Overexpression, purification, crystallization and preliminary X-ray characterization of the fourth scaffoldin A cohesin from *Acetivibrio cellulolyticus* in complex with a dockerin from a family 5 glycoside hydrolase

Pedro Bule, Ana Correia, Kate Cameron, Victor D. Alves, Vânia Cardoso, Carlos M. G. A. Fontes* and Shabir Najmudin*

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

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

Received 26 May 2014
Accepted 5 June 2014

Cellulosomes are cell-bound multienzyme complexes secreted by anaerobic bacteria that play a crucial role in carbon turnover by degrading plant cell walls to simple sugars. Integration of cellulosomal components occurs *via* highly ordered protein–protein interactions between cohesin modules located in a molecular scaffold and dockerin modules found in cellulosomal enzymes. *Acetivibrio cellulolyticus* possesses a complex cellulosome arrangement which is organized by a primary enzyme-binding scaffoldin (ScaA), two anchoring scaffoldins (ScaC and ScaD) and an unusual adaptor scaffoldin (ScaB). A dockerin from a family 5 glycoside hydrolase (GH5), which was engineered to inactivate one of the two putative cohesin-binding interfaces, complexed with one of the ScaA cohesins from *A. cellulolyticus* has been purified and crystallized, and data were processed to a resolution of 1.57 Å in the orthorhombic space group $P2_12_12_1$.

1. Introduction

The efficient degradation of plant cell walls is critical for carbon turnover and has a variety of both environmental and technological applications. Cellulosomes are multi-modular enzymatic molecules produced by anaerobic microorganisms containing a range of cellulases and hemicellulases that act synergistically to deconstruct plant structural carbohydrates. 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 scaffoldins themselves (Bayer *et al.*, 2004; Fontes & Gilbert, 2010).

The mesophilic, anaerobic, Gram-positive *Acetivibrio cellulolyticus*, which was first isolated from sewage sludge (Patel *et al.*, 1980), is well known for its ability to degrade both amorphous and crystalline cellulose (Sanchez *et al.*, 1999). The *A. cellulolyticus* cellulosomal system is very sophisticated, with 143 dockerin-containing proteins and 41 putative cohesin modules distributed across 16 scaffoldins (Dassa *et al.*, 2012). The main components in the organization of the *A. cellulolyticus* cellulosome are the primary enzyme-binding scaffoldin (ScaA), two anchoring scaffoldins (ScaC and ScaD) and an unusual adaptor scaffoldin (ScaB) (Xu *et al.*, 2004). Type I dockerins present in modular cellulosomal catalytic proteins bind to cohesins located in the primary scaffoldin, leading to cellulosome assembly (Carvalho *et al.*, 2003), whereas type II dockerins located in primary scaffoldins bind to adaptor scaffoldins, forming polycellulosomes, or to anchoring scaffoldins, contributing to the cell-surface attachment of the entire complex (Adams *et al.*, 2006). Analysis of type I dockerin sequences of *A. cellulolyticus* cellulosomal enzymes shows that it has two highly homologous 36 amino-acid residue tandem repeats (with 56% sequence identity), which suggests a possible dual binding mode. It is now well established that the presence of two cohesin-binding interfaces in dockerins directly interferes with crystallization (Brás *et al.*, 2012). Thus, in order to promote crystallization, we have engineered a double mutation in the predicted binding site of the first segment (Ser518 to Ile and Ile519 to Asn). The mutation of these two

© 2014 International Union of Crystallography
All rights reserved

Table 1
Macromolecule-production information.

Source organism	<i>A. cellulolyticus</i>
DNA source	Sythesized (NZYTech Ltd)
Cloning vector	pUC-57
Expression vector	pET-28a
Expression host	<i>E. coli</i>
Complete amino-acid sequences of the constructs produced	
AcGH5Doc†	MGSSHHHHHSSGLVPRGSHMASADVKPGVDGNGSININD- DFALMRNYLLGNLKDFFAEDDIKAGDLNGDKSINSLDFAI- MRMYLLGMITKFSV
AcScaACoh‡	MQTGFNLSIDTVEGNPGSSVVVPKLSGISKNGISTADFTVT- YDATKLEYISGDAGSIVTNPVNFINKESDGKLVFLD- YTMSTGYISTDGVFANLNFNIKSSAAIGSKAEVSI SGTPT- FGDSTLTPVVAKVTVNGAVNLE

† WP_010249057.1, region 502–573. The N-terminal tag is underlined. The double mutation is shown in italics. ‡ WP_010247249.1, region 1472–1611. The underlined residues are part of the *XhoI* restriction site (LEHHHHHH in the C-terminal tag).

probable key recognition residues, which are conserved across several *A. cellulolyticus* dockerins, is suggested to force the dockerin to interact with the cohesin *via* a single mode, leading to successful crystal formation. Here, we describe the expression, purification, crystallization and preliminary X-ray analysis of a type I complex of the ScaA cohesin (AcScaACoh) bound to the dockerin from family 5 glycoside hydrolase (AcGH5Doc).

