Expression, purification and crystallization of a novel carbohydrate-binding module from the *Ruminococcus flavefaciens* cellulosome

Immacolata Venditto, Maria S. J. Centeno, Luis M. A. Ferreira, Carlos M. G. A. Fontes and Shabir Najmudin*

CIISA, Faculdade de Medicina Veterinária,
Universidade de Lisboa, Avenida da
Universidade Técnica, 1300-477 Lisbon,
Portugal

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

Received 18 September 2014
Accepted 3 November 2014

Anaerobic bacteria organize carbohydrate-active enzymes into a multi-component complex, the cellulosome, which degrades cellulose and hemicellulose highly efficiently. Genome sequencing of *Ruminococcus flavefaciens* FD-1 offers extensive information on the range and diversity of the enzymatic and structural components of the cellulosome. The *R. flavefaciens* FD-1 genome encodes over 200 dockerin-containing proteins, most of which are of unknown function. One of these modular proteins comprises a glycoside hydrolase family 5 catalytic module (GH5) linked to an unclassified carbohydrate-binding module (CBM-*Rf*1) and a dockerin. The novel CBM-*Rf*1 has been purified and crystallized. The crystals belonged to the trigonal space group *R*32:*H*. The CBM-*Rf*1 structure was determined by a multiple-wavelength anomalous dispersion experiment using *AutoSol* from the *PHENIX* suite using both selenomethionyl-derivative and native data to resolutions of 2.28 and 2.0 Å, respectively.

1. Introduction

The requirement for a consortium of enzymes for the degradation of plant cell-wall polysaccharides reflects the complexity of the carbohydrates within the plant cell wall. Carbohydrate substrates are often insoluble and microorganisms use extracellular enzymes to convert the polysaccharides into soluble sugars that can be transported into cells (Wilson, 2008). Anaerobic microorganisms organize a wide array of multi-modular extracellular cellulases and hemicellulases into a large multi-enzyme complex termed the cellulosome (Bayer *et al.*, 2004; Fontes & Gilbert, 2010). Cellulosome assembly results from the tight interaction established between dockerin modules located in cellulosomal enzymes with reiterated cohesin domains located in a large noncatalytic modular scaffolding protein. A number of anaerobic bacteria have been shown to produce cellulosomes similar to the well characterized cellulosome system of *Clostridium thermocellum*, including *Ruminococcus flavefaciens* FD-1, one of the most important microbes involved in plant cell-wall hydrolysis in the rumen of mammalian herbivores. The *R. flavefaciens* FD-1 genome encodes over 200 dockerin-containing proteins, including a large repertoire of carbohydrate-active enzymes (CAZYmes) and several modules of unknown function (Berg Miller *et al.*, 2009). Considering the vital role that cellulosomes play in the breakdown of structural carbohydrates, it is highly probable that all dockerin-containing proteins are important for polysaccharide degradation. One of these is a modular enzyme of 681 amino acids encoded at locus WP009983134 in the *R. flavefaciens* FD-1 genome. The architectural arrangement of this protein comprises a glycoside hydrolase family 5 catalytic module (GH5) linked to an unclassified carbohydrate-binding module (here termed CBM-*Rf*1) and a C-terminal dockerin. Preliminary experiments involving affinity gel electrophoresis, a microarray platform and isothermal titration calorimetry indicate that CBM-*Rf*1 binds decorated β -1,4-glucans and displays a preference for xyloglucan (unpublished data). There are no structural homologues of CBM-*Rf*1, and *BLAST* analysis (Altschul *et al.*, 1990) shows that this CBM shares more than 25% amino-acid sequence identity with nine other proteins (see Supplementary Fig. S1¹). In

order to gain insight into the structural properties that modulate ligand recognition by CBM-*Rf1*, a recombinant derivative of this domain was expressed, purified and crystallized. Three-dimensional structural determination of CBM-*Rf1* will contribute towards the elucidation of the mechanisms by which highly populated multi-enzyme complexes recognize structural carbohydrates.

