Kate Cameron, Victor D. Alves,
Pedro Bule, Luís M. A. Ferreira,
Carlos M. G. A. Fontes* and
Shabir Najmudin*

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

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

Received 1 June 2012

Accepted 12 July 2012

Purification, crystallization and preliminary X-ray characterization of the *Acetivibrio cellulolyticus* type I cohesin ScaC in complex with the ScaB dockerin

The cellulosome, a highly elaborate extracellular multi-enzyme complex of cellulases and hemicellulases, is responsible for the efficient degradation of plant cell-wall carbohydrates by anaerobic microorganisms. Cohesin and dockerin recognition pairs are integral to the architecture of the cellulosome. Thus, type I cohesin:dockerins are important for attaching the modular enzymatic components to primary scaffoldins to form the cellulosome. In contrast, type II dockerins located in primary scaffoldins bind to anchoring scaffoldins, thus contributing to the cell-surface attachment of the entire complex. Since anchoring scaffoldins usually contain more than one type II cohesin, they contribute to the assembly of polycellulosomes. *Acetivibrio cellulolyticus* possesses an extremely 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 ScaB dockerin mutated to inactivate one of the two putative cohesin-binding interfaces complexed with the ScaC cohesin from *A. cellulolyticus* has been purified and crystallized and data were collected from tetragonal and monoclinic crystal forms to resolutions of 1.5 and 6.0 Å, respectively.

1. Introduction

The plant cell wall represents a major untapped global source of carbon and energy. Organisms in anaerobic ecosystems have evolved multi-enzyme complexes termed cellulosomes, which are comprised of a range of cellulases and hemicellulases that degrade the structural polysaccharides in a highly efficient and concerted way (Bayer *et al.*, 2004; Fontes & Gilbert, 2010). The assembly of cellulosomes occurs *via* highly ordered protein–protein interactions between cohesins, which are located in a macromolecular scaffold (primary scaffoldin), and dockerins, which are found in the enzymes. These high-affinity cohesin–dockerin interactions (type I) are thus responsible for cellulosome assembly. The cellulosomes in turn bind to the plant cell wall through a scaffoldin-borne carbohydrate-binding module. In addition, cellulosomes can also be anchored to the host microbial cell surface through the interaction of type II dockerins located in the primary scaffoldin and type II cohesins found on the cell envelope (Adams *et al.*, 2006; Pinheiro *et al.*, 2012). Here, we have investigated the intermolecular and intramolecular basis of the organization of cellulosomes by the mesophilic Gram-positive bacterium *Acetivibrio cellulolyticus*. The sequencing of the *A. cellulolyticus* cellulosomal gene cluster has shown that it contains four tandem scaffolding genes (*scaA*, *scaB*, *scaC* and *scaD*; Xu *et al.*, 2003, 2004). The primary scaffoldin ScaA contains a C-terminal type II dockerin domain similar to *Clostridium thermocellum* CipA. The ScaA dockerin binds to cohesins located in ScaB, which in turn binds to the anchoring scaffoldin ScaC *via* a third specific cohesin–dockerin interaction. A fourth scaffoldin, ScaD, is also located at the cell surface, but in contrast to ScaC it contains one cohesin that can bind enzymes

directly and two cohesins that are specific for ScaA. Thus, *A. cellulolyticus* has two anchoring scaffoldins: ScaC and ScaD. The multiple cohesins on each of the scaffoldins determine the total number of enzymes that can be incorporated in the highly ordered protein–protein complex. The ScaA–ScaB–ScaC complex can accommodate a total of 96 enzymes. In contrast, since the ScaD scaffoldin can bind

ScaA via its two type II cohesins and can bind to a single enzyme via its third type I cohesin, it provides a scaffold to organize a complex of 15 enzymes. ScaC and ScaD interact with the cell surface through S-layer homology (SLH) modules which enable the entire complex to be anchored to the cell surface by at least two alternative supra-structures (Xu *et al.*, 2004). However, structural and functional studies that provide evidence of the mechanism by which cohesin–dockerin interactions mediate both cellulosome assembly and cell-surface attachment are lacking in *A. cellulolyticus*.

