Overproduction, purification, crystallization and preliminary X-ray characterization of the C-terminal family 65 carbohydrate-binding module (CBM65B) of endoglucanase Cel5A from *Eubacterium cellulosolvens*

Immacolata Venditto,^a Arnaud Baslé,^b Ana S. Luís,^a Max J. Temple,^b Luís M. A. Ferreira,^a Carlos M. G. A. Fontes,^a Harry J. Gilbert^b and Shabir Najmudin^{a*}

^aCIISA – Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal, and ^bInstitute for Cell and Molecular Biosciences, The Medical School, Newcastle University, Newcastle-upon-Tyne NE2 4HH, England

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

Received 21 November 2012

Accepted 16 January 2013

The rumen anaerobic cellulolytic bacterium *Eubacterium cellulosolvens* produces a large range of cellulases and hemicellulases responsible for the efficient hydrolysis of plant cell wall polysaccharides. One of these enzymes, endoglucanase Cel5A, comprises a tandemly repeated carbohydrate-binding module (CBM65) fused to a glycoside hydrolase family 5 (Cel5A) catalytic domain, joined by flexible linker sequences. The second carbohydrate-binding module located at the C-terminus side of the endoglucanase (CBM65B) has been co-crystallized with either cellohexaose or xyloglucan heptasaccharide. The crystals belong to the hexagonal space group $P6_5$ and tetragonal space group $P4_32_12$, containing a single molecule in the asymmetric unit. The structures of CBM65B have been solved by molecular replacement.

1. Introduction

The enzymatic degradation of insoluble polysaccharides is one of the most important reactions occurring on earth. Enzymes that degrade plant cell wall carbohydrates generally contain a catalytic domain associated with one or more non-catalytic carbohydrate-binding modules (CBMs). CBMs bind a variety of polysaccharides playing an important role in potentiating the catalytic function of the appended enzymes (Boraston *et al.*, 2004). Ruminant anaerobic bacteria and fungi have developed a wide range of multi-modular cellulases and hemicellulases that act individually and in organized megadalton multi-enzyme complexes called cellulosomes. Cellulosomes are extremely complex and dynamic extracellular macromolecular nanomachines that actively degrade plant cell wall polysaccharides to soluble sugars (for reviews, see Bayer *et al.*, 2004; Fontes & Gilbert, 2010).

The rumen microbe *Eubacterium cellulosolvens* produces a large consortia of cellulases and hemicellulases responsible for plant cell wall degradation (Flint & Bayer, 2008). The *E. cellulosolvens* endoglucanase Cel5A (*EcCel5A*) is a 1148 amino-acid protein comprising a tandem repeat of a carbohydrate-binding module (CBM65) linked to a glycoside family 5 catalytic module (GH5) separated by a PT-rich linker region (PT). Thus, the architectural arrangement of Cel5A is: N-CBM65A_GH5-1_PT1-CBM65B_GH5-2_PT2-C, flanked by an N-terminal signal peptide (N) and a C-terminal tail of unknown function (C). The two CBM65s show a sequence identity of 73% to each other. They have no structural homologues, but have related sequences to modules found in other endoglucanases (Fig. 1). *EcCel5A* has been shown to have activities against a variety of cellulosic polysaccharides including carboxymethyl cellulose, lichenan, acid-swollen cellulose and oat spelt xylan (Yoda *et al.*, 2005). We have already characterized the structure of CBM65A (previously labelled as CBM-AL1, Luís *et al.*, 2011) and shown that it binds xyloglucans more strongly than cellulosic ligands (Luís *et al.*, 2013). Thus, CBM65A and CBM65B have been assigned as founding members of a new CBM family 65 in the CAZy database (Cantarel *et al.*, 2009). In order to gain insights into the structural properties that

modulate ligand recognition, we aim to determine the crystal structure of *E. cellulosolvens* CBM65B in the presence of cellohexaose (C6) and xyloglucan heptasaccharide (Glc4Xyl3, XXXG). In the present communication, we describe the overproduction, purification, crystallization and preliminary X-ray analysis of the *Ec*CBM5A C-terminal CBM65B module co-crystallized with the ligands cellohexaose or xyloglucan heptasaccharide.

