Crystallization and preliminary X-ray diffraction analysis of a trimodular endo- β -1,4-glucanase (Cel5B) from *Bacillus halodurans*

Immacolata Venditto,^a Helena Santos,^a James Sandy,^b Juan Sanchez-Weatherby,^b Luis M. A. Ferreira,^a Kazuo Sakka,^c Carlos M. G. A. Fontes^{a*} and Shabir Najmudin^{a*}

^aCIISA, Faculdade de Medicina Veterinária, Universidade de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisbon, Portugal, ^bDiamond Light Source, Harwell Science and Innovation Campus, Chilton, Didcot OX11 0DE, England, and ^cGraduate School of Bioresources, Mie University, Tsu 514-8507, Japan

Correspondence e-mail: cafontes@fmv.ulisboa.pt, shabir@fmv.ulisboa.pt

Received 26 August 2014

Accepted 21 October 2014

Cellulases catalyze the hydrolysis of cellulose, the major constituent of plant biomass and the most abundant organic polymer on earth. Cellulases are modular enzymes containing catalytic domains connected, *via* linker sequences, to noncatalytic carbohydrate-binding modules (CBMs). A putative modular endo- β -1,4-glucanase (*BhCel5B*) is encoded at locus BH0603 in the genome of *Bacillus halodurans*. It is composed of an N-terminal glycoside hydrolase family 5 catalytic module (GH5) followed by an immunoglobulin-like module and a C-terminal family 46 CBM (*BhCBM46*). Here, the crystallization and preliminary X-ray diffraction analysis of the trimodular *BhCel5B* are reported. The crystals of *BhCel5B* belonged to the orthorhombic space group $P2_12_12$ and data were processed to a resolution of 1.64 Å. A molecular-replacement solution has been found.

1. Introduction

Plant cell walls are complex matrices composed primarily of structural carbohydrates that provide a robust protective layer (Mohnen, 2008). The degradation of this lignocellulosic matrix requires the action of a large consortia of cellulolytic enzymes, primarily glycoside hydrolases, pectinases, polysaccharide lyases and carbohydrate esterases, which often have different carbohydrate-binding modules (CBMs) attached to them (Gowen & Fong, 2010; Minic & Jouanin, 2006). They are either secreted as free cellulases in aerobic micro-organisms or as part of large multienzyme complexes (cellulosomes; Bayer *et al.*, 2004; Fontes & Gilbert, 2010). CBMs potentiate the catalytic activity of associated catalytic modules by promoting close proximity between enzymes and insoluble substrates. Glycoside hydrolases (GHs) catalyze the cleavage of β -1,4-glycosidic bonds of cellulose (Henrissat & Davies, 1997). The genome of *Bacillus halodurans* (Takami *et al.*, 2000) encodes a putative modular endo- β -1,4-glucanase (*BhCel5B*) composed of an N-terminal glycoside hydrolase family 5 catalytic module (GH5) followed by an immunoglobulin-like domain (Ig) and a C-terminal family 46 CBM (*BhCBM46*) at locus BH0603 (GenBank accession No. BA000004). Recently, we solved the structure of the *BhCBM46* module (PDB entries 4uz8 and 4uzn; Venditto *et al.*, 2014). In the present communication, we describe the crystallization and preliminary X-ray diffraction analysis of the full-length trimodular endo- β -1,4-glucanase from *B. halodurans*.

2. Materials and methods

2.1. Macromolecule production

Endo- β -1,4-glucanase (*BhCel5B*) is a modular enzyme composed of an N-terminal glycoside hydrolase family 5 catalytic module (GH5) followed by an immunoglobulin-like module (Ig) and a C-terminal family 46 CBM. *B. halodurans* genomic DNA was used and the gene encoding *BhCel5B* (GenBank accession No. BA000004) was amplified and inserted into the pET-28a vector (Novagen), generating pCel5B (Wamalwa *et al.*, 2006). The encoded recombinant protein contained an N-terminal His₆ tag. The plasmid DNA, pCel5B, was transformed into *Escherichia coli* BL21 cells by the heat-shock method and the cells were cultured in Luria–Bertani broth at 310 K to mid-exponential phase ($A_{600\text{ nm}} = 0.6$). Recombinant protein

Table 1

Macromolecule-production information.