2. Materials and methods

2.1. Macromolecule production

The gene encoding CBM-*Rf1* (residues 438–586 of the 681-residue full-length protein) 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 used for subsequent subcloning into the pET-28a vector, generating p*Rf1*, which encodes CBM-*Rf1* plus a C-terminal His₆ tag. *E. coli* BL21 cells harbouring p*Rf1* were cultured in Luria–Bertani broth at 310 K to the mid-exponential phase ($OD_{600\text{ nm}} = 0.6$) and recombinant protein overproduction was induced by adding isopropyl β -D-thiogalactopyranoside (1 mM final concentration) with incubation for a further 16 h at 292 K. The His₆-tagged recombinant CBM-*Rf1* was purified from cell-free extracts by immobilized metal-ion affinity chromatography (IMAC) as described previously (Najmudin *et al.*, 2010). Purified CBM-*Rf1* was buffer-exchanged into 50 mM Na HEPES buffer pH 7.5 containing 200 mM NaCl and 5 mM CaCl₂ and subsequently 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* to generate selenomethionylated CBM-*Rf1* (SeMet-CBM-*Rf1*) was performed as described in Najmudin *et al.* (2006) and Venditto *et al.* (2014) and the protein was purified using the same procedures as employed for the native CBM. Purified CBM-*Rf1* was concentrated using an Amicon 10 kDa molecular-mass centrifugal concentrator and washed three times with 0.5 mM CaCl₂, 5 mM DTT (for the SeMet protein) or 0.5 mM CaCl₂ (for native CBM). Recombinant CBM-*Rf1* with the engineered C-terminal His₆ tag (LEHHHHHH) has an approximate molecular mass of 18 kDa. The protein concentration was estimated using a molar extinction coefficient (ϵ) of 23 950 M⁻¹ cm⁻¹ at 280 nm with a Thermo Scientific NanoDrop 2000c and was verified by SDS–PAGE (Fig. 1). Macromolecule-production information is summarized in Table 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 and PEG/Ion 2 (Hampton Research, California, USA) and the JCSG screen (Newman *et al.*, 2005) and an in-house 80 factorial screen using an Oryx8 robotic nanodrop dispensing system (Douglas Instruments). The conditions that gave diffracting crystals (Figs. 2*a*, 2*b* and 2*c*) of native CBM-*Rf1* are given in Table 2. These conditions were then used as part of optimization screens to obtain SeMet-CBM-*Rf1* crystals (Fig. 2*d*). The crystals grew to their maximum size within two weeks. Crystals were cryo-cooled in liquid nitrogen after soaking in cryoprotectant [30%(v/v) glycerol added to the crystallization buffer] for a few seconds.

2.3. Data collection and processing

Data were collected from native CBM-*Rf1* crystals on beamline I02 at Diamond Light Source, Harwell, England. 360° of data were collected with a $\Delta\phi$ of 0.2° and an exposure of 0.2 s (Fig. 3). Data

Table 1

Macromolecule-production information.

Source organism	<i>R. flavefaciens</i> FD-1
DNA source	WP009983134
Forward primer	CACACACCATGGGACTCGCTCCGCTGGCTGATG
Reverse primer	CACACACTCGAGGGCCTCATCGCTGATGC
Cloning vector	pET-28a
Expression vector	pET-28a
Expression host	<i>E. coli</i> BL21
Complete amino-acid sequence of the construct produced†	MGLAPLADGEKLYGKKGSEGTVTFTKAIGDNFAFVEIKTGADT-GFMNGCLGFSESIDGKNYVWVYVQTKKSDTISIDMSSPV-QIAEIIIGTETQEVTDADTIKKLTDKIKTEKSALLQVWYAS-DKTKGQIDPADSASESIEVYIPSADEALEHHHHHH

† Extra residues from cloning artefacts, including the C-terminal His₆ tags, are underlined.

for the SeMet derivative were collected on beamline ID29 at the European Synchrotron Radiation Facility, Grenoble, France. 120° of data were collected with a $\Delta\phi$ of 0.1° and an exposure of 0.04 s. An energy scan around the Se peak was carried out to determine the energies for a multiple-wavelength anomalous diffraction experiment. All crystals were cooled to 100 K using a Cryostream (Oxford Cryosystems) and data were collected using a PILATUS 6M detector (Dectris). The data sets were processed using *iMosflm* (Battye *et al.*, 2011) or *XDS* (Kabsch, 2010) and *AIMLESS* (Evans, 2006) from the *CCP4* suite (Winn *et al.*, 2011), *fast_dp* (Winter, 2010) or *xia2* (Winter *et al.*, 2013). Data-collection statistics are given in Table 3. All of the diffracting CBM-*Rf1* crystals belonged to the trigonal space group *R*32:*H* with two molecules in the asymmetric unit, a solvent content of ~52% and a Matthews coefficient of ~2.57 Å³ Da⁻¹ (Matthews, 1968).