2. Materials and methods

2.1. Protein production and purification

*Ec*CBM5A is a modular enzyme containing an N-terminal CBM followed by a GH5 catalytic domain. The two domains are duplicated in tandem and the enzyme contains an additional C-terminal domain of unknown function. The gene encoding the C-terminal CBM of *Ec*CBM5A (CBM65B, residues 581 to 713) was synthesized (NZYTech Ltd, Portugal) with codon usage optimized for expression in *Escherichia coli*. The synthesized gene contained engineered *NheI* and *XhoI* recognition sequences, at the 5' and 3' ends, respectively, that were used for subcloning into pET28a (Novagen) initially and subsequently into pET21a (Novagen), generating pAL2_28a or pAL2_21a, respectively, which encode CBM65B. Thus, the recombinant CBM65B derivatives produced in this work contained either an N- (pAL2_28a) or C-terminal His₆ tag (pAL2_21a). The N-terminal His₆ tag (pAL2_28a) contains an extra 23 amino-acid residues at the N-terminus (MGSSHHHHHHSSGLVPRGSHMAS), whereas the C-terminal His₆ tag (pAL2_21a) contains just three extra amino-acid residues at the N-terminus (MAS) and eight extra at the C-terminus (LEHHHHHH). *E. coli* Tuner DE3 cells harbouring pAL2_28a or *E. coli* BL21 DE3 cells harbouring pAL2_21a were cultured in Luria-Bertani broth at 310 K to mid-exponential phase (A_{600nm} 0.6) and recombinant protein overproduction was induced by the addition of 0.2 mM isopropyl β -D-1-thiogalactopyranoside and incubation for a further 16 h at 292 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 CBM65B was buffer-exchanged into 50 mM HEPES–

HCl 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. Purified CBM65B was concentrated using an Amicon 10 kDa molecular-mass centrifugal concentrator and washed three times with 1.0 mM CaCl₂ (Fig. 2).

2.2. Crystallization

The crystallization conditions were screened by the hanging-drop vapour-diffusion method using the commercial Crystal Screen, Crystal Screen 2 and PEG/Ion from Hampton Research (California, USA), and the Clear Strategy Screens MD-1 and MD-2 from Molecular Dimensions (UK). Drops of 1 μ l of 20, 40, 60 and 65 mg ml⁻¹ N-terminal His₆-tagged CBM65B (N-His-CBM65B) and 1 μ l of reservoir solution were manually prepared at 292 K. Crystals (maximum dimension ~50 μ m) grew within a week in the following two conditions: (i) 1.0 M KH₂PO₄ and (ii) 0.2 M ammonium tartrate dibasic, pH 7.0, 20% (w/v) PEG 3350. Optimization screens were set up around both these conditions. Hexagonal-plated crystals (maximum dimension ~100 μ m) grew over a period of a few days in 0.5–1.0 M KH₂PO₄ (Fig. 3a). The crystals were cryocooled in liquid nitrogen after soaking in the cryoprotectant [30% (v/v) glycerol added to the crystallization buffer] for a few seconds. Crystals obtained in the first condition gave poor, unsolvable diffraction data and those from the second condition subsequently turned out to be salt. Thus, C-terminal His₆-tagged recombinant CBM65B (C-His-CBM65B) was produced, additionally to the N-terminal His₆-tagged CBM65B, and attempts were made to co-crystallize both with 10 mM of either 1,4- β -D-cellohexaose (C6) or xyloglucan heptasaccharide (Glc4Xyl3, XXXG). The new 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 Clear Strategy Screens I and II, MIDAS and JCSG-*plus* HT96 screens (Molecular Dimensions, UK) using the robotic nanodrop dispensing system Oryx8 (Douglas Instruments). Two drops of 0.7 μ l 30 mg ml⁻¹ of C-His-CBM65B (one with 10 mM C6 and the other with 10 mM XXXG) and 0.7 μ l of

