Pedro Bule,^a Vered Ruimy-
Israeli,^b Vânia Cardoso,^a
Edward A. Bayer,^b
Carlos M. G. A. Fontes^{a*} and
Shabir Najmudin^{a*}

^aCIISA—Faculdade de Medicina Veterinária,
Universidade de Lisboa, Avenida da
Universidade Técnica, 1300-477 Lisboa,
Portugal, and ^bDepartment of Biological
Chemistry, The Weizmann Institute of Science,
Rehovot 76100, Israel

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

Received 16 April 2014
Accepted 30 May 2014

Overexpression, crystallization and preliminary X-ray characterization of *Ruminococcus flavefaciens* scaffoldin C cohesin in complex with a dockerin from an uncharacterized CBM-containing protein

Cellulosomes are massive cell-bound multienzyme complexes tethered by macromolecular scaffolds that coordinate the efforts of many anaerobic bacteria to hydrolyze plant cell-wall polysaccharides, which are a major untapped source of carbon and energy. Integration of cellulosomal components occurs *via* highly ordered protein–protein interactions between cohesin modules, located in the scaffold, and dockerin modules, found in the enzymes and other cellulosomal proteins. The proposed cellulosomal architecture for *Ruminococcus flavefaciens* strain FD-1 consists of a major scaffoldin (ScaB) that acts as the backbone to which other components attach. It has nine cohesins and a dockerin with a fused X-module that binds to the cohesin on ScaE, which in turn is covalently attached to the cell wall. The ScaA dockerin binds to ScaB cohesins allowing more carbohydrate-active modules to be assembled. ScaC acts as an adaptor that binds to both ScaA and selected ScaB cohesins, thereby increasing the repertoire of dockerin-bearing proteins that integrate into the complex. In previous studies, a screen for novel cohesin–dockerin complexes was performed which led to the identification of a total of 58 probable cohesin–dockerin pairs. Four were selected for subsequent structural and biochemical characterization based on the quality of their expression and the diversity in their specificities. One of these is C12D22, which comprises the cohesin from the adaptor ScaC protein bound to the dockerin of a CBM-containing protein. This complex has been purified and crystallized, and data were collected to resolutions of 2.5 Å (hexagonal, $P6_5$), 2.16 Å (orthorhombic, $P2_12_12_1$) and 2.4 Å (orthorhombic, $P2_12_12$) from three different crystalline forms.

1. Introduction

The plant cell wall represents a major untapped global source of carbon and energy. It is especially important for herbivores, who are able to utilize this energy source thanks to the presence of cellulolytic bacteria in their gastrointestinal tract. One of these is *Ruminococcus flavefaciens*, a Gram-positive coccus from the order Clostridiales present in the rumen community (Sijpesteijn, 1951). Although their cellulolytic enzyme systems have been investigated for many years, the mechanisms by which they achieve plant cell-wall breakdown have yet to be fully characterized (Ding *et al.*, 2001). Like many other anaerobic bacteria, they possess multienzyme complexes termed cellulosomes, which comprise a range of cellulases and hemicellulases that degrade the structural polysaccharides in a highly efficient and concerted way.

The assembly of cellulosomes occurs *via* highly ordered protein–protein interactions between cohesins, which are located in a macromolecular scaffold (scaffoldin), and dockerins, which are found in the enzymes or on the scaffoldins themselves (Bayer *et al.*, 2004; Fontes & Gilbert, 2010). Strain FD-1 of *R. flavefaciens* (first isolated from a bolus, containing ruminal microorganisms, used to improve rumen function in calves; Bryant *et al.*, 1958) produces one of the most intricate and potentially versatile cellulosomes described to date. The *R. flavefaciens* FD-1 genome encodes 222 dockerin-bearing proteins, most of them of unknown function (Berg Miller *et al.*, 2009; Rincon *et al.*, 2010). Preliminary analysis of the primary sequences of the dockerins has organized the 222 different proteins into six

different families. Family 1 has four different subfamilies, while families 4 and 6 each have two different subfamilies. There are eight dockerins that were not assigned to any families. In this highly elaborate cellulosome, scaffoldin B (ScaB) acts as the backbone to which other components attach (Rincon *et al.*, 2003). It comprises nine cohesins of two distinct types. Cohesins 1–4 are similar to the two cohesins on ScaA, while cohesins 5–9 are closer to those found in ScaB of *R. flavefaciens* strain 17. It also has a dockerin with an X-module that binds to the cohesin on ScaE, which in turn covalently attaches itself to the cell wall with an LPXTG motif *via* a sortase-mediated mechanism (Rincon *et al.*, 2005). The ScaA dockerin binds to the second type of ScaB cohesin, allowing more carbohydrate-active modules to bind the complex (Rincon *et al.*, 2003). ScaC acts as an adaptor that binds to both ScaA and the first type of ScaB cohesin, thus serving to increase the repertoire of proteins that can be present in the complex (Rincon *et al.*, 2004; Bayer *et al.*, 2008).