2. Materials and methods

2.1. Macromolecule production

To co-express the *A. cellulolyticus* 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; Table 1). The cohesin-encoding gene was positioned at the 3' end and the dockerin-encoding gene at the 5' end. Between the two, the sequences of the T7 terminator (to terminate transcription of the dockerin gene) and the T7 promoter (to control transcription of the cohesin gene) were added (Cameron *et al.*, 2014; Bule *et al.*, 2014). 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 dockerin (by cutting with *NheI*–*SalI*; the primary sequence of the N-terminal engineered tag is MGSSHHHHHSSGLVPRGSHMAS)

Figure 1
SDS–PAGE [14%(w/v)] showing overexpression and purification of AC1_Doc. Lane 1, LMW protein marker (labelled in kDa); lanes 2–6, purified AcScaACoh (top bands, 14.7 kDa) and AcGH5Doc (bottom bands, 10.28 kDa).

Table 2
Crystallization.

Method	Sitting-drop vapour diffusion
Plate type	MRC 2-drop 96-well polystyrene crystallization plate
Temperature (K)	292
Protein concentration (mg ml ^{−1})	12/15.2
Buffer composition of protein solution	0.5 mM CaCl ₂
Composition of reservoir solution	0.2 M sodium thiocyanate, 20%(w/v) PEG 3350 pH 6.9
Volume and ratio of drop	1.4 µl (1:1)
Volume of reservoir (µl)	50

or at the C-terminus of the cohesin (by cutting with *NcoI* and *XhoI*; the primary sequence of the C-terminal engineered tag is LEHHHHHH). Thus, we can generate protein–protein complexes that contain a His tag either on the dockerin (Ac1_Doc) or on the cohesin (Ac1_Coh). Expression screens revealed that the cohesin-tagged complex failed to produce a significant yield of dockerin. Thus, the dockerin-tagged complex was selected for subsequent work.

It was expressed in *E. coli* Origami cells grown at 310 K in an autoinduction medium for 8 h and subsequently at 292 K until an OD₆₀₀ of approximately 10.0 was reached. 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 were 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 for the 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 column (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 with a NanoDrop 2000c spectrophotometer (Thermo Scientific, USA) using a molar extinction coefficient (ε) of 8940 M^{−1} cm^{−1}. The final protein concentration was adjusted to 12 and 15.2 mg ml^{−1} in 0.5 mM CaCl₂. This storage buffer is used because dockerins have two Ca²⁺-binding EF hands which are crucial for stabilizing their fold and thus maintaining their stability when in complex with a cohesin. The purity and molecular mass of the recombinant protein were analyzed 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 from Hampton Research (California, USA), 80 factorial and JCSG-*plus* HT96 screens (Molecular Dimensions, UK) were used for screening (Table 2). Crystal formation was observed in four different conditions within two weeks (Fig. 2; maximum dimensions ~120 × 100 × 30 µm). 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.

Figure 2
Crystals of Ac1_Doc obtained by sitting-drop vapour diffusion. They were gently prised apart from the cluster and data were collected from three of them. The largest crystals are approximately $120 \times 100 \times 30 \mu\text{m}$ in size.

Table 3

Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	ID29, ESRF
Wavelength (Å)	0.97625
Temperature (K)	100
Detector	PILATUS 6M
Crystal-to-detector distance (mm)	291.30
Rotation range per image (°)	0.1
Total rotation range (°)	140
Exposure time per image (s)	0.1
Space group	$P2_12_12_1$
a, b, c (Å)	46.69, 80.1, 112.6
Mosaicity (°)	0.23
Resolution range (Å)	1.57–46.69 (1.57–1.59)
Total No. of reflections	268193 (13657)
No. of unique reflections	59486 (2906)
Completeness (%)	99.3 (99.5)
Multiplicity	4.5 (4.7)
$\langle I/\sigma(I) \rangle$	16.4 (6.5)
$CC_{1/2}^\dagger$	0.998 (0.952)
$R_{\text{p.i.m.}}^\ddagger$	0.027 (0.098)
Overall B factor from Wilson plot (Å ²)	7.9

† $CC_{1/2}$ is the half-data-set correlation coefficient (Diederichs & Karplus, 2013). ‡ $R_{\text{p.i.m.}} = \sum_{hkl} [1/[N(hkl) - 1]]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $\langle I(hkl) \rangle$ is the average of symmetry-related observations of a unique reflection. Note that the resolution limit for the detector edge was set to collect data to 1.6 Å.

