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

Purification, crystallization and crystallographic analysis of *Clostridium thermocellum* endo-1,4-β-D-xylanase 10B in complex with xylohexaose

Shabir Najmudin,[a]* Benedita A. Pinheiro,[b] Maria J. Romão,[a] José A. M. Prates[b] and Carlos M. G. A. Fontes[b]*

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

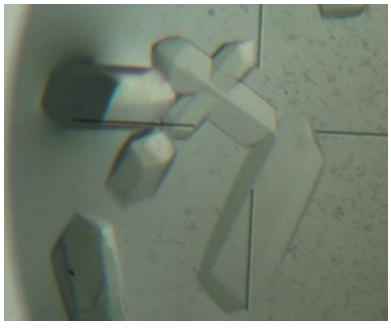
Correspondence e-mail: shabir@dq.fct.unl.pt, cafontes@fmv.utl.pt

Received 1 April 2008
Accepted 27 June 2008

The cellulosome of *Clostridium thermocellum* is a highly organized multi-enzyme complex of cellulases and hemicellulases involved in the hydrolysis of plant cell-wall polysaccharides. The bifunctional multi-modular xylanase Xyn10B is one of the hemicellulase components of the *C. thermocellum* cellulosome. The enzyme contains an internal glycoside hydrolase family 10 catalytic domain (GH10) and a C-terminal family 1 carbohydrate esterase domain (CE1). The N-terminal moiety of Xyn10B (residues 32–551), comprising a carbohydrate-binding module (CBM22-1) and the GH10 E337A mutant, was crystallized in complex with xylohexaose. The crystals belong to the trigonal space group $P3_221$ and contain a dimer in the asymmetric unit. The crystals diffracted to beyond 2.0 Å resolution.

1. Introduction

Cellulose and hemicelluloses are major constituents of plant cell walls (Bayer *et al.*, 2004; see also the Carbohydrate-Active Enzymes website http://www.cazy.org/). Xylanase Xyn10B (formally XynY; EC 3.2.1.8) from *Clostridium thermocellum* is an endo-1,4-β-D-xylan hydrolase involved in the degradation of xylan, one of the most abundant hemicelluloses. Xyn10B comprises two family 22 carbohydrate-binding modules (CBM22-1 and CBM22-2; formerly X6a and X6b, respectively), which flank the glycoside hydrolase GH10 catalytic module, a dockerin sequence and a C-terminal family 1 carbohydrate esterase catalytic module (CE1; Fontes, Hazlewood *et al.*, 1995). Structures have been elucidated of CBM22-2 (Charnock *et al.*, 2000), the dockerin module in complex with a cognate cohesin (Carvalho *et al.*, 2003) and CE1 (Prates *et al.*, 2001), providing important insights into the molecular role of Xyn10B in hemicellulose hydrolysis. CE1 (residues 792–1077) displays the α/β-hydrolase fold with a classical Ser-His-Asp catalytic triad. CE1 has been shown to cleave the ferulate groups involved in the cross-linking of arabinoxylans to lignin (Tarbouriech *et al.*, 2005; Prates *et al.*, 2001). The dockerin module (residues 730–791), which is a repeat of approximately 24 amino acids, was solved in complex with the second cohesin module of *C. thermocellum* cellulosomal scaffolding protein, termed CipA. Both helices of the dockerin helix–turn–helix motif may interact with the cohesin β-sheet, revealing a general dual binding mode of dockerins to cohesins (Carvalho *et al.*, 2003, 2005, 2007). Cellulosome assembly results from the flexible interaction of dockerin modules located on the enzymes with one of the nine cohesin domains of the CipA scaffold (Bayer *et al.*, 2004). Therefore, the cellulosome of *C. thermocellum* comprises nine enzymatic subunits in addition to CipA. CBM22-2 (residues 560–720) has a classic β-jelly-roll fold with a cleft containing three aromatic residues and two polar conserved residues (Charnock *et al.*, 2000). Trp53, Tyr103 and Glu138 play a crucial role in xylan recognition, while Tyr136 and Arg25, although involved in ligand binding, also maintain the structural integrity of the cleft (Xie *et al.*, 2001). The catalytic site of the GH10 module comprises Glu337 (proton donor) and Glu480 (nucleo-



phile) (Fontes, Hall *et al.*, 1995). A mutant construct (CBM22-1–GH10 E337A) was made to capture the substrate xylohexaose in the inactivated site. Here, we report the preliminary crystal structural characterization of the N-terminus of Xyn10B (residues 32–551), comprising the modules CBM22-1 and GH10 E337A, in complex with xylohexaose.

