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

Katie J. Gregg and Alisdair B. Boraston*

Department of Biochemistry and Microbiology, University of Victoria, PO Box 3055 STN CSC, Victoria, British Columbia V8W 3P6, Canada

Correspondence e-mail: boraston@uvic.ca

Received 9 September 2008
Accepted 12 December 2008

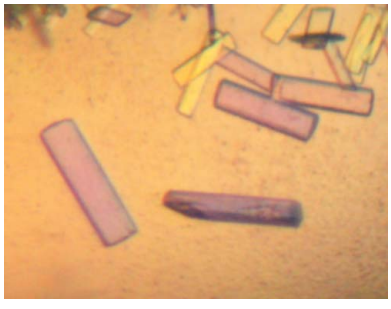
Cloning, recombinant production, crystallization and preliminary X-ray diffraction analysis of a family 101 glycoside hydrolase from *Streptococcus pneumoniae*

Streptococcus pneumoniae is a serious human pathogen that is responsible for a wide range of diseases including pneumonia, meningitis, septicaemia and otitis media. The full virulence of this bacterium is reliant on carbohydrate processing and metabolism, as revealed by biochemical and genetic studies. One carbohydrate-processing enzyme is a family 101 glycoside hydrolase (*Sp*GH101) GH101) that is responsible for catalyzing the liberation of galactosyl β1,3-*N*-acetyl-D-galactosamine (Galβ1,3GalNAc) α-linked to serine or threonine residues of mucin-type glycoproteins. The 124 kDa catalytic module of this enzyme (*Sp*GH101CM) was cloned and overproduced in *Escherichia coli* and purified. Crystals were obtained in space group $P2_1$ and diffracted to 2.0 Å resolution, with unit-cell parameters $a = 81.86$, $b = 88.91$, $c = 88.77$ Å, $\beta = 112.46^\circ$. *Sp*GH101CM also qualitatively displayed good activity towards the synthetic substrate *p*-nitrophenyl-2-acetamido-2-deoxy-3-*O*-(β-D-galactopyranosyl)-α-D-galactopyranoside, which is consistent with the classification of this enzyme as an endo-α-*N*-acetylgalactosaminidase.

1. Introduction

Pneumonia is an acute inflammatory illness of the lungs that is caused by a variety of bacteria and viruses as well as certain fungi and protozoans. *Streptococcus pneumoniae* is a Gram-positive encapsulated diplococcus that is a major causative agent of pneumonia. There are over 90 serotypes of this bacterium, which are defined based on the variable composition of the polysaccharide capsule of *S. pneumoniae*. *S. pneumoniae* is a human commensal that colonizes the nasopharynx of approximately 40% of individuals asymptomatically and has no environmental niche (Kadioglu & Andrew, 2004). The innate and adaptive immune system typically prevents colonization from becoming disease, but host–pathogen homeostasis can be altered, leading to not only pneumonia but also meningitis, septicaemia and otitis media (García-Suárez *et al.*, 2006). Invasive *S. pneumoniae* infections cause more deaths than any other bacterium and are the fifth leading cause of death worldwide (Kadioglu & Andrew, 2004). The importance of *S. pneumoniae* as a pathogen is driving studies of its virulence factors and other aspects of the host–pathogen interaction in the hope that this will ultimately aid in the development of new strategies to deal with infections caused by this bacterium.

Genome sequencing, signature-tagged mutagenesis and other biochemical and genetic studies have revealed the reliance of *S. pneumoniae* on carbohydrate processing and metabolism for full virulence of the bacterium (Tettelin *et al.*, 2001; Hava & Camilli, 2002; Boraston *et al.*, 2006; Shelburne *et al.*, 2008). One component of the extracellular cell-wall-attached armory of enzymes in *S. pneumoniae* is an endo-α-*N*-acetylgalactosaminidase that catalyzes the liberation of galactosyl β1,3-*N*-acetyl-D-galactosamine (Galβ1,3GalNAc) α-linked to serine or threonine residues of mucin-type glycoproteins. Although the identity of the gene encoding this protein has remained unreported in the literature, recent identification and characterization of proteins with very similar activities have suggested that the activity of *S. pneumoniae* is attributable to the hypothetical protein



SP_0368 from *S. pneumoniae* TIGR4 (Bhavanandan *et al.*, 1976; Koutsioulis *et al.*, 2008). This hypothetical protein is classified as a family 101 glycoside hydrolase together with other known endo-α-*N*-acetylgalactosaminidases (http://www.cazy.org/; Henrissat, 1991). Currently, no three-dimensional structure has been determined for a family 101 glycoside hydrolase. In this communication, we report the cloning, recombinant production, crystallization and preliminary X-ray diffraction data of a 124 kDa fragment of the *S. pneumoniae* hypothetical protein SP_0368 (here, the catalytic module fragment is called *Sp*GH101CM), which harbours endo-α-*N*-acetylgalactosaminidase activity.