3. Results and discussion

Isomorphous data were obtained for native CBM-*Rf1* from three different crystallization conditions. X-ray fluorescence scans detected the presence of zinc in the native crystals. Data for the X3 data set were collected at the zinc edge and attempts were made to solve the

Figure 1

A Coomassie Brilliant Blue-stained 16% SDS–PAGE gel evaluation of protein purity (*a*) for native CBM-*Rf1* and (*b*) for the SeMet-CBM-*Rf1* derivative. Lane 1, molecular-mass markers (NZYTech Ltd, Portugal; labelled in kDa); lane 2, protein (labelled in kDa).

structure by SAD, but to no avail. Subsequently, the CBM-Rf1 structure was determined using a SeMet derivative by a multiple-wavelength anomalous dispersion experiment using *AutoSol* (Terwilliger *et al.*, 2009) from the *PHENIX* suite (Adams *et al.*, 2010). Since a large number of data sets had been collected, various combinations of data sets were tested. The best result was obtained using the peak and remote 1 data sets from the SeMet-CBM-Rf1 crystals and the X3 and X9 data sets from the native CBM-Rf1 crystals. All four expected Se-atom sites (two per monomer, corresponding to the well defined internal SeMet in CBM-Rf1) were identified with a figure of merit of 0.21 and a correlation coefficient of 40.6. The final model after *AutoBuild* (Terwilliger *et al.*, 2008) placed 269 amino-acid residues out of a potential 318 in 14 fragments with an

R_{work} and R_{free} of 0.2691 and 0.3359, respectively, in the remote 1 data set to a resolution of 2.28 Å. The structure was improved to 296 amino-acid residues (214 with side chains) in two fragments with an R_{work} and R_{free} of 0.2329 and 0.2982, respectively, using the *Buccaneer* pipeline (Cowtan, 2008). This *Buccaneer* model was placed relative to the correct origin for the native CBM-Rf1 X9 data set (to a resolution of 2.0 Å) by molecular replacement (*Phaser*; McCoy *et al.*, 2007), giving a TFZ of 44.2 and an LLG of 4056.

4. Related literature

The following references are cited in the Supporting Information for this article: Thompson *et al.* (1994).

Table 2
Crystallization.

	SeMet derivative	Native
Method	Hanging-drop vapour diffusion	Sitting-drop vapour diffusion
Plate type	Greiner 24-well hanging-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)	36	83 and 41.5
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	0.3 M ammonium sulfate, 24%(w/v) polyethylene glycol 3350 (Fig. 2d)	Data set X3: 0.2 M potassium sulfate, 20%(w/v) polyethylene glycol 3350 (Fig. 2a) Data set X6: 0.2 M ammonium sulfate, 20%(w/v) polyethylene glycol 3350 (Fig. 2b) Data set X9: 1.2 M trisodium citrate, 0.1 M HEPES pH 7.5 (Fig. 2c)
Volume and ratio of drop	1 µl protein solution, 1 µl reservoir solution	0.7 µl protein solution, 0.7 µl reservoir solution
Volume of reservoir (µl)	600	50

Figure 2
The crystals of native CBM-Rf1 that gave data sets X3 (a), X6 (b) and X9 (c) and crystals of the SeMet-CBM-Rf1 derivative (d). The black scale bar represents 0.1 mm.

Table 3
Data collection and processing.

Values in parentheses are for the outer shell.