2. Materials and methods

2.1. Bacterial strains, plasmids and growth conditions

The *Escherichia coli* strains used in this study were XL10-Gold and BL21 (DE3) (Stratagene). The plasmids used for cloning the truncated derivative of Xyn10B were pGEM T-easy (Promega) and pET-21a (Novagen). Recombinant *E. coli* strains were cultured at 310 K in Luria–Bertani broth supplemented with ampicillin (100 µg ml^{-1}). Xylohexaose was obtained from Megazyme (Bray, Ireland).

2.2. Protein expression and purification

The DNA fragment encoding CBM22-1–GH10 was amplified by PCR from *C. thermocellum* YS genomic DNA using *Pfu* Turbo (Stratagene). Primers (5′-CTC**GCTAGC**GATTATGAAGTGGTTC-ATG-3′ and 5′-CAC**CTCGAG**GGCCGGATTGTTACCGTC-3′) incorporated *Nhe*I and *Xho*I restriction sites (in bold) at the 5′ and 3′ ends of the PCR product, respectively. The resulting DNA fragment was cloned into pGEM T-easy, generating pBP1, and sequenced to ensure that no mutations had occurred during amplification. The plasmid pBP1 was digested with *Nhe*I and *Xho*I and the excised gene was cloned into the similarly restricted expression vector pET-21a such that the recombinant protein contained an N-terminal His_6 tag. The CBM22-1–GH10 nucleophile mutant (Glu→Ala), termed CBM22-1–GH10 E337A, was generated using the PCR-based Quik-Change site-directed mutagenesis kit (Stratagene) according to the manufacturer's instructions. The primers used to generate the nucleic acid mutant were GACGTTGTAAAT**GCG**GCAGTAAGTGAT-GATGC and GCATCATCACTTACTGC**CGC**ATTTACAACGTC (nucleotide substitutions in bold). The mutated DNA sequence was sequenced by MWG (Germany) to ensure that only the appropriate mutation had been incorporated into the nucleic acid. To express the recombinant CBM22-1–GH10 E337A protein, *E. coli* strain BL21 (DE3) harbouring the pET-21a plasmid was cultured in Luria–Bertani broth at 310 K to mid-exponential phase (A_{550} = 0.6). Isopropyl β-D-1-thiogalactopyranoside was then added to a final concentration of 1 m*M* and the cultures were incubated for a further 5 h. Cells were collected by centrifugation and the cell pellet was resuspended in 50 m*M* sodium HEPES buffer pH 7.5 containing 1 *M* NaCl and 10 m*M* imidazole. CBM22-1–GH10 E337A was purified by immobilized metal-ion affinity chromatography as described previously (Najmudin *et al.*, 2006). The protein was buffer-exchanged, using a PD-10 Sephadex G-25M gel-filtration column (Amersham Biosciences), into 50 m*M* HEPES buffer pH 7.5 containing 200 m*M* NaCl (buffer *A*) and concentrated to 20 mg ml^{-1} with Amicon 10 kDa molecular-weight centrifugation membranes. Gel filtration with a HiLoad 16/60 Superdex 75 column (Amersham Biosciences) was used to further improve protein purity and the protein was eluted at 1 ml min^{-1} in buffer *A* (Fig. 1). The purified protein was concentrated as described above and washed three times with water containing 2 m*M* $CaCl_2$. The final protein concentration was adjusted to 60 mg ml^{-1}. The protein sequence of the construct corresponds to amino-acid residues 32–551 (UniProt accession No. P51584) with the E337A mutation. With the additional 20 amino-acid residues (MGSSHHHHHHSSGLVPRGSH) at the N-terminus the expected final molecular weight is 60 kDa.

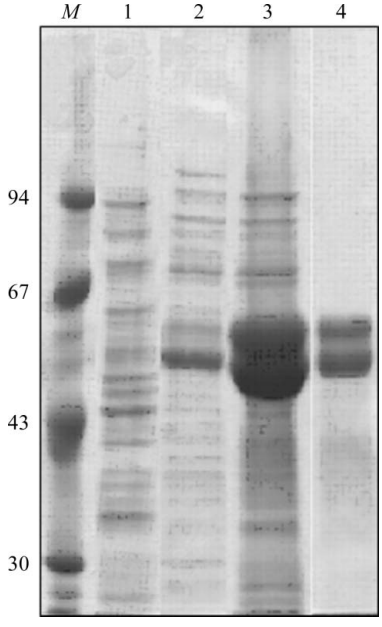


Figure 1
A Coomassie Brilliant Blue-stained 10% SDS–PAGE gel evaluation of protein purity during purification. Lane *M*, molecular-weight markers (kDa); lane 1, cell debris after sonication; lane 2, cell-free extract after sonication; lane 3, purified CBM22-1–GH10 E337A after Ni-affinity chromatography; lane 4, purified CBM22-1–GH10 E337A after gel-filtration chromatography. Although the protein is a highly pure single polypeptidic chain of 60 kDa, two bands were observed in the gel.

