Purification and crystallographic studies of a putative carbohydrate-binding module from the *Ruminococcus flavefaciens* FD-1 endoglucanase Cel5A

Ana José Pires,^a Teresa Ribeiro,^a Andrew Thompson,^b Immacolata Venditto,^a Vânia O. Fernandes,^{a,c} Pedro Bule,^a Helena Santos,^a Victor D. Alves,^a Virginia Pires,^a Luis M. A. Ferreira,^a Carlos M. G. A. Fontes^a and Shabir Najmudin^{a*}

Received 22 December 2014

Accepted 20 May 2015

Edited by B. Hazes, University of Alberta, Canada

Keywords: *Ruminococcus flavefaciens*; carbohydrate-binding module; carbohydrate-active enzymes; glycoside hydrolase; cellulosomes.

^aCIISA, Faculdade de Medicina Veterinária, Universidade de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisbon, Portugal, ^bSynchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin, 91192 Gif-sur-Yvette, France, and ^cNZYTech Genes and Enzymes, Estrada do Paço do Lumiar, Campus do Lumiar, Edifício E, 1649-038 Lisbon, Portugal. *Correspondence e-mail: shabir@fmv.ulisboa.pt

Ruminant herbivores meet their carbon and energy requirements from a symbiotic relationship with cellulosome-producing anaerobic bacteria that efficiently degrade plant cell-wall polysaccharides. The assembly of carbohydrate-active enzymes (CAZymes) into cellulosomes enhances protein stability and enzyme synergistic interactions. Cellulosomes comprise diverse CAZymes displaying a modular architecture in which a catalytic domain is connected, *via* linker sequences, to one or more noncatalytic carbohydrate-binding modules (CBMs). CBMs direct the appended catalytic modules to their target substrates, thus facilitating catalysis. The genome of the ruminal cellulolytic bacterium *Ruminococcus flavefaciens* strain FD-1 contains over 200 modular proteins containing the cellulosomal signature dockerin module. One of these is an endoglucanase Cel5A comprising two family 5 glycoside hydrolase catalytic modules (GH5) flanking an unclassified CBM (termed CBM-Rf2) and a C-terminal dockerin. This novel CBM-Rf2 has been purified and crystallized, and data from cacodylate-derivative crystals were processed to 1.02 and 1.29 Å resolution. The crystals belonged to the orthorhombic space group $P2_12_12_1$. The CBM-Rf2 structure was solved by a single-wavelength anomalous dispersion experiment at the As edge.

1. Introduction

Plant cell-wall polysaccharides represent one of the most abundant sources of biomass and renewable energy on earth (Burton & Fincher, 2014). This complex macromolecular structure is primarily composed of cellulose and hemicellulose, which are often insoluble and highly recalcitrant to hydrolysis (Somerville *et al.*, 2004; Cosgrove, 2005). Anaerobic microorganisms, such as those found in the rumens of polygastric animals, have evolved a supramolecular complex, termed the cellulosome, to efficiently degrade plant structural carbohydrates (Bayer *et al.*, 1998; Gilbert, 2007; Fontes & Gilbert, 2010). The cellulosome comprises a diverse conglomerate of modular carbohydrate-active enzymes (CAZymes) which contain one or more catalytic domains connected to one or more noncatalytic modules. The most common noncatalytic modules appended to CAZymes are carbohydrate-binding modules (CBMs), which target enzymes to their carbohydrate substrates (Boraston *et al.*, 2004). In addition, cellulosomal enzymes contain a C-terminal non-

catalytic module, termed the dockerin, which binds to iterated cohesin domains located in modular scaffoldins. The genome of the rumen cellulolytic bacterium *Ruminococcus flavefaciens* strain FD-1 encodes over 200 dockerin-containing proteins, comprising glycoside hydrolases, carbohydrate esterases and pectate lyases, which contain several modules of unknown function (Berg Miller *et al.*, 2009). One such protein is a bifunctional endoglucanase Cel5A of 947 amino-acid residues termed *RfCel5A*. *RfCel5A* comprises two family 5 glycoside hydrolase catalytic modules (GH5) flanking an unclassified CBM (termed CBM-Rf2) and a C-terminal dockerin. *BLAST* sequence analysis shows that these GH5 modules have over 50% sequence identity to the endoglucanase Cel5A from *R. flavefaciens* 17 and over 30% sequence identity to other GH5 members (Altschul *et al.*, 1990). We have recently experimentally demonstrated their glucanase activity (Pires *et al.*, unpublished work). There are no structural homologues of CBM-Rf2. *BLAST* analysis (Altschul *et al.*, 1990) shows that CBM-Rf2 shares an amino-acid sequence identity of greater than 25% with at least 15 other proteins of unknown function, all of which are from *R. flavefaciens*. In the present communication, we describe the crystallization and crystallographic study of this novel CBM (CBM-Rf2) identified in the endoglucanase Cel5A from the *R. flavefaciens* FD-1 cellulosome.

