Crystallization and preliminary crystallographic studies of a novel noncatalytic carbohydrate-binding module from the *Ruminococcus flavefaciens* cellulosome

Immacolata Venditto,^a Arun Goyal,^b Andrew Thompson,^c Luis M. A. Ferreira,^a 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, ^bDepartment of Biotechnology, Indian Institute of Technology Guwahati, Assam, India, and ^cSynchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin, 91192 Gif-sur-Yvette, France

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

Received 24 September 2014
Accepted 22 November 2014

Microbial degradation of the plant cell wall is a fundamental biological process with considerable industrial importance. Hydrolysis of recalcitrant polysaccharides is orchestrated by a large repertoire of carbohydrate-active enzymes that display 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 potentiating catalysis. The genome of the most abundant ruminal cellulolytic bacterium, *Ruminococcus flavefaciens* strain FD-1, provides an opportunity to discover novel cellulosomal proteins involved in plant cell-wall deconstruction. It encodes a modular protein comprising a glycoside hydrolase family 9 catalytic module (GH9) linked to two unclassified tandemly repeated CBMs (termed CBM-Rf6A and CBM-Rf6B) and a C-terminal dockerin. The novel CBM-Rf6A from this protein has been crystallized and data were processed for the native and a selenomethionine derivative to 1.75 and 1.5 Å resolution, respectively. The crystals belonged to orthorhombic and cubic space groups, respectively. The structure was solved by a single-wavelength anomalous dispersion experiment using the CCP4 program suite and SHELXC/D/E.

1. Introduction

The microbial degradation of plant cell-wall polysaccharides is a fundamental biological process required for carbon turnover that has considerable industrial importance. Different mechanisms have evolved for the hydrolysis of plant cell walls, which are composed primarily of cellulose and hemicellulose and are the most abundant sources of terrestrial biomass and renewable energy. Anaerobic microorganisms organize carbohydrate-active enzymes (CAZymes) into a supramolecular complex termed the 'cellulosome' that degrades cellulose and hemicellulose highly efficiently (Bayer *et al.*, 1998; Gilbert, 2007; Fontes & Gilbert, 2010). CAZymes have been classified into sequence-based families (Lombard *et al.*, 2014; <http://www.cazy.org/>). CAZymes are modular enzymes which contain one or more catalytic domains connected *via* linker sequences to one or more noncatalytic modules. The most prevalent noncatalytic modules appended to CAZymes are carbohydrate-binding modules (CBMs), which target enzymes to the carbohydrate substrates (Boraston *et al.*, 2004). Rumen cellulolytic bacteria have developed a wide array of multi-modular cellulases and other proteins involved in plant cell-wall breakdown. The genome of the ruminal cellulolytic bacterium *Ruminococcus flavefaciens* strain FD-1 has been sequenced, providing an opportunity to discover novel cellulosomal enzymes. The *R. flavefaciens* FD-1 genome encodes over 200 dockerin-containing proteins, most of them of unknown function (Berg Miller *et al.*, 2009). Encoded at locus WP009984389 in the *R. flavefaciens* FD-1 genome is a modular protein of 925 amino-acid residues termed RfCel9A. This modular protein comprises an N-terminal family 9 glycoside hydrolase catalytic module (GH9) linked to two unclassified CBMs (termed CBM-Rf6A and CBM-Rf6B) and a C-terminal dockerin (Fig. 1). CBM-Rf6A and CBM-Rf6B share a sequence identity of 95%. There are no structural homologues of CBM-Rf6. BLAST analysis (Altschul *et al.*, 1990) shows that CBM-Rf6 shares an amino-acid sequence identity of 25% or more with 20 other proteins of unknown function. In the present communication, we describe the

Figure 1

Schematic showing the modular architecture of the full-length *R. flavefaciens* FD-1 *RfCel9A*. SP is the N-terminal signal peptide, GH9 is the catalytic module belonging to glycoside hydrolase family 9 and CBM-*Rf6A* and CBM-*Rf6B* are the putative tandem carbohydrate-binding modules, with the dockerin module (DOC) at the C-terminus. The CBM-*Rf6A* construct used in this study covers the residue range 495–621. The linker regions between the defined modules are expected to be flexible.

Table 1

Macromolecule-production information.

