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

Purification, crystallization and preliminary X-ray characterization of the pentamodular arabinoxylanase *Ct*Xyl5A from *Clostridium thermocellum*

Joana L. A. Brás,[a] Márcia A. S. Correia,[a,b] Maria J. Romão,[b] José A. M. Prates,[a] Carlos M. G. A. Fontes[a]* and Shabir Najmudin[a]*

[a]CIISA – Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal, and [b]REQUIMTE, Departamento de Química, FCT-UNL, 2829-516 Caparica, Portugal

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

Received 11 May 2011
Accepted 30 May 2011

The cellulosome, a highly elaborate extracellular multi-enzyme complex of cellulases and hemicellulases, is responsible for the degradation of plant cell walls. The xylanase *Ct*Xyl5A (Cthe_2193) is a multimodular arabinoxylanase which is one of the largest components of the *Clostridium thermocellum* cellulosome. The N-terminal catalytic domain of *Ct*Xyl5A, which is a member of glycoside hydrolase family 5 (GH5), is responsible for the hydrolysis of arabinoxylans. Appended after it are three noncatalytic carbohydrate-binding modules (CBMs), which belong to families 6 (CBM6), 13 (CBM13) and 62 (CBM62). In addition, *Ct*Xyl5A has a fibronectin type III-like (Fn3) module preceding the CBM62 and a type I dockerin (DOK) module following it which allows the enzyme to be integrated into the cellulosome through binding to a cohesin module of the protein scaffold CipA. Crystals of the pentamodular enzyme without the DOK module at the C-terminus, with the domain architecture *Ct*GH5-CBM6-CBM13-Fn3-CBM62, have been obtained. The structure of this pentamodular xylanase has been determined by molecular replacement to a resolution of 2.64 Å using coordinates of *Ct*GH5-CBM6, Fn3 and CBM62 from the PDB as search models.

1. Introduction

The plant cell wall is one of the largest repositories of intractable and fixed carbon on earth. It comprises myriads of interlocking polysaccharides displaying a high physical and chemical complexity. Thus, a very large repertoire of enzymes is required to obtain its total degradation. Certain microorganisms have evolved a highly elaborate megadalton extracellular multi-enzyme complex of cellulases and hemicellulases, termed the cellulosome, to efficiently carry out this biological conversion of complex polysaccharides to simple monosaccharides (for reviews, see Bayer *et al.*, 2004; Fontes & Gilbert, 2010). The cellulosomal enzymes are multimodular, with a variable architecture and size. However, each has a dockerin (DOK) module which allows their integration into the cellulosome through the binding of a cohesin module (COH) of the protein scaffold. The *Clostridium thermocellum* protein scaffold, CipA, has nine COH modules (Carvalho *et al.*, 2004) and the genome sequence of the bacterium revealed the presence of 72 DOK-containing proteins. One such enzyme is the xylanase *Ct*Xyl5A (Cthe_2193), a multimodular arabinoxylanase that is one of the largest components of the *C. thermocellum* cellulosome, comprising 948 amino-acid residues (molecular mass 103 kDa). The N-terminal catalytic domain of *Ct*Xyl5A, which is a member of glycoside hydrolase family 5 (GH5), is responsible for the hydrolysis of arabinoxylans [chemically and structurally complex polysaccharides comprising a backbone of β-1,4-xylose residues decorated with arabinofuranose (Ara*f*) moieties]. Appended after the enzyme catalytic domain are three noncatalytic carbohydrate-binding modules (CBMs), which belong to families 6 (CBM6), 13 (CBM13) and 62 (CBM62). The structure of the