The cloning artefact containing the N-terminal His₆ tag, the putative boundaries between the GH5 and Ig modules and between the Ig and CBM46 modules at positions Ser365/Ser366 and Thr459/Pro460, respectively, are underlined.

Source organism	<i>B. halodurans</i>
DNA source	BA000004
Cloning vector	pET-28a
Expression host	<i>E. coli</i> BL21
Complete amino-acid sequence of the construct produced	MGSSHHHHSSGLVPRGSHMASMKWMSMAWLAVVLVSVFV- APAVSSAHEDVKTLDIQSYVRDMQPGWNLGNTFDVAGQDE- TAWGNPRVTRELIEQIADEGYKSIRIPVTWENRIGGAPDY- PIDPQFLNRVDQVQWALEEDLYVMINLHSDSLWLYEME- HNYNGVMVKYRSLWEQLSNHFKDYPTKLMFESVNEPKFSQ- NWGEIRENHALLDDLNTVFVEIVRQSGGQNDIRPLVLPT- METATSQPLLNNLYQTIDKLDDPNLIATVHYGFVFPFSVN- IAGYTRFEENSKEIIEAFDRVHHTFVARGIPVVLGEFGL- LGFDKHTGVIIQGEKLFKEFLIHHLNERDITHMLWDNGQ- HFNRTHTYEWYDQELFDMMRASWEGRSSVAESNFYILKQGD- RIADATVSLQLHGNETGLRANGQRLTPGQDYELNGERLT- VKAHVLSAIASSGTLGTNGMVTAEFNRGADWHFRVNTYRT- PVLQSTQGHVSNFSPASFNGNSLATMEAVYVDGGNAGPQ- DWTSKFEFGYAFSPSYDANEMKLEAFFREVRDGEVRLTF- HFWSGETVNYTIKNGNQVTGIAAQTTNSKNKNKK

Table 2

Crystallization.

Method	Sitting-drop vapour-diffusion
Plate type	MRC 96-well sitting-drop crystallization plate (Molecular Dimensions, UK)
Temperature (K)	292
Protein concentration (mg ml ⁻¹)	22
Buffer composition of protein solution	Deionized water (Sigma), 0.5 M CaCl ₂ buffer
Composition of reservoir solution	0.2 M calcium acetate, 0.1 M cacodylate pH 6.5, 8% polyethylene glycol 8000
Volume and ratio of drop	0.7 µl 22 mg ml ⁻¹ protein solution, 0.7 µl reservoir solution
Volume of reservoir (µl)	50

overproduction was induced by adding isopropyl β-D-1-thiogalactopyranoside (1 mM final concentration) with incubation for a further 16 h at 292 K. The cells were collected after centrifugation at 5000 rev min⁻¹ for 10 min at 277 K. The pellets were frozen for 30 min and resuspended in 10 mM imidazole, 1 M NaCl, 5 mM CaCl₂, 50 mM Na HEPES pH 7.5. The cells were lysed by sonication. After further centrifugation at 12 000 rev min⁻¹ for 30 min at 277 K, the His₆-tagged recombinant protein was purified from cell-free extracts by immobilized metal-ion affinity chromatography (IMAC) as described previously (Najmudin *et al.*, 2006). Purified *Bh*Cel5B was buffer-exchanged into 50 mM Na HEPES buffer pH 7.5 containing 200 mM NaCl and 5 mM CaCl₂, and then subjected to gel filtration using a HiLoad 16/60 Superdex 75 column (GE Healthcare) at a flow rate of 1 ml min⁻¹. *Bh*Cel5B was concentrated using an Amicon 10 kDa molecular-mass centrifugal concentrator and washed three times with 20 ml water (Sigma) containing 1 mM CaCl₂. The protein concentration was estimated using a molar extinction coefficient (ϵ) of 124 790 M⁻¹ cm⁻¹ at 280 nm with a Thermo Scientific NanoDrop 2000c and was verified by SDS-PAGE (Fig. 1). Macromolecule-production information is summarized in Table 1.