Figure 1
Sequence comparison of CBM65 family members. *Ec*CBM65A and *Ec*CBM65B share ~30% of similarity with the CBMs of *Cellulosilyticum ruminicola* (CrCBM65) and *Clostridium lentocellum* (C1CBM65). The aromatic amino acids conserved between all CBM65s are shaded in yellow. Amino acids that are conserved in both *Ec*CBM65 sequences are shaded in turquoise. Some endoglucanases of *Clostridium acetobutylicum*, *Clostridium cellulovorans* and *Ruminococcus albus* also have putative CBM65s, but are not included here for the sake of clarity. The alignment was made using *ClustalW* (Thompson *et al.*, 1994).

Figure 2
A Coomassie Brilliant Blue-stained 14% PAGE gel evaluation of protein purity. Lane 1: molecular-mass markers (kDa); lane 2: C-His-CBM65B.

reservoir solution were prepared at 292 K. Crystals were seen in six different conditions. For the XXXG complex crystals were seen in 0.1 M sodium citrate tribasic dihydrate pH 5.6, 30%(w/v) PEG 4000 (Fig. 3b), 2.0 M ammonium sulfate (Fig. 3c), 0.2 M potassium sodium tartrate tetrahydrate, 0.1 M sodium citrate tribasic dihydrate pH 5.6, 2 M ammonium sulfate (Fig. 3d) and 0.01 M zinc sulfate heptahydrate, 0.1 M MES monohydrate pH 6.5, 25% PEG monomethyl ether 550 (Fig. 3e); and for the C6 complex crystals were seen in 0.2 M ammonium acetate, 0.1 M sodium acetate trihydrate pH 4.6, 30%(w/v) PEG 400 (Fig. 3f) and 2.0 M ammonium sulfate (Fig. 3g). Crystals were also obtained in 0.2 M ammonium acetate, 0.1 M sodium citrate tribasic dihydrate pH 5.6, 30%(w/v) PEG 4000 for the N-His-CBM65B at a protein concentration of 1.4 mM and XXXG concentration of 14 mM. Though CBM65A and CBM65B show just a 27% sequence identity difference, this is enough to change their crystallization behaviour. Unlike CBM65A, we were unable to obtain suitable diffracting crystals with just N-terminal His₆-tagged CBM65B, but had to use C-His-CBM65B or add suitable ligands.

2.3. Data collection and processing

Initial data from crystals obtained from the N-terminal His₆-tagged CBM65B were collected on beamline PROXIMA-1 at SOLEIL (Orsay, France) using a Quantum 315r charge-coupled device detector (ADSC, USA) with the crystal cooled at 100 K using a Cryostream (Oxford Cryosystems Ltd). Data for the CBM65B co-crystallized with either C6 or XXXG were collected on beamline IO2 at the Diamond Light Source (Harwell, UK) using a PILATUS 6M detector (Dectris, Baden, Switzerland) with the crystal cooled at 100 K using a Cryostream (Oxford Cryosystems Ltd). All data sets were processed using the programs *iMOSFLM* (Battye *et al.*, 2011)