Here, we describe the expression, purification and crystallization of the cohesin from ScaC (*R/ScaCCoh*; C12) bound to the dockerin from an uncharacterized protein (*R/CBM6Doc*; D22) that contains a family 6 carbohydrate-binding module (CBM6).

2. Materials and methods

2.1. Protein expression and purification

To co-express the *R. flavefaciens* dockerin and cohesin in the same cells, the genes encoding the two proteins optimized for expression in *Escherichia coli* were synthesized *in vitro* (NZYTech Ltd, Portugal; Cameron *et al.*, 2012). The dockerin-encoding gene was positioned at the 3' end and the cohesin-encoding gene at the 5' end. Between the two, the sequences of the T7 terminator (to terminate transcription of the cohesin gene) and the T7 promoter (to control transcription of the dockerin gene) were added. This construct also contained specifically tailored *NheI* and *NcoI* recognition sites at the 5' end and *XhoI* and *SalI* at the 3' end to allow subcloning into pET-28a (Novagen) such that the sequence encoding a six-residue His tag could be introduced either at the N-terminus of the cohesin (by cutting with *NheI* and *SalI*, with the additional sequence MGSSHHHHHHSSGLVPRGSHMAS) or at the C-terminus of the dockerin (by cutting with *NcoI* and *XhoI*, with the additional sequence LEHHHHHH). Thus, by subcloning the initial construct in two different ways in pET-28a we could generate protein–protein

complexes that could contain a His tag either on the cohesin (C12D22_Coh) or the dockerin (C12D22_CDco).

A second construct was designed by switching the positions of the dockerin- and cohesin-encoding sequences, which resulted in either a complex with an N-terminally tagged dockerin (C12D22_NDoc) or a C-terminally tagged cohesin. Expression screens revealed that the cohesin-tagged complexes failed to produce a significant yield of dockerin. Thus, the dockerin-tagged complexes were selected for subsequent work.

Both C12D22_Doc cohesin–dockerin complexes were expressed in *E. coli* BL21 cells grown at 310 K to an OD₆₀₀ of 0.5. Recombinant protein expression was induced by the addition of 0.2 mM isopropyl β -D-1-thiogalactopyranoside and incubation at 292 K for 16 h. The recombinant complex was purified by immobilized metal-ion affinity chromatography (IMAC) using Sepharose columns loaded with nickel (HisTrap, GE Healthcare, UK) and following conventional protocols (Najmudin *et al.*, 2006). Since the levels of expression of the cohesin are much higher than those of the dockerin and the His tag is on the dockerin, we assumed that most of the purified molecules were complexed. The buffer of all fractions containing the purified cohesin–dockerin complex was exchanged into 50 mM HEPES buffer pH 7.5 containing 200 mM NaCl, 5 mM CaCl₂ using a PD-10 Sephadex G-25M gel-filtration column (Amersham Pharmacia Biosciences, UK). A further purification step by gel-filtration chromatography was performed by loading the samples onto a HiLoad 16/60 Superdex 75 (GE Healthcare, UK) at a flow rate of 1 ml min^{−1}. Fractions containing the purified complex were then concentrated with Amicon Ultra-15 centrifugal devices with a 10 kDa cutoff membrane (Millipore, USA) and washed three times with 0.5 mM CaCl₂. The protein concentration was estimated using a molar extinction coefficient (ϵ) of 26 025 M^{−1} cm^{−1} with a NanoDrop 2000c spectrophotometer (Thermo Scientific, USA). The final protein concentrations were adjusted to a maximum of 38 mg ml^{−1} for the C-terminal tagged dockerin complex (C12D22_CDco) and 81 mg ml^{−1} for the N-terminally tagged dockerin complex (C12D22_NDoc) with 0.5 mM CaCl₂. The purity and molecular mass of the recombinant protein were analysed by 14%(w/v) SDS–PAGE (Fig. 1).

2.2. Crystallization

The crystallization conditions were set up using the sitting-drop vapour-diffusion method with an Oryx8 robotic nanodrop dispensing system (Douglas Instruments, UK). The commercial kits Crystal Screen, Crystal Screen 2, PEG/Ion and PEG/Ion 2 (Hampton Research, California, USA), JCSG+ HT96 (Molecular Dimensions, UK) and an in-house screen (80 factorial) were used for the screening. 0.7 μ l drops of 15 and 30 mg ml^{−1} C12D22_CDco were mixed with 0.7 μ l reservoir solution at room temperature per well containing 50 μ l of the crystallization solution. The same process was repeated with C12D22_NDoc with protein concentrations of 40 and 81 mg ml^{−1}. The resulting plates were then stored at 292 K. Crystal formation was observed in nine different conditions with the C-terminal tagged complex (maximum dimensions $\sim$ 300 $\times$ 20 $\times$ 20 μ m) and 35 conditions with the N-terminal tagged protein (maximum dimensions $\sim$ 100 $\times$ 20 $\times$ 20 μ m). All the crystals were obtained from the initial screens. These crystals were cryoprotected with mother solution containing 20–30% glycerol or with 100% Paratone-N (Hampton Research, USA) and flash-cooled in liquid nitrogen.