2.3. Crystallization

Crystallization conditions were screened by the hanging-drop vapour-phase diffusion method using an in-house modified version of the sparse-matrix method of Jancarik & Kim (1991) and the commercial screens Crystal Screen, Crystal Screen 2 and PEG/Ion Screen from Hampton Research (California, USA). Drops consisting of 1 µl of 20, 40 and 60 mg ml^{-1} CBM22-1–GH10 E337A and 1 µl reservoir solution were prepared and equilibrated against 1 ml reservoir solution at 292 K. Protein crystals were obtained at 292 K

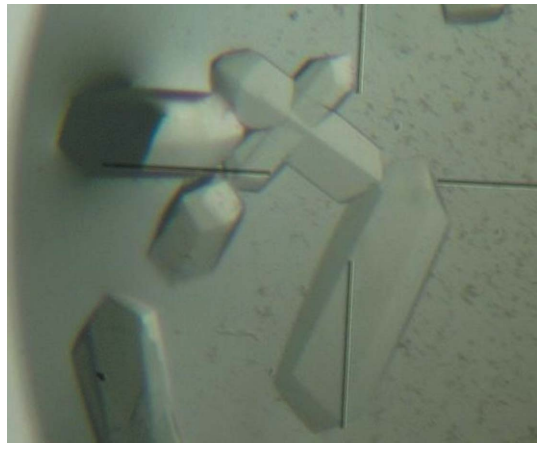


Figure 2
Crystals of CBM22-1–GH10 E337A obtained by hanging-drop vapour diffusion in the presence of 1 *M* sodium acetate, 0.1 *M* HEPES pH 7.5 and 0.05 *M* $CdSO_4$. The largest crystals are approximately 0.2 mm in their longest dimension.

Table 1
Matthews coefficient calculations.

No. of molecules per ASU	Matthews coefficient ($Å^3$ Da^{-1})	Solvent content (%)	Probability
2	4.74	74.07	0.01
3	3.16	61.10	0.17
4	2.37	48.14	0.68
5	1.90	35.17	0.13

within 4–6 d using the following conditions: 1 *M* sodium acetate, 0.1 *M* HEPES pH 7.5, 0.05 *M* $CdSO_4$ and 10 m*M* xylohexaose (Fig. 2). The crystals were flash-cooled in liquid nitrogen after soaking in a cryoprotectant solution [the crystallization mother liquor containing 30%(*v*/*v*) glycerol] for a few seconds.

2.4. Data collection and processing

Data were collected on beamline ID14-EH1 at the ESRF (Grenoble, France) using a Quantum-4 charge-coupled device detector (ADSC) with the crystal cooled at 100 K using a Cryostream (Oxford Cryosystems Ltd). All data sets were processed using the programs *MOSFLM* (Leslie, 1992) and *SCALA* (Kabsch, 1978) from the *CCP*4 suite (Collaborative Computational Project 4, Number 4, 1994). The crystal belonged to the trigonal space group $P3_121$ or $P3_221$ as indicated by *POINTLESS* (Evans, 2006). Plots of the acentric and centric moments and the cumulative intensity distribution from the *SCALA* output indicated that the crystal form was indeed $P3_121$ or $P3_221$ and not the the lower symmetry twin parental *P*32. An anisotropy analysis plot calculated by *TRUNCATE* (French & Wilson, 1978) showed the data to be isotropic. Calculation of the Matthews coefficient indicated the possibility of a range of oligomers from a dimer to a pentamer in the asymmetric unit (Table 1; Matthews, 1968). The *CCP*4 programs *ECALC*, *POLARRFN* and *RFCORR* (Collaborative Computational Project 4, Number 4, 1994) and *AMoRE* (Navaza, 2001) were used to resolve this ambiguity.