Data set	SeMet MAD				Native		
	Peak	Inflection point	Remote 1	Remote 2	X3	X6	X9
Diffraction source	ID29, ESRF				I02, Diamond Light Source		
Wavelength (Å)	0.97916	0.97939	0.96112	0.97685	1.2855	0.97949	0.97949
Space group	R32:H				R32:H		
a, b, c (Å)	131.9, 131.9, 104.8	132.1, 132.1, 105.0	131.7, 131.7, 104.7	132.1, 132.1, 105.3	132.3, 132.3, 105.0	132.1, 132.1, 104.2	131.1, 131.1, 105.2
Mosaicity (°)	0.37	0.31	0.54	0.23	0.44	0.47	0.56
Resolution range (Å)	77.24–2.36 (2.45–2.36)	77.35–2.54 (2.65–2.54)	77.13–2.28 (2.35–2.28)	77.46–2.87 (3.03–2.87)	66.13–2.69 (2.76–2.69)	50.14–2.59 (2.66–2.59)	30.17–2.00 (2.05–2.00)
Total No. of reflections	85948 (9328)	56562 (6824)	99432 (9474)	30719 (4582)	192100 (13900)	199342 (15471)	398258 (26385)
No. of unique reflections	14504 (1508)	11693 (1420)	16115 (1470)	7651 (1115)	9947 (733)	10960 (789)	23568 (1731)
Completeness (%)	99.9 (99.9)	99.6 (99.7)	99.9 (100)	94.1 (95.2)	99.8 (99.7)	99.9 (99.2)	100 (100)
Multiplicity	5.9 (6.2)	4.8 (4.8)	6.2 (6.4)	4.0 (4.1)	19.3 (19.0)	18.2 (19.6)	16.9 (15.2)
CC _{1/2} †	0.998 (0.715)	0.995 (0.603)	0.999 (0.837)	0.980 (0.554)	0.996 (0.922)	0.998 (0.916)	0.997 (0.847)
⟨I/σ(I)⟩	11.2 (2.1)	8.1 (1.8)	16.4 (3.2)	5.1 (1.6)	14.1 (3.2)	15.5 (3.6)	12.5 (1.7)
R _{merge} ‡ (%)	8.9 (81.7)	10.4 (76.7)	6.5 (54.2)	16.5 (74.2)	15.3 (116.5)	12.4 (131.4)	11.7 (170.7)
R _{p.i.m.} § (%)	3.8 (34.6)	5.5 (41.6)	2.7 (22.7)	9.0 (41.6)	3.6 (27.4)	3.0 (30.6)	2.9 (44.9)

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

Figure 3
Representative diffraction pattern for crystal X9, with the outer resolution ring shown as a dashed circle.

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 No. Biostruct-X_4399). 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 Diamond Light Source (beamline I02), Harwell, England and ESRF (beamline ID29), Grenoble, France for provision of synchrotron-radiation facilities. We would also like to thank Cecília Bonifácio

and Pedro Bule for help with crystallization and Drs Juan Sanchez-Weatherby, Filipe Freire, Catarina Coelho and Joana L. A. Brás for help with data collection.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
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., Belaich, J.-P., Shoham, Y. & Lamed, R. (2004). *Annu. Rev. Microbiol.* **58**, 521–554.
Berg Miller, M. E. *et al.* (2009). *PLoS One*, **4**, e6650.
Cowtan, K. (2008). *Acta Cryst.* **D64**, 83–89.
Diederichs, K. & Karplus, P. A. (2013). *Acta Cryst.* **D69**, 1215–1222.
Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
Fontes, C. M. G. A. & Gilbert, H. J. (2010). *Annu. Rev. Biochem.* **79**, 655–681.
Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
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.
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.
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.
Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H. & Hung, L.-W. (2009). *Acta Cryst.* **D65**, 582–601.
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.* **D64**, 61–69.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Venditto, I., Santos, H., Ferreira, L. M. A., Sakka, K., Fontes, C. M. G. A. & Najmudin, S. (2014). *Acta Cryst.* **F70**, 754–757.
Wilson, D. B. (2008). *Ann. N. Y. Acad. Sci.* **1125**, 289–297.
Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.
Winter, G., Lobley, C. M. C. & Prince, S. M. (2013). *Acta Cryst.* **D69**, 1260–1273.