2.3. Data collection and processing

Data were collected on beamline ID29 at the European Synchrotron Radiation Facility, Grenoble, France. *iMosflm* (Battye *et al.*, 2011) was used for strategy calculation during data collection. All data sets were processed using *iMosflm* and *AIMLESS* (Evans, 2006) from the *CCP4* suite (Collaborative Computational Project, Number 4, 1994; Winn *et al.*, 2011). Data-collection statistics are given in Table 3. The Ac1_Doc crystals obtained from the condition 0.2 M sodium thiocyanate, 20% (w/v) PEG 3350 pH 6.9 (Fig. 2) diffracted beyond the detector limit and data were processed to a resolution of 1.57 Å. The crystals from the other three conditions did not diffract at all. The crystal belonged to the orthorhombic space group $P2_12_12_1$ with a dimer in the asymmetric unit ($V_M = 2.11 \text{ Å}^3 \text{ Da}^{-1}$ and a solvent content of 42%; Matthews, 1968) composed of two dockerins and two cohesins. *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 *Clostridium cellulolyticum* (PDB entries 2vn5 and 2vn6 with a sequence identity of 36.9% for the cohesin and 32.8% for the dockerin; Pinheiro *et al.*, 2008), which gave an R factor and R_{free} of 35.7 and 40.6%, respectively, and a Q factor of 0.719 after *REFMAC5* (Murshudov *et al.*, 2011) at the end of the *BALBES* run for a dimer of the cohesin–dockerin complex. An *ARP/wARP* (Langer *et al.*, 2008) run after *BALBES* gave a model of 400 residues in four chains, with an estimated correctness of 99.9%. Structure refinement and analysis are ongoing for the complex.

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 the ESRF, Grenoble for provision of synchrotron-radiation facilities (beamline ID29). We would also like to thank Cecília Bonifácio for help with crystallization, Cristiano de Sousa Mota for help with data collection and the European Community's Seventh Framework Programme (FP7/2007–2013) under BioStruct-X (grant agreement No. 283570, proposal No. Biostruct-X_4399) for funding.

References

- Adams, J. J., Pal, G., Jia, Z. & Smith, S. P. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 305–310.
- 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.
- Brás, J. L. A., Carvalho, A. L., Viegas, A., Najmudin, S., Alves, V. D., Prates, J. A. M., Ferreira, L. M. A., Romão, M. J., Gilbert, H. J. & Fontes, C. M. G. A. (2012). *Methods Enzymol.* **510**, 395–415.
- Bule, P., Ruimy-Israeli, V., Cardoso, V., Bayer, E. A., Fontes, C. M. G. A. & Najmudin, S. (2014). *Acta Cryst.* **F70**, 1061–1064.
- Cameron, K., Alves, V. D., Bule, P., Ferreira, L. M. A., Fontes, C. M. G. A. & Najmudin, S. (2014). *Acta Cryst.* **F70**, 656–658.
- Carvalho, A. L., Dias, F. M. V., Prates, J. A. M., Nagy, T., Gilbert, H. J., Davies, G. J., Ferreira, L. M. A., Romão, M. J. & Fontes, C. M. G. A. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 13809–13814.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Dassa, B., Borovok, I., Lamed, R., Henrissat, B., Coutinho, P., Hemme, C. L., Huang, Y., Zhou, J. & Bayer, E. A. (2012). *BMC Genomics*, **13**, 210.
- 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.
- 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.
- 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.
- Patel, G. B., Khan, A. W., Agnew, B. J. & Colvin, J. R. (1980). *Int. J. Syst. Bacteriol.* **30**, 179–185.
- Pinheiro, B. A., Proctor, M. R., Martinez-Fleites, C., Prates, J. A. M., Money, V. A., Davies, G. J., Bayer, E. A., Fontes, C. M. G. A., Fierobe, H.-P. & Gilbert, H. J. (2008). *J. Biol. Chem.* **283**, 18422–18430.
- Sanchez, C. R., Peres, C. S. & Barbosa, H. R. (1999). *Rev. Microbiol.* **30**, 310–314.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Xu, Q., Barak, Y., Kenig, R., Shoham, Y., Bayer, E. A. & Lamed, R. (2004). *J. Bacteriol.* **186**, 5782–5789.