N-terminal bimodular *Ct*GH5-CBM6 component showed that *Ct*GH5 displays a canonical $(\alpha/\beta)_8$-barrel fold, with the enzyme catalytic domain establishing a tight hydrophobic interaction with the CBM6 module (Correia *et al.*, 2011). CBM62 binds to D-galactose and L-arabinopyranose and mediates calcium-dependent enzyme oligomerization (Montanier *et al.*, 2011). In addition, *Ct*Xyl5A has a fibronectin type III-like (Fn3) module preceding the CBM62 module (Alahuhta *et al.*, 2010) and a type I dockerin (DOK) module following it. We have obtained crystals of the pentamodular derivative of *Ct*Xyl5A excluding the C-terminal DOK module. The molecular architecture of the crystallized enzyme is *Ct*GH5-CBM6-CBM13-Fn3-CBM62. In order to obtain insights into the structural properties that govern interdomain interactions in multimodular enzymes, we aim to determine the crystal structure of *C. thermocellum* arabinoxylanase *Ct*Xyl5A. In the present communication, we describe the overproduction, purification, crystallization and preliminary X-ray analysis of *Ct*Xyl5A excluding the C-terminal dockerin.

2. Materials and methods

2.1. Protein expression and purification

*Ct*Xyl5A is a modular enzyme containing an N-terminal GH5 catalytic domain followed by CBM6 and CBM13 modules, a fibronectin type III-like (Fn3) domain, a CBM62 module and a type I dockerin module (Fig. 1). The gene encoding the N-terminus of *Ct*Xyl5A (residues 21–885, lacking the DOK module; molecular mass 91 kDa) was amplified by PCR from *C. thermocellum* genomic DNA using NZYlong DNA polymerase (NZYTech Ltd, Portugal) and primers 5′-CTC GCT AGC AGC CCG CAA CGT GGC CGG and 3′-CAC CTC GAG ATG CAC ATC ATC ATT CTC C, which contained *Nhe*I and *Xho*I restriction sites, respectively. The DNA

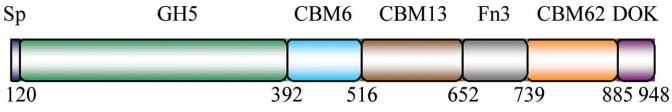

Figure 1
Domain architecture of *Ct*Xyl5A. The modules that were not included in our construct are abbreviated as follows: Sp, signal peptide; DOK, dockerin type I module (adapted from Montanier *et al.*, 2011).

Figure 2
Crystals of *Ct*Xyl5A obtained by hanging-drop vapour diffusion in the presence of 40%(*v*/*v*) 2-methyl-2,4-pentanediol and 10–20% 2-propanol. The largest crystals in the needle clusters are approximately 200 × 10 × 10 µm in size.

product was cloned into the *Nhe*I/*Xho*I sites of the *Escherichia coli* expression vector pet21a (Novagen) to generate *Ct*Xyl5A. The recombinant *Ct*Xyl5A contained a C-terminal His_6 tag. *E. coli* Tuner DE3 cells harbouring pXyl5A were cultured in Luria–Bertani broth at 310 K to mid-exponential phase ($A_{600\,nm}$ = 0.6) and recombinant protein expression was induced by the addition of 0.2 m*M* isopropyl β-D-1-thiogalactopyranoside and incubation for a further 16 h at 292 K. The His_6-tagged recombinant protein was purified from cell-free extracts by immobilized metal-ion affinity chromatography (IMAC) as described previously (Najmudin *et al.*, 2005). Purified *Ct*Xyl5A was buffer-exchanged into 50 m*M* Na HEPES buffer pH 7.5 containing 200 m*M* NaCl and 5 m*M* $CaCl_2$ 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}. Preparation of *E. coli* for the generation of seleno-L-methionine-substituted *Ct*Xyl5A was performed as described in Carvalho *et al.* (2004) and the protein was purified using the same procedures as employed for native *Ct*Xyl5A. Purified *Ct*Xyl5A was concentrated using an Amicon 10 kDa molecular-weight cutoff centrifugal concentrator and washed three times with 5 m*M* DTT containing 2 m*M* $CaCl_2$ (for the SeMet protein) or 2 m*M* $CaCl_2$ (for the native protein).