Source organism	<i>R. flavefaciens</i> FD-1
DNA source	WP009984389
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†	MASDGYTIKPNKKVTYSALGEDERMIGFSYKDFGISSEKITEVQVNI SANKNIGKYVGGFTST-TDSANGYWAMGDEITQSI SGNSGTITWKVPSDISSIIQTQ-GDEITQSI SGNSGTITWKVPSDISSIIQTQYGGEIKFGVW-WIDCDEFTIDS VVLKLEHHHHHH

† The cloning artefact is underlined and the Met/SeMet residues are shown in *italics*.

crystallization and preliminary crystallographic studies of this novel CBM (CBM-*Rf6A*) identified in the *R. flavefaciens* FD-1 cellulosome.

2. Materials and methods

2.1. Macromolecule production

The gene encoding CBM-*Rf6A* (residues 495–621 of *RfCel9A*) 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, was subsequently subcloned into the pET-28a expression vector (Novagen), generating p*Rf6*. Recombinant CBM-*Rf6A* contained a C-terminal His₆ tag (LEHHHHHH; Table 1). *E. coli* BL21 cells harbouring p*Rf6* were cultured in Luria–Bertani broth at 310 K to mid-exponential phase ($OD_{600\text{ nm}} = 0.6$) and recombinant protein 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 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-*Rf6* was buffer-exchanged into 50 mM Na HEPES buffer pH 7.5 containing 200 mM NaCl and 5 mM CaCl₂, and was subjected to gel filtration on a HiLoad 16/60 Superdex 75 column (GE Healthcare) at a flow rate of 1 ml min⁻¹. The methionine-auxotrophic *E. coli* strain B834 (DE3) was used to generate selenomethionine-derivatized (SeMet) CBM-*Rf6A*. The cells were washed with the seleno-L-methionine medium and incubated overnight at 310 K to mid-exponential phase ($OD_{600\text{ nm}} = 0.6$), and the protein was purified as described previously (Najmudin *et al.*, 2006; Venditto *et al.*, 2014). Purified CBM-*Rf6A* was concentrated using an Amicon 10 kDa molecular-weight cutoff centrifugal concentrator and washed three times with 0.5 mM CaCl₂, 5 mM DTT for the SeMet protein or 0.5 mM CaCl₂ buffer for the native CBM-*Rf6A*. Recombinant CBM-*Rf6A*, including the C-terminal His₆ tag (LEHHHHHH), has an approximate molecular mass of 15 kDa. The protein concentration was estimated with a Thermo Scientific NanoDrop 2000c using a molar extinction coefficient (ϵ) of 30 940 M⁻¹ cm⁻¹ at 280 nm and was verified by SDS–PAGE (Fig. 2).

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 (Hampton Research, California, USA), the JCSG screen 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 1 μ l 117 mg ml⁻¹ native CBM-*Rf6A* and 1 μ l reservoir solution and the other consisting of 1 μ l 50 mg ml⁻¹ native CBM-*Rf6A* and 1 μ l reservoir solution. Crystals grew in one condition: 1 M trisodium citrate, 0.1 M 2-(*N*-

Figure 2

A Coomassie Brilliant Blue-stained 16% SDS–PAGE evaluation of protein purity. Lane 1, molecular-mass markers (labelled in kDa); lane 2, native CBM-*Rf6A*. Similar purity was obtained for SeMet CBM-*Rf6A*.

Figure 3

Crystals of native CBM-*Rf6A*. Similar crystals were obtained for the selenomethionine derivative that gave data sets X5 and X7 in Table 3. The black scale bar represents 0.1 mm.

Table 2
Crystallization of CBM-*Rf6A*.

	Selenomethionine derivative	Native
Method	Sitting-drop vapour diffusion	Sitting-drop vapour diffusion
Plate type	MRC 96-well 2-drop sitting-drop crystallization plate (Molecular Dimensions)	MRC 96-well 2-drop sitting-drop crystallization plate (Molecular Dimensions)
Temperature (K)	292	292
Protein concentration (mg ml ⁻¹)	46 and 23	117 and 50
Buffer composition of protein solution	Deionized water (Sigma), 0.5 mM CaCl ₂ , 5 mM DTT	Deionized water (Sigma), 0.5 mM CaCl ₂
Composition of reservoir solution	Crystal X5: 0.2 M trisodium citrate, 2 M ammonium sulfate Crystal X7: 0.2 M ammonium acetate, 0.1 M Tris pH 7.5, 1.5 M dipotassium phosphate, 1.5 M sodium dihydrogen phosphate	1 M trisodium citrate, 0.1 M MES pH 6.5
Volume and ratio of drop	0.7 µl protein solution and 0.7 µl reservoir solution.	1 µl protein solution and 1 µl reservoir solution
Volume of reservoir (µl)	50	50

Table 3
Data collection and processing.