2.2. Crystallization

Crystallization conditions were screened by the sitting-drop vapour-diffusion method using the commercial kits Crystal Screen, Crystal Screen 2, PEG/Ion and PEG/Ion 2 from Hampton Research, California, USA and an in-house 80 factorial screen using an Oryx8 robotic nanodrop dispensing system (Douglas Instruments). The best

Table 3

Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	102, Diamond
Wavelength (Å)	0.9795
Temperature (K)	100
Detector	PILATUS 6M detector (DECTRIS)
Rotation range per image (°)	0.2
Total rotation range (°)	200
Exposure time per image (s)	0.2
Space group	<i>P</i> 2 ₁ 2 ₁ 2
<i>a</i> , <i>b</i> , <i>c</i> (Å)	74.22, 141.84, 50.82
α , β , γ (°)	90, 90, 90
Mosaicity (°)	0.77
Resolution range (Å)	51.27–1.64 (1.68–1.64)
Total No. of reflections	480050 (33097)
No. of unique reflections	66600 (4833)
Completeness (%)	99.9 (99.8)
Multiplicity	7.2 (6.8)
$\langle I/\sigma(I) \rangle$	19.7 (2.6)
$R_{\text{merge}}^{\dagger}$ (%)	7.3 (74.3)
$R_{\text{p.i.m.}}^{\ddagger}$ (%)	4.4 (45.7)
CC _{1/2} [§]	0.99 (0.60)
Overall Wilson <i>B</i> factor (Å ²)	13.99
$V_M^{\P}$ (Å ³ Da ⁻¹)	2.22
Solvent content (%)	45

[†] $R_{\text{p.i.m.}} = \sum_{hkl} \{1/[N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $\langle I(hkl) \rangle$ is the average of symmetry-related observations of a unique reflection. [‡] $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th intensity measurement of reflection *hkl*, including symmetry-related reflections, and $\langle I(hkl) \rangle$ is its average. [§] CC_{1/2} is the half data set correlation coefficient (Diederichs & Karplus, 2013). [¶] The Matthews coefficient (Matthews, 1968) based on the molecular weight from the sequence indicates the presence of a single molecule in the asymmetric unit.

diffracting crystals were obtained from the in-house 80 factorial screen (Table 2). The crystals were vitrified in liquid nitrogen after soaking in cryoprotectant [30%(v/v) glycerol added to the crystallization buffer] for a few seconds.

2.3. Data collection and processing

The crystals were tested on beamlines I04 and 102 at the Diamond Light Source, Harwell, England. A systematic grid search was carried out on each of these crystals to select the best diffracting part of the crystal. *EDNA* (Winter & McAuley, 2011) and *iMosflm* (Battye *et al.*,

Figure 1

SDS-PAGE [14%(w/v)] showing overexpression and purification of *Bh*Cel5B. Lane 1, LMW protein marker (NZytech); lane 2, purified *Bh*Cel5B.

2011) were used for strategy calculation during data collection. All data sets were processed using the *fast_dp* (Winter, 2010) and *xia2* (Winter *et al.*, 2013) packages, which use the programs *XDS* (Kabsch, 2010), *POINTLESS* (Evans, 2006) and *SCALA* (Evans, 2006) from the *CCP4* suite (Winn *et al.*, 2011). Data-collection statistics are given in Table 3 for the best data.