2. Materials and methods

2.1. Cloning, production and purification of *Sp*GH101

The gene fragment encoding the GH101 catalytic module was PCR-amplified from *S. pneumoniae* TIGR4 genomic DNA (ATCC BAA-334D) using the following oligonucleotide primers: 5′-GGC AGC CAT ATG GAA AAA GAA ACA GGT CCT G-3′ and 5′-GGA TCC CTC GAG TTA CAA CAT CTT ACC TG-3′. The PCR-amplified gene fragment was obtained using standard PCR methods using Phusion High-Fidelity DNA Polymerase (New England Biolabs). The product was digested with *Nde*I and *Xho*I restriction endonucleases and ligated to similarly digested pET-28a(+) (Novagen) using standard cloning procedures. The resultant plasmid encodes a polypeptide consisting of residues 317–1425 of the unprocessed sequence preceded by MGSSHHHHHHSSGLVPR-GSH: an N-terminal six-histidine tag followed by a thrombin protease cleavage site.

The *Sp*GH101 catalytic module was produced in 4 l cultures of *Escherichia coli* BL21 Star (DE3) (Invitrogen) in Luria–Bertani medium containing 50 µg ml^{-1} kanamycin (Sigma). Cells were harvested using centrifugation and were resuspended in 30 ml 25% sucrose in 20 m*M* Tris–HCl pH 8.0. 10 mg lysozyme was added to the resuspended cells and stirred for 10 min. 60 ml of 1% deoxycholate, 1% Triton X-100, 20 m*M* Tris–HCl pH 7.5, 100 m*M* NaCl was then added to the cells and stirred for an additional 10 min. Finally, 0.5 mg DNase (Sigma) and 5 m*M* $MgCl_2$ was added to the lysed cells and allowed to spin for another 10 min. Cell debris was pelleted using centrifugation at 27 000*g* for 45 min. The polypeptide was purified from cell-free extract using immobilized metal-affinity chromatography (IMAC). The supernatant was loaded onto a nickel resin (Sigma His-Select) and protein elution began with a stepwise gradient of imidazole. The purity of the fractions was assessed using SDS–PAGE and those deemed to be greater than 95% pure were pooled. The pooled polypeptides were concentrated and exchanged into 20 m*M* Tris–HCl pH 8.0 in a stirred ultrafiltration unit (Amicon) using a 10 kDa molecular-weight cutoff membrane (Filtron). The protein was further purified by size-exclusion chromatography using Sephacryl S-200 (GE Biosciences) in 20 m*M* Tris–HCl pH 8.0. The concentration of purified protein was determined from the UV absorbance at 280 nm using a calculated molar extinction coefficient of 240 420 M^{-1} cm^{-1}.

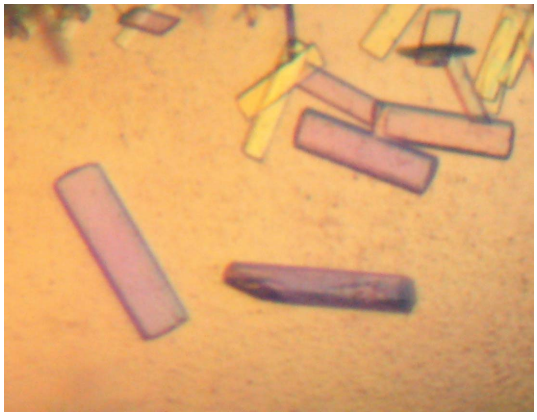


Figure 1
Crystals of the GH101 catalytic module from *S. pneumoniae* TIGR4 grown in 25% polyethylene glycol 1500.

Table 1
Data-collection statistics.

Values in parentheses are for the highest resolution bin.

Wavelength (Å)	1.5418
Space group	$P2_1$
Unit-cell parameters (Å, °)	a = 81.86, b = 88.91, c = 88.77, β = 112.46
Resolution range (Å)	19.98–2.00 (2.07–2.00)
Total No. of reflections	300953
No. of unique reflections	76040
Average redundancy	3.96 (3.88)
Completeness (%)	95.8 (93.1)
$R_{\rm merge}$†	0.113 (0.398)
Reduced χ^2	0.98 (1.08)
$I/\sigma(I)$	7.9 (3.1)

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

2.2. Crystallization and X-ray data collection

Prior to crystallization, the *Sp*GH101 catalytic module was concentrated to 15 mg ml^{-1} in 20 m*M* Tris–HCl pH 8.0. Crystals were grown within one week by adding 1 µl 25% polyethylene glycol (PEG) 1500 (Hampton Research) to 1 µl protein solution using the hanging-drop vapour-diffusion method at 292 K. Removal of the six-histidine tag was unnecessary for crystallization. Crystals were cryoprotected in 1 µl 33% PEG 1500 supplemented with 6% MPD (Hampton Research) and flash-cooled directly in a nitrogen-gas stream at 113 K. Diffraction experiments were performed on a 'home-beam' Micromax 002 X-ray source equipped with Osmic Blue Optics, an Oxford Cryo 700 System and an R-AXIS IV^{++} area detector. 410 images were collected at 0.5° intervals with an exposure time of 2 min; *d*TREK* was used for data processing (Pflugrath, 1999).