2. Materials and methods

2.1. Macromolecule production

The gene encoding the putative CBM-Rf2 (residues 495–621 of *RfCel5A*) was synthesized (NZYTech Ltd, Portugal) with codon usage optimized for expression in *Escherichia coli*. The synthesized gene, containing engineered *NcoI* and *XhoI* restriction sites at the 5' and 3' ends, respectively, derived from the primers given in Table 1, was subsequently subcloned into the pET-28a expression vector (Novagen), generating p*Rf2*,

Figure 1

A Coomassie Brilliant Blue-stained 16% SDS-PAGE evaluation of protein purity. Lane 1, molecular-mass markers (labelled in kDa; NZYTech Ltd, Portugal); lane 2, native CBM-Rf2 (labelled in kDa).

Table 1

Macromolecule-production information.

Source organism	<i>R. flavefaciens</i> FD-1
DNA source	cazy236510 3252
Forward primer	CACACACCATGGGAGACGGTTACACAATCAAG
Reverse primer	CACACACTCGAGCTTGAGAACTACAGAGTC
Cloning vector	pET-28a
Expression vector	pET-28a
Expression host	<i>E. coli</i> BL21
Complete amino-acid sequence of the construct produced†	MGAEEDTAILYPFTISGNDNRNGNFTINFKGTNP-STNNGCIGYSYNGDWKIEWEGSCDGNGLVV-EVPMSKIPAGVTSGEIQIWWHSGDLKMTDYKA-LEHHHHHH

† The cloning artefacts are underlined.

Table 2

Crystallization.

Method	Sitting-drop vapour diffusion
Plate type	MRC 96-well 2-drop sitting-drop crystallization plate (Molecular Dimensions, UK)
Temperature (K)	292
Protein concentration (mg ml ⁻¹)	20
Buffer composition of protein solution	Deionized water (Sigma), 0.5 mM CaCl ₂
Composition of reservoir solution	0.2 M sodium acetate, 0.1 M cacodylic acid pH 6.5, 30% (w/v) polyethylene glycol 8000
Volume and ratio of drop	0.7:0.7 or 1.0:1.0 µl protein solution: reservoir solution
Volume of reservoir (µl)	50

which also encodes a C-terminal His₆ tag (LEHHHHHH; Table 1). *E. coli* BL21 cells harbouring p*Rf2* were cultured in Luria–Bertani broth at 310 K to mid-exponential phase (OD_{600 nm} = 0.6) and recombinant protein overproduction was induced by adding isopropyl β-D-1-thiogalactopyranoside to a final concentration of 1 mM with 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.*, 2010). Purified CBM-Rf2 was buffer-exchanged into 50 mM Na HEPES buffer pH 7.5 containing 200 mM NaCl and 5 mM CaCl₂ and subjected to gel filtration using a HiLoad 16/60 Superdex 75 column (GE Healthcare) at a flow rate of 1 ml min⁻¹. Purified CBM-Rf2 was concentrated using an Amicon 10 kDa molecular-mass centrifugal concentrator. A full buffer replacement was carried out to avoid protein precipitation and the protein was finally washed with 0.5 mM CaCl₂ buffer. Recombinant CBM-Rf2, including the C-terminal His₆ tag (LEHHHHHH), has an approximate molecular mass of 12 kDa. The protein concentration was estimated using a molar extinction coefficient (ε) of 30 940 M⁻¹ cm⁻¹ at 280 nm with a Thermo Scientific NanoDrop 2000c and was verified by SDS-PAGE (Fig. 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, PEG/Ion 2 (Hampton Research, California, USA) and the JCSG screen (Newman *et*

Table 3
Data collection and processing.

Values in parentheses are for the outer shell.