and *SCALA* (Evans, 2006) from the *CCP4* suite (Winn *et al.*, 2011) or *XDS* (Kabsch, 2010). Data-collection statistics are given in Table 1. Initial attempts to solve the apo-form of the N-His-CBM65B resulted in crystals that gave poor, unsolvable diffraction data or turned out to be salt. The best data set was to a resolution of 3.5 Å in the hexagonal space group (*P*6₂) with a very long *c* axis (over 500 Å; see Table 1). Thus, C-His-CBM65B was produced, additionally to the N-His-CBM65B. Crystals of C-His-CBM65B grown with 10 mM XXXG in 0.1 M sodium citrate tribasic dihydrate pH 5.6, 2 M ammonium sulfate and 0.01 M zinc sulfate heptahydrate, 0.1 M MES monohydrate, pH 6.5, 25% PEG monomethyl ether 550 gave very poor or no diffraction. However, two types of diffracting crystals were obtained when C-His-CBM65B was co-crystallized with either cellohexaose or xyloglucan heptasaccharide in the other conditions. Crystals with the hexagonal form with point group *P*6 were obtained with both ligands and diffracted to very high resolution, up to 1.42 Å. A second form of crystals in the tetragonal *P*4₂ space group was obtained when either N-His-CBM65B or C-His-CBM65B were co-crystallized with XXXG. These crystals diffracted to medium resolution, up to 2.6 Å. The Matthews coefficient ($V_M = 2.61 \text{ Å}^3 \text{ Da}^{-1}$ for the *P*6 form and $V_M = 3.58 \text{ Å}^3 \text{ Da}^{-1}$ for the *P*4₂ form) indicated the presence of one molecule in the asymmetric unit for both and a solvent content of 50 and 65%, respectively (Matthews, 1968). Initial phasing for structure solution was obtained using the molecular-replacement program *Phaser* (McCoy *et al.*, 2007). The atomic coordinates of native CBM65A (PDB code 4afm, Luis *et al.*, 2013) were used as a search model. For the highest-resolution data (1.42 Å) obtained for the C-His-CBM65B co-crystallized with C6, testing all six alternate hexagonal space groups, a successful solution was obtained in space group *P*6₅ with a TFZ score of 20.8 and LLG of 318. For the highest-resolution data (*i.e.* including all data to 2.5 Å)

Figure 3

Crystals of CBM65B obtained by hanging/sitting-drop vapour diffusion. (a) The apo-form in the presence of 0.5–1.0 M KH_2PO_4 (data set N-His-CBM65B); (b) the XXXG complex in 0.1 M sodium citrate tribasic dihydrate pH 5.6, 30%(w/v) PEG 4000 (data set x1); (c) 2.0 M ammonium sulfate (data set x9); (d) 0.2 M potassium sodium tartrate tetrahydrate, 0.1 M sodium citrate tribasic dihydrate pH 5.6, 2 M ammonium sulfate (no diffraction); and (e) 0.01 M zinc sulfate heptahydrate, 0.1 M MES monohydrate pH 6.5, 25% PEG ether 550 (poor diffraction); (f) the C6 complex in 0.2 M ammonium acetate, 0.1 M sodium acetate trihydrate, pH 4.6, 30%(w/v) PEG 400 (data set x4); and (g) 2.0 M ammonium sulfate (data set x8). The N-His-CBM65B with XXXG was crystallized in 0.2 M ammonium acetate, 0.1 M sodium citrate tribasic dihydrate pH 5.6, 30%(w/v) PEG 4000 (data set N-His-CBM65B-XXXG).

Table 1
Data-collection statistics.

Values in parentheses are for the lowest/highest-resolution shells, with the ranges being 49.04–11.09/3.70–3.51, 41.82–5.06/1.69–1.6, 58.96–9.49/3.16–3.0, 36.75–4.49/1.50–1.42, 41.70–5.06/1.69–1.6 and 2.48–2.35 for each data set, respectively.

