choi *et al.*, 2009; Li *et al.*, 2002). While the *nagA* gene is present for the conversion of chito-oligosaccharides into monomeric GlcNAc, no chitinase gene has been identified for the degradation of chitin into chito-oligosaccharide in the *T. maritima* genome, suggesting the cohabitation of this bacterium with other species enabling the transformation of chitin to chito-oligosaccharides for chitin utilization (Choi *et al.*, 2009).

To explain the discrepancy of substrate specificities between β -N-acetylglucosaminidase from *T. maritima* and other GH3 family enzymes, it has been suggested that the unique C-terminal domain of β -N-acetylglucosaminidase of *T. maritima*, which exhibits low sequence homology to other GH3-type enzymes, is critical in determining the substrate specificities (Yoon *et al.*, 2012). In order to facilitate further structural comparisons among GH3-type enzymes,

we are interested in determining the three-dimensional structure of the NagA protein from *T. maritima* (Tm0809), which shares 63% amino-acid sequence identity with the structurally uncharacterized *T. neapolitana* NagA proteins. NagA protein from *T. maritima* has been overexpressed in *Escherichia coli* and crystallized. Its crystallization conditions and the collection of X-ray crystallographic data are reported here.

2. Materials and methods

2.1. Protein expression and purification

The Tm0809 gene encoding NagA of *T. maritima* was amplified from the genomic DNA by polymerase chain reaction. The forward and reverse oligonucleotide primers were 5'-G G CTA TAT **GGA TCC** ATG GAC GTG GAT CTT GGA AAG C-3' and 5'-G CTA ATT **GCG GCC GCT** CATG AAT CAC ACC TCC CTA AAA AAT G-3', respectively. The bases in bold represent the *Bam*HI and *Not*I restriction enzyme cleavage sites. The amplified DNA was digested with *Bam*HI and *Not*I and was then inserted into the *Bam*HI/*Not*I-digested expression vector pGST2 parallel vector (Sheffield *et al.*, 1999). The plasmid was transformed into *E. coli* strain BL21(DE3) star cells for protein expression. The cells were grown at 310 K to an OD₆₀₀ of 0.5 in Luria-Bertani medium containing 50 μ g ml⁻¹ ampicillin and protein expression was induced by 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG). Cell growth continued at 298 K for 18 h after IPTG induction and the cells were harvested by centrifugation at 5000 rev min⁻¹ (Vision V10004A rotor) for 30 min at 277 K. The cell pellet was resuspended in ice-cold lysis buffer A (20 mM Tris-HCl pH 8.0, 200 mM NaCl) containing 1 mM phenylmethylsulfonyl fluoride and was homogenized with an ultrasonic processor. The crude cell extract was centrifuged at 12 000 rev min⁻¹ (Vision V506CA rotor) for 30 min at 277 K. The supernatant was applied onto a glutathione-Sepharose column (GE Healthcare). The fusion protein was eluted with elution buffer (15 mM glutathione, 150 mM Tris-HCl pH 8.0) and cleaved with His₆-tagged TEV protease while bound to the matrix at 277 K overnight. Further purification was performed by gel filtration on a HiLoad 16/60 Superdex 200 prep-grade column (GE Healthcare), which was previously equilibrated with buffer A. The homogeneity of the purified protein was assessed

Figure 1
Coomassie gel of NagA protein from *T. maritima* isolated by size-exclusion chromatography: lane 1, size marker (labelled in kDa); lanes 2–8, fractions of NagA protein.

Figure 2
Crystals and diffraction patterns of β -N-acetylglucosaminidase NagA from *T. maritima*. (a) NagA crystals with approximate dimensions of 0.05 $\times$ 0.05 $\times$ 0.5 mm. (b) A representative diffraction image collected from a single NagA crystal. The concentric red circles show the resolution limits.

Table 1

(a) Data collection and refinement statistics.

Values in parentheses are for the highest-resolution shell.

X-ray wavelength (Å)	0.97951
Detector	ADSC Quantum 270 CCD
No. of crystals	1
Crystal-to-detector distance (mm)	300
Rotation range per frame (°)	1
Total rotation range (°)	360
Space group	C2
Unit-cell parameters (Å, °)	$a = 231.15$, $b = 133.62$, $c = 140.88$, $\beta = 89.97$
Resolution range (Å)	50–3.8 (3.87–3.80)
Total/unique reflections	323 389/43 210
Completeness (%)	99.9 (100)
Mean $I/\sigma(I)$	7.5 (6.0)
$R_{\text{merge}}^{\dagger}$ (%)	13.2 (50.0)
Average mosaicity (°)	0.53

(b) Cell-content analysis.

No. of monomers in asymmetric unit	4	6	8	10
Solvent content (%)	76	64	52	40
V_M (Å ³ Da ^{−1})	5.15	3.43	2.58	2.06

$\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection hkl , $\sum_{hkl}$ is the sum over all reflections, and $\sum_i$ is the sum over i measurements of reflection hkl .

by polyacrylamide gel electrophoresis in the presence of 0.1% (w/v) sodium dodecyl sulfate (Laemmli, 1970). The protein solution was concentrated using a YM10 ultrafiltration membrane (Amicon) to about 7 mg ml^{−1}. The protein concentration was estimated by measuring the absorbance at 280 nm, employing the calculated molar extinction coefficient of 32 320 M^{−1} cm^{−1} (Swiss-Prot; <http://www.expasy.ch/>).