Data set	As edge (x6)	High resolution (x5)
Beamline	PROXIMA 1, SOLEIL	PROXIMA 1, SOLEIL
Space group	$P2_12_12_1$	$P2_12_12_1$
Wavelength (Å)	1.04408	0.82656
Unit-cell parameters (Å)	$a = 43.39, b = 45.21, c = 49.56$	$a = 43.27, b = 45.02, c = 49.45$
Resolution limits (Å)	33.40–1.29 (1.31–1.29)	45.02–1.02 (1.04–1.02)
No. of observations	176443 (6683)	263876 (10674)
No. of unique observations	25196 (1180)	42277 (1959)
Multiplicity	7.0 (5.7)	6.2 (5.4)
Completeness (%)	99.5 (93.6)	99.7 (94.6)
$\langle I/\sigma(I) \rangle$	22.1 (8.4)	16.2 (1.5)
$CC_{1/2}^\dagger$	0.997 (0.978)	0.999 (0.706)
$R_{\text{merge}}^\ddagger$	0.057 (0.152)	0.05 (0.941)
$R_{\text{p.i.m.}}^\S$	0.023 (0.069)	0.021 (0.427)

† $CC_{1/2}$ is the correlation between intensities from random half data sets (Diederichs & Karplus, 2013). ‡ $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, 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. § $R_{\text{p.i.m.}} = \sum_{hkl} \{1/[N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $\langle I(hkl) \rangle$ is the average of symmetry-related observations of a unique reflection.

al., 2005) and an in-house 80-factorial screen using an Oryx8 robotic nanodrop dispensing system (Douglas Instruments) (Table 2). Two drops per well containing 50 μ l reservoir solution were prepared: one consisting of 0.7 μ l 20 mg ml $^{-1}$ CBM-Rf2 and 0.7 μ l reservoir solution and one consisting of 1 μ l 20 mg ml $^{-1}$ CBM-Rf2 and 1 μ l reservoir solution. Crystals grew in the following two conditions: 0.2 M sodium acetate, 0.1 M cacodylic acid pH 6.5, 30% (w/v) polyethylene glycol 8000 (Fig. 2) and 0.03 M citric acid, 0.07 M bis-tris propane pH 7.6, 20% (w/v) polyethylene glycol 3350 (which gave poor diffraction). All crystals grew to maximum size within a week. The crystals were cryocooled in liquid nitrogen after soaking in cryoprotectant [either 30% (v/v) glycerol added to the crystallization buffer or Paratone-N oil] for a few seconds.

2.3. Data collection and processing

Data were collected on the PROXIMA 1 beamline at SOLEIL, Orsay, France using CBM-Rf2 crystals (Fig. 2, Table 2) cooled to 100 K using a Cryostream (Oxford Cryosystems) and a Pilatus 6M detector (Dectris). The presence of As in the crystals was confirmed using X-ray fluorescence, and an energy scan of the As absorption edge was performed with expected values of the anomalous scattering factors calculated using *CHOOCH* (Evans & Pettifer, 2001). The presence of bound As was confirmed by examining the presence of anomalous signal as shown by the anomalous correlation and SigAno reported by *XDS*. After indexing and strategy calculation, in order to solve the crystallographic phase problem, 100° of data were collected with a $\Delta\varphi$ of 0.1° and an exposure of 0.1 s interleaved every 10° of rotation with a further 100° using inverse-beam geometry (data collections starting at φ and $\varphi + 180^\circ$) at the As absorption-edge peak (an X-ray energy of 11.875 keV, $f' = -8.08$ and $f'' = 8.03$) in a single-wavelength anomalous diffraction experiment using the first

crystal (x6 in Table 3, Fig. 2b) as described previously (Venditto *et al.*, 2015). In order to mitigate the effects of radiation damage to the crystal during the phasing experiment, three measurements were taken: collection of data in an interleaved inverse-beam geometry as described above, reduction of the incident X-ray dose by attenuation of the beam to 10% of its nominal value and collection of data at a lower resolution than the diffraction limit of the crystal. A second, higher resolution (1.02 Å) data set comprising 180° of data was collected from a fresh crystal (x5 in Table 3, Fig. 2a), a shorter X-ray wavelength being chosen for data collection in order to increase the resolution attainable at the minimum sample-to-detector distance available at the beamline. All data sets were processed using *XDS* (Kabsch, 2010) via the command-line interface *xdsme* (<https://code.google.com/p/xdsme/>) and *AIMLESS* (Evans, 2006) from the *CCP4* suite (Collaborative Computational Project, Number 4, 1994; Winn *et al.*, 2011). Data-collection statistics (as reported by

(a)

(b)

Figure 2
Crystals of CBM-Rf2 used in data collection: (a) x5 and (b) x6. Glycerol was used as a cryoprotectant for both these crystals. The black scale bar represents 0.1 mm.