3. Results and discussion

*Sp*GH101 is a large multimodular protein, as is common for glycoside hydrolases, comprising 1767 amino acids in three definable domains or modules sandwiched by an N-terminal secretion signal peptide and a C-terminal LP*X*TG cell-wall attachment motif. The first module following the signal peptide comprises 278 amino acids and is of unknown identity. The following amino acids 317–1425 comprise the catalytic domain of *Sp*GH101, here called *Sp*GH101CM, and neighbouring the catalytic module is a carbohydrate-binding module. In an effort to characterize the structure of this *S. pneumoniae* protein, we cloned the gene fragment that we predicted to contain the catalytic module (*Sp*GH101CM), recombinantly produced the 1109-amino-acid 124 kDa polypeptide in *E. coli* and purified it in high yields of near 30 mg per litre of culture. The resulting polypeptide qualitatively displayed good activity towards the synthetic substrate *p*-nitrophenyl-2-acetamido-2-deoxy-3-*O*-(β-D-galactopyranosyl)-α-D-galactopyranoside (Toronto Research Chemical Inc.). This was consistent with the classification of *Sp*GH101 as an endo-α-*N*-acetylgalactosaminidase

and strongly suggests that this hypothetical protein is indeed the previously characterized endo-α-*N*-acetylgalactosaminidase from *S. pneumoniae* that is available in commercial preparations and has been referred to as EngSP (Bhavanandan *et al.*, 1976; Koutsioulis *et al.*, 2008).

Diffraction-quality crystals of *Sp*GH101CM were grown within 1–2 weeks of setting up the crystallization experiment (Fig. 1). The crystals diffracted to 2.0 Å resolution and belonged to space group $P2_1$, with unit-cell parameters a = 81.86, b = 88.91, c = 88.77 Å, $\beta = 112.46^\circ$ (Table 1). Analysis of the contents of the asymmetric unit indicated that it contains only one 124 kDa *Sp*GH101CM molecule with a predicted solvent content of ∼49%. Native Patterson and self-rotation function analyses did not reveal any peaks above the background, which is consistent with the presence of a single molecule of SpGH101CM in the asymmetric unit.

Structural studies of family 101 glycoside hydrolases have been lacking, which is likely to be a consequence of their generally very large size which makes them recalcitrant to crystallization. We have dissected *Sp*GH101 into a smaller fragment that retains catalytic activity and crystallizes readily. These crystals are of sufficient quality to enable the determination of a high-resolution crystal structure of this protein. Determining the three-dimensional structure of this enzyme will not only help to determine the structure and catalytic mechanism of this particular enzyme but will also provide considerable insight into this uncharacterized family of glycoside hydrolases.

Note added in proof: during the review of this manuscript, Caines *et al.* (2008) published the 2.9 Å resolution crystal structure of a protein that comprised residues 4–1567 of *Sp*GH101 from *S. pneumoniae* R6. Our crystals are of a different form, are of higher quality and diffract to substantially higher resolution and will thus be of utility in studying the structural basis of substrate and inhibitor recognition by this protein. Structure solution by molecular replacement is ongoing.

This work was supported by a grant from the Canadian Institutes of Health Research. ABB is a Canada Research Chair in Molecular Interactions and a Michael Smith Foundation for Health Research Scholar.

References

Bhavanandan, V. P., Umemoto, J. & Davidson, E. A. (1976). *Biochem. Biophys. Res. Commun.* **70**, 738–745.
Boraston, A. B., Wang, D. & Burke, R. D. (2006). *J. Biol. Chem.* **281**, 35263–35271.
Caines, M. E. C., Zhu, H., Vuckovic, M., Willis, L. M., Withers, S. G., Wakarchuk, W. W. & Strynadka, N. C. J. (2008). *J. Biol. Chem.* **283**, 31279–31283.
García-Suárez, M. d. M., Vazquez, F. & Mendez, F. J. (2006). *Enferm. Infecc. Microbiol. Clin.* **24**, 512–517.
Hava, D. L. & Camilli, A. (2002). *Mol. Microbiol.* **45**, 1389–1406.
Henrissat, B. (1991). *Biochem. J.* **280**, 309–316.
Kadioglu, A. & Andrew, P. W. (2004). *Trends Immunol.* **25**, 143–149.
Koutsioulis, D., Landry, D. & Guthrie, E. P. (2008). *Glycobiology*, **10**, 799–805.
Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.
Shelburne, S. A., Davenport, M. T., Keith, D. B. & Musser, J. M. (2008). *Trends Microbiol.* **16**, 318–325.
Tettelin, H. *et al.* (2001). *Science*, **293**, 498–506.