2.2. Crystallization and X-ray data collection

Crystallization experiments were carried out using the sitting-drop vapour-diffusion method at 296 K using 96-well plates (Greiner). A sitting drop was prepared by mixing equal volumes (1 µl each) of the protein solution and the reservoir solution. The sitting drop was placed over 70 µl reservoir solution. Initial crystallization conditions were established using screening kits from Qiagen (JCSG Core Suites I, II, III and IV) and from Emerald BioSystems (Wizard Classic 1, 2, 3 and 4). Crystals of *T. maritima* NagA obtained using ammonium sulfate as a precipitant were further optimized by pH screening kits (Hampton Research).

Crystals were flash-cooled in a liquid-nitrogen stream employing 100% Paratone-N as a cryoprotectant. X-ray diffraction data of *T. maritima* NagA protein were collected at 100 K on the BL-5C experimental station of the Pohang Light Source, Pohang, Republic of Korea, with an ADSC Quantum 270 CCD detector system (Area Detector Systems Corporation, Poway, California, USA). The wavelength of the synchrotron X-rays was 0.97951 Å and a 0.3 mm collimator was used. The crystal was rotated through a total of 360° with 1.0° oscillation range per frame and 1 s exposure time. The raw

data were processed and scaled using the program *HKL-2000* (Otwinowski & Minor, 1997).

3. Results

The intact form of the NagA protein from *T. maritima* was over-expressed in a soluble form with a yield of ~15 mg homogeneous protein per litre of culture (Fig. 1). The best crystals were produced by mixing equal volumes (1 µl each) of the protein solution (7 mg ml^{−1} in 20 mM Tris–HCl pH 8.0, 200 mM NaCl) and reservoir solution [100 mM phosphate–citrate pH 4.2 and 8–10% (w/v) polyethylene glycol 4000]. Crystals grew to maximum dimensions of 0.05 × 0.05 × 0.5 mm within several days (Fig. 2). A set of X-ray diffraction data was collected to 3.80 Å resolution at 100 K. A total of 323 389 measured reflections were merged into 43 210 unique reflections, giving an R_{merge} of 13.2% and a completeness of 99.9%. The space group was determined to be C2 on the basis of systematic absences and the symmetry of diffraction intensities. The unit-cell parameters are $a = 231.15$, $b = 133.62$, $c = 140.88$ Å, $\beta = 89.97^\circ$. Table 1 summarizes the statistics of data collection. The molecular mass of the recombinant NagA protein was estimated to be ~100 kDa by size-exclusion chromatography (Fig. 1), indicating that the enzyme exists as a dimer in solution (calculated monomer mass = 52 781 Da).

The authors thank Bum Jun Park for the initial crystallization of NagA protein from *T. maritima*. The authors also thank the staff at beamline BL-5C of the Pohang Light Source for assistance during X-ray experiments. This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST; grant Nos. 2012-0004058 and 2012K1A3A1A09033383).

References

- Achenbach-Richter, L., Gupta, R., Stetter, K. O. & Woese, C. R. (1987). *Syst. Appl. Microbiol.* **9**, 34–39.
- Bhattacharya, D., Nagpure, A. & Gupta, R. K. (2007). *Crit. Rev. Biotechnol.* **27**, 21–28.
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.
- Cheng, Q., Li, H., Merdek, K. & Park, J. T. (2000). *J. Bacteriol.* **182**, 4836–4840.
- Chitlaru, E. & Roseman, S. (1996). *J. Biol. Chem.* **271**, 33433–33439.
- Choi, K.-H., Seo, J. Y., Park, K.-M., Park, C.-S. & Cha, J. (2009). *J. Biosci. Bioeng.* **108**, 455–459.
- Huber, R., Langworthy, T., König, H., Thomm, M., Woese, C., Sleytr, U. & Stetter, K. (1986). *Arch. Microbiol.* **144**, 324–333.
- Keyhani, N. O. & Roseman, S. (1999). *Biochim. Biophys. Acta*, **1473**, 108–122.
- Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
- Li, H., Morimoto, K., Katagiri, N., Kimura, T., Sakka, K., Lun, S. & Ohmiya, K. (2002). *Appl. Microbiol. Biotechnol.* **60**, 420–427.
- Li, X. & Roseman, S. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 627–631.
- Nelson, K. E. *et al.* (1999). *Nature (London)*, **399**, 323–329.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Shaikh, S. A. & Deshpande, M. V. (1993). *World J. Microbiol. Biotechnol.* **9**, 468–475.
- Sheffield, P., Garrard, S. & Derewenda, Z. (1999). *Protein Exp. Purif.* **15**, 34–39.
- Yoon, B.-Y., Jiao, L., Moon, H. R., Cha, J. & Ha, N.-C. (2012). *Acta Cryst.* **F68**, 56–58.