To understand the mechanisms that modulate the different specificities expressed by *A. cellulolyticus* cohesin–dockerin complexes, we have expressed, purified and crystallized the type I complex established between the dockerin module of ScaB and the third cohesin module of ScaC of *A. cellulolyticus*. It has now been well established that type I dockerins usually display an internal twofold symmetry that supports the evolution of two similar cohesin-binding interfaces (Carvalho *et al.*, 2004, 2007). Residues that dominate cohesin recognition are usually located at positions 11 and 12 of either calcium-binding loop (Fontes & Gilbert, 2010). In ScaB dockerin these residues are conserved in the two duplicated segments (IN), suggesting that this type I dockerin contains two analogous cohesin-binding interfaces and as such should indeed express a dual binding mode (Xu *et al.*, 2004). To prevent the formation of a pool of heterogeneous complexes that could prevent complex crystallization, we engineered a mutation in the N-terminal helix (IN to SI) of the dockerin to promote a single binding mode and thereby facilitate crystallization (Fig. 1). Here, we describe the expression, purification and crystallization of the first *A. cellulolyticus* cohesin–dockerin complex. We present these results together with a description of the preliminary data collection.

Figure 1

A Coomassie Brilliant Blue-stained 14% PAGE evaluation of protein purity. Lane M, low-molecular-weight (LMW) Protein Marker (NZYTech Ltd); lane F1, AC3_coh separated into the cohesin (top) and dockerin (bottom).

Figure 2

Crystals of AC3_coh obtained by both the sitting-drop and the hanging-drop vapour-diffusion methods. (a) is the single (salt) crystal obtained in the initial screen from the conditions 0.5 M ammonium sulfate, 0.1 M HEPES pH 7.5, 30% (v/v) MPD; (b) and (c) are from optimization trials from the conditions 0.7 M ammonium sulfate, 0.1 M HEPES pH 7.5, 20% (v/v) MPD (b) and 0.7 M ammonium sulfate, 0.1 M HEPES pH 7.5, 35% (v/v) MPD (c). The crystals in (d), (e) and (f) were obtained in the original screening from conditions 0.1 M HEPES pH 7.5, 70% (v/v) MPD (d), 20% ethanol, 0.1 M Tris pH 8.5 (e) and 10% PEG 20K, 2% dioxane, 0.1 M Bicine pH 9.0 (f). The length of the scale bar represents ~200 µm.

2. Materials and methods

2.1. Protein expression and purification

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). To separate the sequences of the two genes, with the dockerin-encoding gene at the 5' end and the cohesin-encoding gene

at the 3' end, we introduced the sequence of the T7 terminator (to terminate transcription at the dockerin gene) and the T7 promoter (to control transcription of the cohesin gene). In addition, this construct contained specifically tailored *NheI* and *NcoI* recognition sites at the 5' end and *XhoI* and *SalI* sites at the 3' end to allow subcloning into pET28a (Novagen) such that a sequence encoding a six-residue His tag could be introduced either at the N-terminus of the dockerin (*NheI*–*SalI*) or at the C-terminus of the cohesin (*NcoI*–*XhoI*) (Brás *et al.*, 2012). Thus, by subcloning the initial construct in two different ways in pET28a we could generate protein–protein complexes that contained a His₆ tag either on the cohesin (AC3_coh) or on the dockerin (AC3_dok). Expression screens revealed that the cohesin-tagged complex gave higher expression levels and thus this was selected for subsequent work. The AC3_coh cohesin–dockerin complex was expressed in *E. coli* Tuner cells grown at 310 K to an OD₆₀₀ of 0.5. Recombinant protein expression was induced by the addition of isopropyl β -D-1-thiogalactopyranoside to a final concentration of 0.2 mM and incubation at 282 K for 16 h. Since the expression levels of the dockerin were much higher than those of the cohesin, we assumed that most of the cohesin molecules were complexed with dockerin. Thus, the recombinant complex and unbound cohesin were purified by immobilized metal-ion affinity chromatography (IMAC) using Sepharose columns charged with nickel (HisTrap) following conventional protocols (Najmudin *et al.*, 2006). Fractions containing the purified cohesin–dockerin complex were buffer-exchanged using PD-10 Sephadex G-25M gel-filtration columns (Amersham Pharmacia Biosciences) into 20 mM Tris–HCl pH 8.0 containing 5 mM CaCl₂. A further purification step consisting of anion-exchange chromatography was performed using a column loaded with Source 30Q media and a gradient elution with 0–1 M NaCl (Amersham Pharmacia Biosciences). This chromatography step enabled separation of the protein complex from unbound cohesin as a minor contaminant. Fractions containing the purified complex were then concentrated using Amicon Ultra-15 centrifugal devices with a 10 kDa cutoff membrane (Millipore) and washed three times with 1 mM CaCl₂. The final protein concentration was adjusted to 42 g l^{−1} in a storage buffer consisting of 1 mM CaCl₂.