Figure 3
Representative diffraction pattern of the high-resolution (x5) crystal with resolution rings shown as dashed circles.

AIMLESS) are given in Table 3, with anomalous mates counted separately.

3. Results and discussion

Crystals with glycerol as a cryoprotectant gave much better diffraction data than those with Paratone-N oil. All of the diffracting CBM-Rf2 crystals belonged to the orthorhombic space group $P2_12_12_1$, with a single molecule in the asymmetric unit, a solvent content of $\sim 34\%$ and a Matthews coefficient of $\sim 1.85 \text{ \AA}^3 \text{ Da}^{-1}$ (Matthews, 1968). The CBM-Rf2 structure was determined using a cacodylate derivative by a single-wavelength anomalous dispersion experiment using the *SHELX* suite (Sheldrick, 2008) via the *HKL2MAP* graphical interface (Pape & Schneider, 2004). Combined normal and inverse-beam data at the peak wavelength corresponding to the As absorption edge were used to determine heavy-atom sites using *SHELXD*. A single strong site and four other minor sites were suggested. These five sites could be associated, after examination of the anomalous difference Fourier map, as coming from a single cacodylate molecule (As) and the S atoms found in the two Cys and two Met residues, respectively. These sites were then used to calculate phases using *Phaser* in SAD mode (McCoy *et al.*, 2004, 2007) in the *CCP4* suite (Winn *et al.*, 2011) followed by density improvement using *Parrot* (Zhang *et al.*, 1997). The quality of the electron-density maps was excellent. Automatic model building using *ARP/wARP* (Langer *et al.*, 2008) interspersed with *REFMAC5* (Murshudov *et al.*, 2011) placed 96 residues out of 106 with an *R* value of 20.5%. The second data set,

processed to 1.02 Å resolution, is currently being used for refinement of this model.

Acknowledgements

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. Part of this work was supported by the European Union Seventh Framework Programme (FP7 2007-2013) under the WallTraC project (grant agreement No. 263916) and under BioStruct-X (grant agreement No. 283570, proposal Nos. 4399 and 7093). 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 SOLEIL (beamline PROXIMA 1, proposal 20131210) for the provision of synchrotron-radiation facilities. We would also like to thank Cecília Bonifácio for help with crystallization and Dr Teresa Santos-Silva and Marino Santos for help with data collection.

References

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.
- Bayer, E. A., Shimon, L. J. W., Shoham, Y. & Lamed, R. (1998). *J. Struct. Biol.* **124**, 221–234.
- Berg Miller, M. E. *et al.* (2009). *PLoS One*, **4**, e6650.
- Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. (2004). *Biochem. J.* **382**, 769–781.
- Burton, R. & Fincher, G. B. (2014). *Curr. Opin. Biotechnol.* **26**, 79–84.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Cosgrove, D. J. (2005). *Nature Rev. Mol. Cell Biol.* **6**, 850–861.
- Diederichs, K. & Karplus, P. A. (2013). *Acta Cryst.* **D69**, 1215–1222.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Evans, G. & Pettifer, R. F. (2001). *J. Appl. Cryst.* **34**, 82–86.
- Fontes, C. M. G. A. & Gilbert, H. J. (2010). *Annu. Rev. Biochem.* **79**, 655–681.
- Gilbert, H. J. (2007). *Mol. Microbiol.* **63**, 1568–1576.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
- 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.
- McCoy, A. J., Storoni, L. C. & Read, R. J. (2004). *Acta Cryst.* **D60**, 1220–1228.
- 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., Pinheiro, B. A., Prates, J. A. M., Gilbert, H. J., Romão, M. J. & Fontes, C. M. G. A. (2010). *J. Struct. Biol.* **172**, 353–362.
- Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I. & Perrakis, A. (2005). *Acta Cryst.* **D61**, 1426–1431.
- Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., Osborne, E., Paredes, A., Persson, S., Raab, T., Vorwerk, S. & Youngs, H. (2004). *Science*, **306**, 2206–2211.
- Venditto, I., Goyal, A., Thompson, A., Ferreira, L. M. A., Fontes, C. M. G. A. & Najmudin, S. (2015). *Acta Cryst.* **F71**, 45–48.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Zhang, K. Y. J., Cowtan, K. & Main, P. (1997). *Methods Enzymol.* **277**, 53–64.