2.2. Crystallization

The crystallization conditions were screened by the hanging-drop vapour-diffusion method using the commercial kits Crystal Screen, Crystal Screen 2, PEG/Ion and PEG/Ion 2 (Hampton Research, California, USA). Drops consisting of 1 µl 20 mg ml^{-1} *Ct*Xyl5A and 1 µl reservoir solution were prepared at 292 K using both native and selenomethionine-containing protein. No hits were seen in any conditions in countless trials over a year. However, after a breakdown of the air conditioning of the 292 K crystallization room (which caused a fluctuation in the temperature of anything in the range 292–310 K over a few days) a year after the drops had been set up, six crystals of seleno-L-methionine-containing protein appeared in a single drop

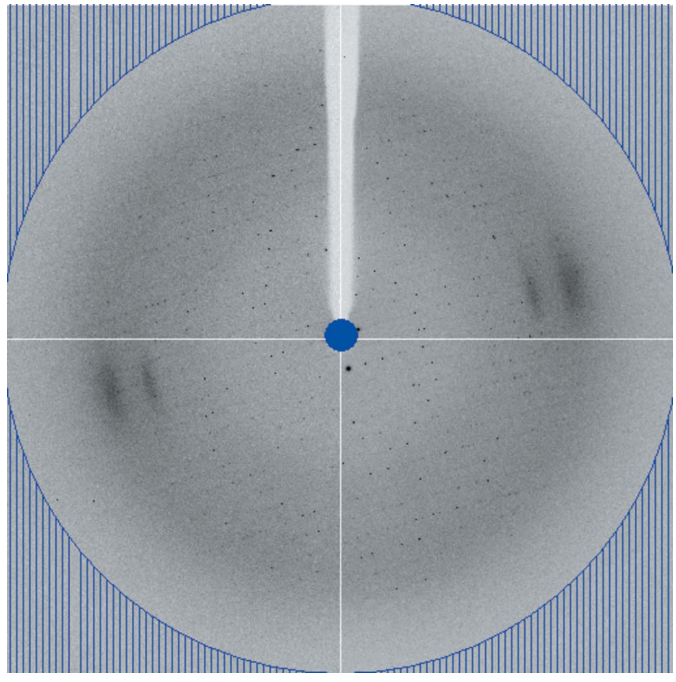

Figure 3
Representative diffraction pattern of a *Ct*Xyl5A crystal (the outer circle corresponds to 2.64 Å resolution).

Table 1
Data-collection statistics.

Values in parentheses are for the lowest/highest resolution shells.

Data set	Fixed	Peak	Inflection point
Beamline	ESRF ID14-EH1	ESRF ID14-EH4	ESRF ID14-EH4
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$
Wavelength (Å)	0.9334	0.9790	0.9795
Unit-cell parameters			
a (Å)	147.4	147.3	147.6
b (Å)	191.7	191.1	191.3
c (Å)	50.7	50.7	50.6
Resolution limits (Å)	50.7–2.64 (50.7–8.75/2.78–2.64)	80.15–2.98 (80.15–9.42/3.14–2.98)	80.26–2.98 (80.26–9.42/3.14–2.98)
No. of observations	244475 (9421/29324)	135258 (3685/19377)	139396 (3604/20027)
No. of unique observations	42246 (1525/5920)	29061 (945/4124)	29500 (906/4227)
Multiplicity	5.8 (6.2/5.0)	4.7 (3.9/4.7)	4.7 (4.0/4.7)
Completeness (%)	98.5 (99.6/96.4)	96.7 (88.8/95.9)	98.0 (85.3/98.0)
$\langle I/\sigma(I)\rangle$	8.0 (21.6/2.0)	9.9 (14.1/4)	11.6 (15.2/6.2)
$R_{\rm merge}$†	16.5 (4.5/69.5)	9.4 (5.0/23.5)	8.5 (5.8/17.9)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith measurement of reflection hkl and $\langle I(hkl)\rangle$ is the mean value for all i measurements.