Values in parentheses are for the outer shell. Note: the X7 data were processed to a resolution of 1.6 Å rather than 1.5 Å based on a $CC_{1/2}$ value of approximately 0.5.

	SeMet CBM- <i>Rf6A</i>		
	Data set X7	Data set X5	Native CBM- <i>Rf6A</i>
Beamline	PROXIMA1, SOLEIL	PROXIMA1, SOLEIL	BM30, ESRF
Detector	PILATUS 6M, Dectris	PILATUS 6M, Dectris	Quantum 415r CCD, ADSC
Space group	$I2_13$	$I2_12_12_1$	$I2_12_12_1$
Wavelength (Å)	0.95372	0.97895	0.9792
a, b, c (Å)	104.05, 104.05, 104.05	102.30, 103.35, 109.07	102.25, 102.52, 109.46
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
V_M † (Å ³ Da ⁻¹)	2.89	3.19	3.17
Solvent content (%)	58	61	61
Molecules in asymmetric unit	1	3	3
Resolution limits (Å)	42.48–1.6 (1.69–1.60)	43.63–2.0 (2.11–2.00)	30.19–1.75 (1.84–1.75)
No. of observations	277268 (39552)	212034 (8675)	809754 (95080)
No. of unique observations	24839 (3559)	37050 (3981)	58166 (8389)
Multiplicity	11.2 (11.1)	5.7 (2.2)	13.9 (11.3)
Completeness (%)	100.0 (100.0)	93.8 (69.9)	100 (100)
$\langle I/\sigma(I) \rangle$	14.5 (1.3)	5.2 (0.9)	15.8 (1.8)
$CC_{1/2}$ ‡	0.999 (0.462)	0.976 (0.500)	0.999 (0.625)
R_{merge} §	0.092 (1.890)	0.158 (0.694)	0.127 (1.383)
$R_{\text{p.i.m.}}$ ¶	0.032 (0.617)	0.070 (0.606)	0.035 (0.418)

† Matthews coefficient (Matthews, 1968). ‡ $CC_{1/2}$ is the correlation between intensities from random half data sets (Karplus & Diederichs, 2012). § $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, 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.

morpholino)ethanesulfonic acid (MES) pH 6.5 (Fig. 3). Crystals of the SeMet protein were obtained by the sitting-drop vapour-diffusion method, repeating the full screen with commercial kits. Two drops per well containing 50 µl reservoir solution were prepared: one consisting of 0.7 µl 46 mg ml⁻¹ CBM-*Rf6A* and 0.7 µl reservoir solution and the other consisting of 0.7 µl 23 mg ml⁻¹ CBM-*Rf6A* and 0.7 µl reservoir solution. Crystals grew in the following conditions: 0.2 M trisodium citrate, 2 M ammonium sulfate (crystal X5) and 0.2 M ammonium acetate, 0.1 M Tris pH 7.5, 1.5 M dipotassium phosphate, 1.5 M sodium dihydrogen phosphate (crystal X7). All crystals grew to maximum size within a week. The crystals were cryocooled in liquid nitrogen after soaking in cryoprotectant [30% (v/v) glycerol added to the crystallization buffer or just Paratone-N] for a few seconds.

2.3. Data collection and processing

Data for native CBM-*Rf6A* were collected on beamline I02 at Diamond Light Source, Harwell, England and on beamline BM30 at the European Synchrotron Radiation Facility, Grenoble, France from crystals cooled to 100 K using a Cryostream (Oxford Cryosystems). 180° of data were collected with a $\Delta\varphi$ of 0.2° and an exposure of 0.2 s. Data for the selenomethionine derivatives were collected on the PROXIMA1 beamline at SOLEIL, Orsay, France. 200° of data were collected with a $\Delta\varphi$ of 0.2° and an exposure of 0.2 s and a further 360° of data were collected using inverse-beam geometry (two equivalent