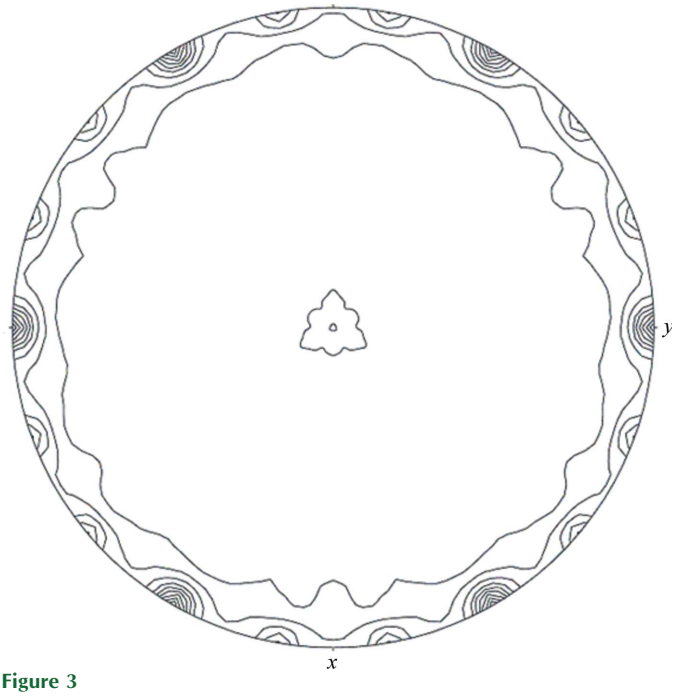


Figure 3
Plot of a 180° section of a self-rotation function from *POLARRFN*. The plot was calculated with a radius of 35 Å.

Table 2
X-ray crystallography data-collection statistics.

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

Data set	CBM22-1–GH10 E337A
X-ray source	ID14-EH1, ESRF
Wavelength (Å)	0.9340
Space group	$P3_221$
Unit-cell parameters	
$a = b$ (Å)	173.2
c (Å)	131.8
Resolution limits (Å)	65.8–2.0
No. of observations	2939405 (100685/232571)
No. of unique observations	153122 (5107/22116)
Multiplicity	19.2 (19.7/10.5)
Completeness (%)	99.9 (99.7/99.7)
$\langle I/\sigma(I)\rangle$	17.3 (51.2/1.7)
$R_{\rm merge}$†	12.0 (4.5/>100)
$R_{\rm p.i.m.}$‡	2.8 (1.1/44.1)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is the mean value of $I_i(hkl)$ for all *i* measurements. ‡ $R_{\rm p.i.m.} = \sum_{hkl}[1/(N-1)]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$ and is a measure of the quality of the data after averaging the multiple measurements.

ECALC converts *F* values to *E* values and gives beautifully sharp maps. *POLARRFN* was used to calculate a fast rotation function in polar angles at different radii (20, 25, 30, 35 Å). The self-rotation function (SRF) indeed has clear noncrystallographic symmetry (NCS) peaks on the 180° section (Fig. 3); the axes are in the *xy* plane at 20° and 40° from the *x* axis (and obviously repeated by the crystallographic symmetry) and it is possible that there is also one at 60° under the crystallographic twofold. However, there is no twofold peak along the *z* axis, which would appear to rule out a 222 tetramer. There are also clear peaks corresponding to rotations about the *z* axis at χ = 40°, 80° and 160° (and also obviously at 120° for the crystallographic threefold). These rotations probably relate molecules in different asymmetric units. Thus, every peak can be explained by having just a single NCS twofold. An NCS trimer or tetramer would give many more peaks. *AMoRE* was used to compute both the cross-rotation function (XRF) and SRF maps in Eulerian space and they were fed into *RFCORR*. *RFCORR* relates pairs of peaks in the XRF to the peaks found in the SRF. Only two of the XRF solutions were consistent with the SRF, leaving no doubt that there is just a single NCS twofold, indicating the presence of a dimer in the asymmetric unit with an ∼75% solvent content.

The programs *Phaser* (McCoy *et al.*, 2005) and *MOLREP* (Vagin & Teplyakov, 2000) were used in attempts to solve the native structure by molecular replacement using an ensemble of GH10 domains [PDB codes 1xyz (Dominguez *et al.*, 1995), 1nq6 (Canals *et al.*, 2003), 1r86 (Bar *et al.*, 2004) and 1n82 (Solomon *et al.*, 2007); sequence identities 30–35%] and an ensemble of CBM22-2 domains [PDB codes 1dyo (Charnock *et al.*, 2000), 1h6x and 1h6y (Xie *et al.*, 2001); sequence identity 22%]. Both programs found good solutions for two GH10 domains in space group $P3_221$, but not for the CBM22-1 domains. Automated model building using *ARP*/*wARP* (Cohen *et al.*, 2004) and the solutions from molecular replacement and the X-ray data located 300 amino acids with a sequence coverage of 28% and an estimated correctness of the model of approximately 65%. Completion of the structure is ongoing.