in Clear Strategy Screen II condition No. 13 [40%(*v*/*v*) 2-methyl-2,4-pentanediol (MPD)]. These crystals (maximum dimensions of ∼200 × 50 × 50 µm) were immediately cryocooled in liquid nitrogen and taken to the ESRF for data collection. Numerous attempts were made to optimize these conditions in order to obtain better crystals. However, the best crystals of *Ct*Xyl5A obtained subsequently were obtained in the presence of 40%(*v*/*v*) MPD and 10–20% 2-propanol. The largest crystals in the needle clusters were approximately 200 × 10 × 10 µm in size (Fig. 2). None of these crystals diffracted as well as those obtained from the one-off anomaly.

2.3. Data collection and processing

Initially, a data set was collected from an SeMet-labelled crystal on beamline ID14-1 at the ESRF (Grenoble, France) using a Quantum 315r charge-coupled device detector (ADSC) with the crystal cooled to 100 K using a Cryostream (Oxford Cryosystems Ltd). The crystal diffracted to a resolution of 2.64 Å (Fig. 3). A second, MAD data set (at the selenium edge) was also collected on beamline ID14-4 at the ESRF (Grenoble, France) from another crystal to 2.98 Å resolution. The data were collected at wavelengths of 0.9795 Å (inflection point, $f' = -10.09$, $f'' = 3.44$), 0.9790 Å (peak, $f' = -5.78$, $f'' = 6.56$) and 0.9770 Å (remote). The crystals were delicate and suffered radiation damage after peak data collection. All data sets were processed using the programs *iMOSFLM* (Battye *et al.*, 2011) and *SCALA* (Evans, 2006) from the *CCP*4 suite (Winn *et al.*, 2011). The crystal belonged to the orthorhombic space group $P2_12_12$ according to *POINTLESS* (Evans, 2006). Data-collection statistics are given in Table 1. The Matthews coefficient ($V_{\rm M}$ = 3.66 Å^3 Da^{-1}) indicated the presence of one molecule in the asymmetric unit and a solvent content of 68% (Matthews, 1968). Attempts to solve the structure using all of the available phasing programs proved unfruitful. It was difficult to locate any of the five expected Se sites, with the best figures of merit being less than 0.2. The structure was eventually solved by molecular replacement (MR) using independently solved structures of some of the modules of *Ct*Xyl5A: *Ct*GH5-CBM6 (PDB entry 2y8k; Correia *et al.*, 2011), Fn3 (PDB entry 3mpc; Alahuhta *et al.*, 2010) and *Ct*CBM62 (PDB entries 2y8m, 2y9i and 2y9s; Montanier *et al.*, 2011) when they became available. *Phaser* was used to carry out MR in a stepwise manner (McCoy *et al.*, 2007). The structure of *Ct*GH5-CBM6 (2y8k) was initially used to find a solution, which gave RFZ and TFZ scores of 27.7 and 55.8, respectively, with an LLG of 2081. This solution was fixed and a second round of *Phaser* was performed using the Fn3 (3mpc) structure. The RFZ and TFZ scores were 3.3 and 22.6, respectively, with an LLG of 2391. In the third round of *Phaser* the solution obtained using both *Ct*GH5-CBM6 and Fn3 was fixed and *Ct*CBM62 was used as a search model. However, the LLG decreased to 1984 and a sensible solution could not be obtained for the *Ct*CBM62 domain. A *BLAST* search (Altschul *et al.*, 1990) showed that the best structural homologues of the *Ct*CBM13 module were the four ricin B-type lectin modules of the mosquitocidal toxin from *Bacillus sphaericus* (PDB entry 2vsa; Treiber *et al.*, 2008), which have sequence identities in the range 22–25%. These four domains were superposed on top of each other and also used as a search ensemble in the third round. 23 solutions were found with LLG scores ranging from 2434 to 2489. Fixing this result and searching for the *Ct*CBM62 domain once again gave no solutions. Thus, the *Ct*CBM62 domain could not be located at this stage even though SDS–PAGE analysis clearly showed that the protein had not suffered proteolysis (Fig. 4). Separate autobuilding runs were carried out using *Buccaneer* (Cowtan, 2006) and *PHENIX* (Terwilliger *et al.*, 2008) using the phases from the third round of *Phaser* with the *Ct*GH5-CBM6 and Fn3 domains as starting models in their fixed positions, but the 2vsa model was not used, thus removing model bias. *PHENIX* built 637 amino-acid residues in 11 fragments with an $R_{\rm work}$ of 28.8% and an