complete interleaved data collections starting at $\varphi = 0^\circ$ and $\varphi = 180^\circ$) at the Se peak edge (an energy of 12.665 keV, $f' = -7.63$ and $f'' = 4.99$) for a single-wavelength anomalous diffraction experiment for the first crystal form (crystal X5) and slightly above the Se absorption edge for the second crystal form (crystal X7). As the structure had already been solved using crystal X5, a higher resolution data set (to 1.5 Å resolution) was collected from crystal X7 using inverse-beam geometry. This way, phase information that may prove useful in later refinement can be obtained without recourse to an optimized SAD data collection. All data sets were processed using *iMosflm* (Battye *et al.*, 2011) or *XDS* (Kabsch, 2010) via the command-line interface *xdsmc* (<https://code.google.com/p/xdsmc/>) and *AIMLESS* (Evans, 2006) from the CCP4 suite (Collaborative Computational Project, Number 4, 1994; Winn *et al.*, 2011). Data-collection statistics (as reported by *AIMLESS*) are given in Table 3.

3. Results and discussion

The CBM-*Rf6A* structure was determined using a selenomethionine derivative by a single-wavelength anomalous dispersion experiment for two different crystal forms using the *SHELX* suite (Sheldrick, 2008) via the *HKLMAP* graphical interface (Pape & Schneider, 2004). Inverse-beam data from the orthorhombic crystal form at the peak wavelength corresponding to the Se absorption edge were used

to determine the heavy-atom sites using *SHELXD*. Both integral Se sites were located in each monomer. These sites were then used to calculate phases using *Phaser* in SAD mode (McCoy *et al.*, 2004) in the *CCP4* suite (Winn *et al.*, 2011) followed by density improvement using *Parrot* (Zhang *et al.*, 1997), taking into account the threefold NCS. The quality of the electron-density maps was excellent, and *Buccaneer* (Cowtan, 2006, 2008) interspersed with *REFMAC5* (Murshudov *et al.*, 2011) placed almost 90% of the residues automatically, with *R* and *R*_{free} values of 28.3 and 31.4%, respectively. The SeMet cubic crystal form diffracted to a higher resolution and also exhibited significant anomalous signal to approximately 1.9 Å resolution, as estimated by *XDS* (Kabsch, 2010), *phenix.xtriage* (Zwart *et al.*, 2005) and *SHELXC* (Sheldrick, 2008). The electron-density map calculated by *SHELXE* (Sheldrick, 2008) via the *HKL2MAP* interface (Pape & Schneider, 2004) was of high quality, and automatic model building using *Buccaneer* (Cowtan, 2006, 2008) modelled 92% of the residues, with *R* and *R*_{free} values of 24.8 and 26.5%, respectively. The SeMet orthorhombic (crystal X5) model was placed relative to the correct origin for the native CBM-Rf6A data set (to a resolution of 1.75 Å) by molecular replacement using *Phaser* (McCoy *et al.*, 2007), giving a translation-function *Z*-score of 62.2 and a log-likelihood gain of 16 019. Further structure refinement and analysis are ongoing.

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 under BioStruct-X (grant agreement No. 283570, proposal Biostruct-X_4399). This paper reflects the views of the authors only. The European Community is not liable for any use that may be made of the information contained herein. We acknowledge Diamond Light Source (beamline I02), Harwell, England, SOLEIL (beamline PROXIMA-1; proposal 20120372), Orsay, France and ESRF (beamline BM30), Grenoble, France for provision of synchrotron-radiation facilities. We would also like to thank Cecília

Bonifácio for help with crystallization and Drs James Sandy, Victor D. Alves, Pedro Bule and Teresa Ribeiro 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.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Bayer, E. A., Shimon, L. J., 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.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Cowtan, K. (2006). *Acta Cryst.* **D62**, 1002–1011.
- Cowtan, K. (2008). *Acta Cryst.* **D64**, 83–89.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- 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.
- Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. (2014). *Nucleic Acids Res.* **42**, D490–D495.
- 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., 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.
- 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.
- Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- Venditto, I., Santos, H., Ferreira, L. M. A., Sakka, K., Fontes, C. M. G. A. & Najmudin, S. (2014). *Acta Cryst.* **F70**, 754–757.
- 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.
- Zwart, P., Grosse-Kunstleve, R. & Adams, P. (2005). *CCP4 Newsl. Protein Crystallogr.* **43**, contribution 10.