2.2. Crystallization

Crystallization conditions were screened by the sitting-drop vapour-phase diffusion method using the commercial kits Crystal Screen, Crystal Screen 2, PEG/Ion, PEG/Ion 2 (Hampton Research, California, USA), Clear Strategy Screens I and II, MIDAS and JCSG-plus HT96 screens (Molecular Dimensions, UK) using an Oryx8 robotic nanodrop dispensing system (Douglas Instruments). Three drops of 0.7 μ l each with different dilutions (10, 20 and 42 mg ml^{−1} in 1 mM CaCl₂) of AC3_coh and 0.7 μ l reservoir solution were prepared at 292 K. Crystals were observed in five different conditions: Clear Strategy Screen II condition No. G5 [0.5 M ammonium sulfate, 0.1 M HEPES pH 7.5, 30%(v/v) ($\pm$)-2-methyl-2,4-pentanediol (MPD); Fig. 2a], Clear Strategy Screen II condition No. G11 [0.1 M HEPES pH 7.5, 70%(v/v) MPD; Fig. 2d], Clear Strategy Screen II condition No. H3 (0.2 M magnesium chloride hexahydrate, 0.1 M Tris pH 8.5, 3.4 M 1,6-hexanediol), JCSG-plus condition No. C8 (20% ethanol, 0.1 M Tris pH 8.5; Fig. 2e) and JCSG-plus condition No. C10 (10% PEG 20K, 2% dioxane, 0.1 M bicine pH 9.0; Fig. 2f). An optimization screen was carried out by the hanging-drop vapour-phase diffusion method using 1 μ l drops of 10, 20 or 42 mg ml^{−1} AC3_coh and 1 μ l reservoir solution at 292 K. The screen was based on the Clear Strategy Screen II hit 0.5 M ammonium sulfate, 0.1 M HEPES pH 7.5, 30%(v/v) MPD (Figs. 2b and 2c).

Figure 3 Representative diffraction patterns of (a) the best tetragonal AC3_coh crystal (the outer circle corresponds to 1.49 Å resolution) and (b) the monoclinic AC3_coh crystal form (the outer circle corresponds to 6.0 Å resolution). 151 frames of data with $\Delta\phi = 0.55^\circ$ were collected for the tetragonal crystal and 90 frames of data with $\Delta\phi = 2.0^\circ$ were collected for the monoclinic crystal at 100 K.

Table 1

Data-collection statistics.

Values in parentheses are for the lowest/highest resolution shells.

Data set	High resolution	Low resolution
Beamline	I04, Diamond	I04, Diamond
Space group	$P4_12_12$	$C2$
Wavelength (Å)	0.9795	0.9795
Unit-cell parameters		
a (Å)	107.75	202.50
b (Å)	107.75	68.78
c (Å)	100.81	137.55
β (°)	90	109.44
Resolution limits (Å)	76.19–1.49 (29.25–6.66/1.53–1.49)	129.71–6.50 (129.71–20.55/6.85–6.50)
No. of observations	641814 (7385/43638)	12447 (316/1931)
No. of unique observations	96788 (1234/7024)	3625 (115/530)
Multiplicity	6.6 (6.0/6.2)	3.4 (2.7/3.6)
Completeness (%)	99.9 (99.8/99.0)	98.4 (88.6/99.8)
$\langle I/\sigma(I) \rangle$	15.9 (33.8/2.7)	1.7 (1.4/1.1)
$R_{\text{merge}}^\dagger$ (%)	6.6 (3.6/58.7)	29.8 (14.4/83.5)
$R_{\text{p.i.m.}}^\ddagger$ (%)	3.1 (1.7/27.6)	20.0 (12.5/43.5)

$^\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the measurement of reflection hkl and $\langle I(hkl) \rangle$ is the mean value of $I_i(hkl)$ for all i measurements. $^\ddagger R_{\text{p.i.m.}} = \sum_{hkl} \{1/[N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ and is a measure of the quality of the data after averaging the multiple measurements.