This work was supported in part by Fundação para a Ciência e a Tecnologia (Lisbon, Portugal) through grant POCTI/BIA-PRO/59118/2004 and the individual grants SFRH/BPD/20357/2004 (SN) and SFRH/BD/25439/2005 (BAP). The authors would like to thank

Dr Ana L. Carvalho for help with data collection and Dr Ian Tickle for help with *POLARRFN* and *RFCORR*.

References

Bar, M., Golan, G., Nechama, M., Zolotnitsky, G., Shoham, Y. & Shoham, G. (2004). *Acta Cryst.* D**60**, 545–549.

Bayer, E. A., Belaich, J. P., Shoham, Y. & Lamed, R. (2004). *Annu. Rev. Microbiol.* **58**, 521–554.

Canals, A., Vega, M. C., Gomis-Rüth, F. X., Díaz, M., Santamaría, R. I. & Coll, M. (2003). *Acta Cryst.* D**59**, 1447–1453.

Carvalho, A. L., Dias, F. M., Nagy, T., Prates, J. A., Proctor, M. R., Smith, N., Bayer, E. A., Davies, G. J., Ferreira, L. M., Romao, M. J., Fontes, C. M. & Gilbert, H. J. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 3089–3094.

Carvalho, A. L., Dias, F. M., Prates, J. A., Nagy, T., Gilbert, H. J., Davies, G. J., Ferreira, L. M., Romao, M. J. & Fontes, C. M. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 13809–13814.

Carvalho, A. L., Pires, V. M., Gloster, T. M., Turkenburg, J. P., Prates, J. A., Ferreira, L. M., Romao, M. J., Davies, G. J., Fontes, C. M. & Gilbert, H. J. (2005). *J. Mol. Biol.* **349**, 909–915.

Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M., Davies, G. J. & Fontes, C. M. (2000). *Biochemistry*, **39**, 5013–5021.

Cohen, S. X., Morris, R. J., Fernandez, F. J., Ben Jelloul, M., Kakaris, M., Parthasarathy, V., Lamzin, V. S., Kleywegt, G. J. & Perrakis, A. (2004). *Acta Cryst.* D**60**, 2222–2229.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Dominguez, R., Souchon, H., Spinelli, S., Dauter, Z., Wilson, K. S., Chauvaux, S., Beguin, P. & Alzari, P. M. (1995). *Nature Struct. Biol.* **2**, 569–576.

Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.

Fontes, C. M., Hall, J., Hirst, B. H., Hazlewood, G. P. & Gilbert, H. J. (1995). *Appl. Microbiol. Biotechnol.* **43**, 52–57.

Fontes, C. M., Hazlewood, G. P., Morag, E., Hall, J., Hirst, B. H. & Gilbert, H. J. (1995). *Biochem. J.* **307**, 151–158.

French, S. & Wilson, K. (1978). *Acta Cryst.* A**34**, 517–525.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Kabsch, W. (1978). *Acta Cryst.* A**34**, 827–828.

Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.

McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Najmudin, S., Guerreiro, C. I., Carvalho, A. L., Prates, J. A., Correia, M. A., Alves, V. D., Ferreira, L. M., Romao, M. J., Gilbert, H. J., Bolam, D. N. & Fontes, C. M. (2006). *J. Biol. Chem.* **281**, 8815–8828.

Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.

Prates, J. A., Tarbouriech, N., Charnock, S. J., Fontes, C. M., Ferreira, L. M. & Davies, G. J. (2001). *Structure*, **9**, 1183–1190.

Solomon, V., Teplitsky, A., Shulami, S., Zolotnitsky, G., Shoham, Y. & Shoham, G. (2007). *Acta Cryst.* D**63**, 845–859.

Tarbouriech, N., Prates, J. A. M., Fontes, C. M. G. A. & Davies, G. J. (2005). *Acta Cryst.* D**61**, 194–197.

Vagin, A. & Teplyakov, A. (2000). *Acta Cryst.* D**56**, 1622–1624.

Xie, H., Gilbert, H. J., Charnock, S. J., Davies, G. J., Williamson, M. P., Simpson, P. J., Raghothama, S., Fontes, C. M., Dias, F. M., Ferreira, L. M. & Bolam, D. N. (2001). *Biochemistry*, **40**, 9167–9176.