Data set	N-His-CBM65B	Cellohexaose x4	XXXG x1	Cellohexaose x8	XXXG x9	N-His-CBM65B-XXXG
Beamline	PROXIMA-1 SOLEIL	Diamond IO2	Diamond IO2	Diamond IO2	Diamond IO2	Diamond IO4
Space group	<i>P</i> 622	<i>P</i> 6 ₅	<i>P</i> 4 ₃ 2 ₁ 2	<i>P</i> 6 ₅	<i>P</i> 6 ₅	<i>P</i> 4 ₃ 2 ₁ 2
Wavelength (Å)	0.9793	0.9795	0.9795	0.9795	0.9795	0.9795
Rotation per frame (°)	1.0	0.3	0.3	0.3	0.3	0.5
Total rotation (°)	165	96	150	105	150	90
Exposure per frame (s)	0.5	0.2	0.2	0.3	0.3	0.5
Mosaicity (°)	3.0	0.4	1.9	0.5	0.4	0.8
Unit-cell parameters						
<i>a</i> = <i>b</i> (Å)	62.6	83.64	58.96	83.57	83.41	57.94
<i>c</i> (Å)	573.7	36.76	117.1	36.75	36.76	116.77
Resolution limits (Å)	49.04–3.51	41.82–1.6	58.96–3.0	36.75–1.42	41.7–1.6	58.38–2.35
No. of observations	20901 (2072/1.231)	93062 (3108/13103)	16385 (552/2486)	146829 (4874/21492)	142256 (4632/19967)	51328 (6897)
No. of unique observations	5985 (337/585)	19582 (667/2835)	4315 (116/606)	27755 (929/4003)	19500 (663/2831)	7993 (1126)
Multiplicity	3.5 (5.5/2.1)	4.8 (4.7/4.6)	3.8 (3.3/4.1)	5.3 (5.2/5.4)	7.3 (7.0/7.1)	6.4 (6.1)
Completeness (%)	62.6 (90.4/45.3)	99.9 (99.9/99.9)	96.5 (91.1/97.4)	99.8 (99.3/99.9)	100 (99.9/99.2)	100.0 (100.0)
<i>I</i> / <i>σ</i> (<i>I</i>)	4.7 (11.1/0.7)	8.7 (26.4/1.5)	3.5 (5.6/1.3)	11.9 (40.1/1.8)	9.4 (22.9/2.2)	14.6 (3.1)
<i>R</i> _{merge} [†] (%)	23.4 (17.7/172)	13.4 (4.0/162.0)	22.8 (11.8/94.4)	6.1 (1.9/69.2)	15.6 (4.2/198)	7.5 (51.9)
<i>R</i> _{p.i.m.} [‡] (%)	15.3 (9.2/157)	6.8 (2.0/84.9)	13.3 (7.0/51.7)	2.8 (0.8/31.4)	6.2 (1.7/79.9)	3.2 (22.5)

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

obtained for the C-His-CBM65B co-crystallized with XXXG, searching in all alternate tetragonal space groups, a successful solution was obtained in space group *P*4₃2₁2 with a TFZ score of 19.6 and LLG of 273. Structure completion and analysis are ongoing. This work was supported in part by Fundação para a Ciência e a Tecnologia (Lisbon, Portugal) through grant Nos. PTDC/BIA-PRO/103980/2008 and PTDC/QUI-BIO/100359/2008, and by the European Union Seventh Framework Programme (FP7 2007–2013) under the WallTraC project (grant agreement No. 263916). 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. The authors would like to thank Dr Ana Luisa Carvalho, Dr Victor D. Alves and Rita Cardoso, and Drs Marco Mazzorana at Diamond and Andrew Thompson at SOLEIL for their help with data collection and Cecília Bonifacio for help with the 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.

Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. (2004). *Biochem. J.* **382**, 769–781.
Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.
Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
Flint, H. J. & Bayer, E. A. (2008). *Ann. NY Acad. Sci.* **1125**, 280–288.
Fontes, C. M. & Gilbert, H. J. (2010). *Annu. Rev. Biochem.* **79**, 655–681.
Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
Luís, A. S., Alves, V. D., Romão, M. J., Prates, J. A. M., Fontes, C. M. G. A. & Najmudin, S. (2011). *Acta Cryst.* **F67**, 491–493.
Luís, A. S., Venditto, I., Prates, J. A. M., Ferreira, L. M. A., Temple, M. J., Rogowski, A., Baslé, A., Xue, J., Knox, J. P., Najmudin S., Fontes C. M. G. A. & Gilbert, H. J. (2013). *J. Biol. Chem.*. In the press. DOI 10.1074/jbc.M112.432781.
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.
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., Romão, M. J., Gilbert, H. J., Bolam, D. N. & Fontes, C. M. (2006). *J. Biol. Chem.* **281**, 8815–8828.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
Yoda, K., Toyoda, A., Mukoyama, Y., Nakamura, Y. & Minato, H. (2005). *J. Appl. Environ. Microbiol.* **71**, 5787–5793.