These crystals (maximum dimensions $\sim 300 \times 200 \times 100 \mu\text{m}$; Fig. 2) were cryocooled in liquid nitrogen. In cases where the crystallization condition did not contain MPD, 30% glycerol was added to the crystallization buffer to make the solution a cryoprotectant.

2.3. Data collection and processing

Data were collected on beamline I04 at the Diamond Light Source, Harwell, England using a Quantum 315r charge-coupled device detector (ADSC) from a crystal cooled to 100 K using a Cryostream (Oxford Cryosystems Ltd). What we assumed to be the best crystal (Fig. 2a) turned out to be a salt crystal. However, the crystals obtained from an optimization screen based around this condition [0.5 M ammonium sulfate, 0.1 M HEPES pH 7.5, 30% (v/v) MPD; Fig. 2c] diffracted to resolutions between 3.0 and 1.49 Å (Fig. 3a). The crystals from the other conditions did not diffract at all or diffracted in the low-resolution range from 10 to 6 Å (Fig. 3b). The best of the crystals that diffracted to low resolution was obtained from the condition 10% PEG 20K, 2% dioxane, 0.1 M Bicine pH 9.0 (Fig. 2f). Currently, optimization screens have been set up for the other hits. 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 *MOSFLM* (Leslie, 1992) 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 (Collaborative Computational Project, Number 4, 1994; Winn *et al.*, 2011). The best protein crystal belonged to the tetragonal space group $P4_12_12$ or $P4_32_12$ and the low-resolution crystal belonged to the monoclinic space group $C2$. Data-collection statistics are given in Table 1. The Matthews coefficient ($V_M = 3.05$ or $2.03 \text{ Å}^3 \text{ Da}^{-1}$) indicated the presence of a dimer or a trimer in the

asymmetric unit for the high-resolution data and a solvent content of 60 or 40% (Matthews, 1968) and anything between a tetramer and a decamer for the low-resolution data. *BALBES* was used to carry out molecular replacement (Long *et al.*, 2008). A total of 42 structures with a sequence identity greater than 15% for either cohesion or dockerin or both modules were found. The best solution was found using the mutant cohesin–dockerin complex from *C. thermocellum* (PDB entry 2ccl; Carvalho *et al.*, 2007). This gave a dimer in space group $P4_12_12$ with a final R factor and R_{free} of 39.8% and 41.2%, respectively, and a Q -factor of 0.706 after *REFMAC5* at the end of the *BALBES* run. Structure completion and analysis are ongoing.

This work was supported in part by Fundação para a Ciência e a Tecnologia (Lisbon, Portugal) through grants PTDC/BIA-PRO/103980/2008 and PTDC/QUI-BIO/100359/2008 and by the European Union Seventh Framework Programme (FP7 2007–2013) under the WallTraC project (Grant Agreement No. 263916). 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. The authors would like to thank Dr Pierpaolo Romano for help with data collection at Diamond Light Source.

References

- Adams, J. J., Pal, G., Jia, Z. & Smith, S. P. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 305–310.
- Bayer, E. A., Belaich, J.-P., Shoham, Y. & Lamed, R. (2004). *Annu. Rev. Microbiol.* **58**, 521–554.
- Brás, J. L., 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.
- 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.
- Carvalho, A. L., Goyal, A., Prates, J. A. M., Bolam, D. N., Gilbert, H. J., Pires, V. M. R., Ferreira, L. M. A., Planas, A., Romão, M. J. & Fontes, C. M. G. A. (2004). *J. Biol. Chem.* **279**, 34785–34793.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- 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.
- Leslie, A. G. W. (1992). *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **26**.
- 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.
- 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.
- Pinheiro, B. A., Brás, J. L. A., Najmudin, S., Carvalho, A. L., Ferreira, L. M. A., Prates, J. A. M. & Fontes, C. M. G. A. (2012). *Biocatal. Biotrans.* **30**, 309–315.
- 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.
- Xu, Q., Barak, Y., Kenig, R., Shoham, Y., Bayer, E. A. & Lamed, R. (2004). *J. Bacteriol.* **186**, 5782–5789.
- Xu, Q., Gao, W., Ding, S.-Y., Kenig, R., Shoham, Y., Bayer, E. A. & Lamed, R. (2003). *J. Bacteriol.* **185**, 4548–4557.