3. Results and discussion

The best *BhCel5B* crystal diffracted to a resolution beyond 1.5 Å. This crystal belonged to the orthorhombic space group $P2_12_12$ and data were processed to a resolution of 1.64 Å as chosen by *xia2*. *BALBES* was used to carry out molecular replacement (Long *et al.*, 2008). The best solution was found using the GH5 catalytic domain (without the β -side domain) of endoglucanase D from *Clostridium cellulovorans* (PDB entry 3ndz, with a sequence identity of 31.1%; C. M. Bianchetti, R. W. Smith, C. A. Bingman & G. N. Phillips Jr, unpublished work), giving an *R* factor and *R*_{free} of 47.2 and 48.8%, respectively, and a *Q* factor (Krissinel & Henrick, 2004) of 0.555 after *REFMAC5* (Murshudov *et al.*, 2011) at the end of the *BALBES* run. An *ARP/wARP* (Langer *et al.*, 2008) run after *BALBES* gave an almost complete model with 521 amino-acid residues out of 541 identified in a single chain, with an estimated correctness of 98%.

The authors acknowledge financial support from Fundação para a Ciência e a Tecnologia, Portugal through projects PTDC/BIA-PRO/103980/2008 and PTDC/BIAPRO/100359/2008. This work was supported by the European Union Seventh Framework Programme (FP7 2007-2013) under the WallTraC project (grant agreement No. 263916) and BioStruct-X (grant agreement No. 283570, Biostruct-X_4399). This paper reflects the authors' views only. The European Community is not liable for any use that may be made of the information contained herein. We acknowledge Diamond Light Source, Oxfordshire for provision of synchrotron-radiation facilities (beam-

lines I04 and I02). We would also like to thank Cecília Bonifácio and Pedro Bule for help with crystallization.

References

- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Bayer, E. A., Belaich, J.-P., Shoham, Y. & Lamed, R. (2004). *Annu. Rev. Microbiol.* **58**, 521–554.
- Diederichs, K. & Karplus, P. A. (2013). *Acta Cryst.* **D69**, 1215–1222.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Fontes, C. M. G. A. & Gilbert, H. J. (2010). *Annu. Rev. Biochem.* **79**, 655–681.
- Gowen, C. M. & Fong, S. S. (2010). *Chem. Biodivers.* **7**, 1086–1097.
- Henrissat, B. & Davies, G. (1997). *Curr. Opin. Struct. Biol.* **7**, 637–644.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* **D60**, 2256–2268.
- Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
- Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* **D64**, 125–132.
- Matthews, B. W. (1968). *J. Mol. Biol.* **63**, 491–497.
- Minic, Z. & Jouanin, L. (2006). *Plant Physiol. Biochem.* **44**, 435–449.
- Mohnen, D. (2008). *Curr. Opin. Plant Biol.* **11**, 266–277.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Najmudin, S., Guerreiro, C. I. P. D., Carvalho, A. L., Prates, J. A. M., Correia, M. A. S., Alves, V. D., Ferreira, L. M. A., Romão, M. J., Gilbert, H. J., Bolam, D. N. & Fontes, C. M. G. A. (2006). *J. Biol. Chem.* **281**, 8815–8828.
- Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., Fuji, F., Hiram, C., Nakamura, Y., Ogasawara, N., Kuhara, S. & Horikoshi, K. (2000). *Nucleic Acids Res.* **28**, 4317–4331.
- Venditto, I., Santos, H., Ferreira, L. M. A., Sakka, K., Fontes, C. M. G. A. & Najmudin, S. (2014). *Acta Cryst.* **F70**, 754–757.
- Wamalwa, B. M., Sakka, M., Shiundu, P. M., Ohmiya, K., Kimura, T. & Sakka, K. (2006). *Appl. Environ. Microbiol.* **72**, 6851–6853.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.
- Winter, G., Lobley, C. M. C. & Prince, S. M. (2013). *Acta Cryst.* **D69**, 1260–1273.
- Winter, G. & McAuley, K. E. (2011). *Methods*, **55**, 81–93.