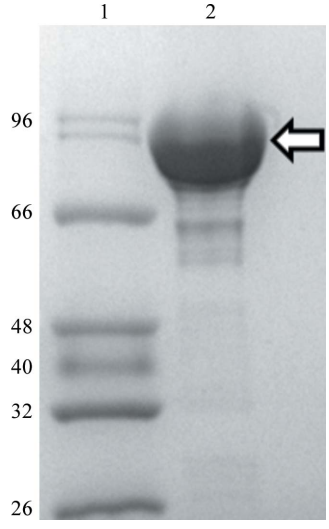

Figure 4
Coomassie Brilliant Blue-stained 14% SDS–PAGE gel evaluation of protein purity. Lane 1, low-molecular-weight protein marker (NZYtech Ltd; labelled in kDa); lane 2, *Ct*Xyl5A (marked by an arrow). The protein sample used here had been stored at 277 K for more than a year.

R_{free} of 34.0%, whereas *Buccaneer* located 657 residues in ten fragments with an R_{work} of 32.5% and an R_{free} of 37.8%. The two models were superimposed in *Coot* (Emsley & Cowtan, 2004) using *SSM* (Krissinel & Henrick, 2004) and were used for further manual rebuilding. Structure completion and analysis are ongoing.

This work was supported in part by Fundação para a Ciência e a Tecnologia (Lisbon, Portugal) through grants PTDC/BIA-PRO/103980/2008, PTDC/BIA-PRO/69732/2006 and SFRH/BD/38667/2007. The authors would like to thank Dr Ana Luisa Carvalho and Filipe Freire for their help with data collection.

References

Alahuhta, M., Xu, Q., Brunecky, R., Adney, W. S., Ding, S.-Y., Himmel, M. E. & Lunin, V. V. (2010). *Acta Cryst.* F**66**, 878–880.
Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Bayer, E. A., Belaich, J. P., Shoham, Y. & Lamed, R. (2004). *Annu. Rev. Microbiol.* **58**, 521–554.
Carvalho, A. L., Goyal, A., Prates, J. A., Bolam, D. N., Gilbert, H. J., Pires, V. M., Ferreira, L. M., Planas, A., Romão, M. J. & Fontes, C. M. (2004). *J. Biol. Chem.* **279**, 34785–34793.
Correia, M. A. S., Mazumder, K., Brás, J. L. A., Firbank, S. J., Zhu, Y., Lewis, R. J., York, W. S., Fontes, C. M. G. A. & Gilbert, H. J. (2011). *J. Biol. Chem.* **286**, 22510–22520.
Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Fontes, C. M. & Gilbert, H. J. (2010). *Annu. Rev. Biochem.* **79**, 655–681.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
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.
Montanier, C. Y., Correia, M. A. S., Flint, J. E., Zhu, Y., McKee, L. S., Prates, J. A. M., Polizzi, S. J., Coutinho, P. M., Lewis, R. J., Henrissat, B., Fontes, C. M. G. A. & Gilbert, H. J. (2011). *J. Biol. Chem.* **286**, 22499–22509.
Najmudin, S., Guerreiro, C. I. P. D., Ferreira, L. M. A., Romão, M. J. C., Fontes, C. M. G. A. & Prates, J. A. M. (2005). *Acta Cryst.* F**61**, 1043–1045.
Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst.* D**64**, 61–69.
Treiber, N., Reinert, D. J., Carpusca, I., Aktories, K. & Schulz, G. E. (2008). *J. Mol. Biol.* **381**, 150–159.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.