2.3. Data collection and processing

Data were collected on beamline I04 at the Diamond Light Source, Harwell, England using a PILATUS 6M detector (Dectris Ltd) from

Figure 1
SDS–PAGE [14%(w/v)] showing overexpression and purification of C12D22_CDco. Lane 1, LMW protein marker (labelled in kDa); lanes 2–6, purified C12 (top bands, 18.5 kDa) and D22 (bottom bands, 8.2 kDa). A similar gel was obtained for the C12D22_NDoc.

Table 1

Data-collection statistics.

Values in parentheses are for the highest resolution shell.

Data set	C12D22_CDdoc monomer	C12D22_CDdoc dimer	C12D22_CDdoc monomer
Beamline	I04, Diamond	I04, Diamond	I04, Diamond
Space group	$P6_5$	$P2_12_12_1$	$P2_12_12$
Wavelength (Å)	0.9795	0.9795	0.9795
Unit-cell parameters			
a (Å)	81.87	59.49	58.79
b (Å)	81.87	66.7	108.3
c (Å)	82.14	109.6	32.81
α (°)	90	90	90
β (°)	90	90	90
γ (°)	120	90	90
V_M † (Å ³ Da ⁻¹)	3.02	2.04	1.96
Solvent content (%)	59	40	37
No of molecules in asymmetric unit	1 Coh–Doc complex	2 Coh–Doc complexes	1 Coh–Doc complex
Resolution limits (Å)	70.91–2.50 (2.64–2.50)	52.28–2.16 (2.22–2.16)	58.79–2.40 (2.46–2.40)
No. observations	274754 (39906)	103107 (7228)	37219 (2868)
No. unique observations	11052 (1594)	23625 (1733)	8616 (629)
Multiplicity	24.9 (25.0)	4.4 (4.2)	4.3 (4.6)
Completeness (%)	100.0 (100.0)	98.5 (99.0)	99.1 (99.2)
$\langle I/\sigma(I) \rangle$	15.1 (1.2)	12.6 (2.3)	15.3 (1.9)
$CC_{1/2}$ ‡	0.999 (0.70)	0.998 (0.787)	0.999 (0.737)
R_{merge} §	0.135 (3.695)	0.086 (0.646)	0.066 (0.645)
$R_{\text{p.i.m.}}$ ¶	0.028 (0.746)	0.045 (0.352)	0.036 (0.394)

† Matthews coefficient (Matthews, 1968). ‡ $CC_{1/2}$ is the half-data-set correlation coefficient (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.

crystals cooled to 100 K using a Cryostream (Oxford Cryosystems Ltd). A systematic grid search was carried out on all of these crystals to select the best diffracting part of the crystal. *EDNA* (Winter & McAuley, 2011) and *iMosflm* (Battye *et al.*, 2011) were used for strategy calculation during data collection. All data sets were processed using the *Fast_dp* and *xia2* (Winter, 2010) packages, which use the programs *XDS* (Kabsch, 2010), *POINTLESS* (Evans, 2006) and *SCALA* (Evans, 2006) from the *CCP4* suite (Winn *et al.*, 2011). Data-collection statistics are given in Table 1.

3. Results and discussion

The C12D22_CDdoc crystals obtained from condition JCSG+ 1.5 [0.2 M magnesium formate, 20%(w/v) PEG 2000 MME] (Fig. 2a) diffracted to a resolution of 2.5 Å. Crystals from condition JCSG+ 2.33 [0.1 M potassium thiocyanate, 30%(w/v) PEG 2000 MME] (Fig. 2b) diffracted to a resolution of 2.16 Å. The crystals from all the other conditions did not diffract at all or diffracted to a low-resolution range from 10 to 4 Å. All the crystals that diffracted well were obtained from the construct that contained the His tag at the C-terminus of the dockerin. The best crystal from the JCSG+ 1.5 condition belonged to the hexagonal space group $P6_5$. *BALBES* was used to carry out molecular replacement (Long *et al.*, 2008). The best solution was found using the type I cohesin–dockerin complex from *C. thermocellum* [PDB entry 2ccl; Carvalho *et al.*, 2007], the cohesin of which gave a sequence identity of 30.0% and the dockerin of 31.7%, with an R factor and R_{free} of 39.8 and 41.2%, respectively, and a Q -factor of 0.506 after *REFMAC5* (Murshudov *et al.*, 2011) at the end of the *BALBES* run. An *ARP/wARP* (Langer *et al.*, 2008) run after *BALBES* gave a model of 90 residues in 13 chains, with an estimated correctness of 54.7%. Two crystalline forms were obtained

(a)

(b)

Figure 2

(a) Crystals of C12D22_CDdoc obtained by sitting-drop vapour diffusion. The largest crystals are approximately $300 \times 20 \times 20 \mu\text{m}$ in size. (b) Crystals of C12D22_NDdoc obtained by sitting-drop vapour diffusion. The largest crystals are approximately $100 \times 20 \times 20 \mu\text{m}$ in size.

from the same JCSG+ 2.33 condition in the orthorhombic space group, $P2_12_12_1$ and $P2_12_12$, with protein concentrations of 30 and 15 mg ml⁻¹, respectively. The model from the hexagonal form was used as a search model in *Phaser* (McCoy *et al.*, 2007) to solve the orthorhombic forms by molecular replacement. The final LLG and TFZ scores were 2762 and 25.5, respectively, for the $P2_12_12_1$, and 775 and 11.1, respectively, for the $P2_12_12$ form. Initial refinement with *REFMAC5* gave R -factor and R_{free} values of 24.45 and 30.58%, respectively, for the $P2_12_12_1$ form, and 25.03% and 30.38%, respectively, for the $P2_12_12$ form. Structure refinement and analysis are ongoing for all three structures.

The authors acknowledge financial support from Fundação para a Ciência e a Tecnologia (Portugal) through projects PTDC/BIA-PRO/103980/2008 and EXPL/BIA-MIC/1176/2012 and individual grant SFRH/BD/86821/2012 (to PB). We acknowledge Diamond Light Source, Oxfordshire for provision of synchrotron-radiation facilities (beamline I04). We would also like to thank Cecília Bonifácio for help with crystallization and James Sandy, José Brandão-Neto, Catarina Coelho, Immacolata Venditto and Hugo Correia for help

with data collection, and the European Community's Seventh Framework Programme (FP7/2007-2013) under BioStruct-X (grant agreement No. 283570, proposal Biostruct-X_4399) for funding.

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.
- Bayer, E. A., Lamed, R., White, B. A. & Flint, H. J. (2008). *Chem. Rec.* **8**, 364–377.
- Berg Miller, M. E. *et al.* (2009). *PLoS One*, **4**, e6650.
- Bryant, M. P., Small, N., Bouma, C. & Robinson, I. M. (1958). *J. Bacteriol.* **76**, 529–537.
- Cameron, K., Alves, V. D., Bule, P., Ferreira, L. M. A., Fontes, C. M. G. A. & Najmudin, S. (2012). *Acta Cryst.* **F68**, 1030–1033.
- Carvalho, A. L., Dias, F. M. V., Nagy, T., Prates, J. A. M., Proctor, M. R., Smith, N., Bayer, E. A., Davies, G. J., Ferreira, L. M. A., Romão, M. J., Fontes, C. M. G. A. & Gilbert, H. J. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 3089–3094.
- Diederichs, K. & Karplus, P. A. (2013). *Acta Cryst.* **D69**, 1215–1222.
- Ding, S.-Y., Rincon, M. T., Lamed, R., Martin, J. C., McCrae, S. I., Aurilia, V., Shoham, Y., Bayer, E. A. & Flint, H. J. (2001). *J. Bacteriol.* **183**, 1945–1953.
- 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.
- Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
- Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* **D64**, 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.
- 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.
- Rincon, M. T., Čepeljnik, T., Martin, J. C., Lamed, R., Barak, Y., Bayer, E. A. & Flint, H. J. (2005). *J. Bacteriol.* **187**, 7569–7578.
- Rincon, M. T., Dassa, B., Flint, H. J., Travis, A. J., Jindou, S., Borovok, I., Lamed, R., Bayer, E. A., Henrissat, B., Coutinho, P. M., Antonopoulos, D. A., Berg Miller, M. E. & White, B. A. (2010). *PLoS One*, **5**, e12476.
- Rincon, M. T., Ding, S.-Y., McCrae, S. I., Martin, J. C., Aurilia, V., Lamed, R., Shoham, Y., Bayer, E. A. & Flint, H. J. (2003). *J. Bacteriol.* **185**, 703–713.
- Rincon, M. T., Martin, J. C., Aurilia, V., McCrae, S. I., Rucklidge, G. J., Reid, M. D., Bayer, E. A., Lamed, R. & Flint, H. J. (2004). *J. Bacteriol.* **186**, 2576–2585.
- Sijpesteijn, A. K. (1951). *J. Gen. Microbiol.* **5**, 869–879.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.
- Winter, G. & McAuley, K. E. (2011). *Methods*, **55**, 81–93.