

**Markus Alahuhta, Qi Xu, Roman
 Brunecky, William S. Adney,
 Shi-You Ding, Michael E.
 Himmel and Vladimir V. Lunin***

BioSciences Center, National Renewable Energy
 Laboratory, 1617 Cole Boulevard, Golden,
 Colorado 80401-3305, USA

Correspondence e-mail:
 vladimir.lunin@nrel.gov

Received 29 April 2010
 Accepted 11 June 2010

PDB Reference: fibronectin type III-like module,
 3mpc.

Structure of a fibronectin type III-like module from *Clostridium thermocellum*

The 1.6 Å resolution structure of a fibronectin type III-like module from *Clostridium thermocellum* (PDB code 3mpc) with two molecules in the asymmetric unit is reported. The crystals used for data collection belonged to space group $P2_12_12_1$, with unit-cell parameters $a = 35.43$, $b = 45.73$, $c = 107.72$ Å, and the structure was refined to an R factor of 0.166. Structural comparisons found over 800 similar structures in the Protein Data Bank. The broad range of different proteins or protein domains with high structural similarity makes it especially demanding to classify these proteins. Previous studies of fibronectin type III-like modules have indicated that they might function as ligand-binding modules, as a compact form of peptide linkers or spacers between other domains, as cellulose-disrupting modules or as proteins that help large enzyme complexes remain soluble.

1. Introduction

Utilization of cellulosomes is a strategy that is employed by some bacteria to deconstruct and metabolize cellulosic substrates (Bayer *et al.*, 2008). In the cellulosomal system, multi-domain enzymes are attached to scaffoldin proteins that are bound to the bacterial cell wall. Cellulosomes are known to contain carbohydrate-binding modules (CBMs), glycoside hydrolases (GHs) and some domains of unknown function. These modules are usually annotated as X-domains, X-modules or fibronectin type III-like [Fn(III)-like] modules.

Bacterial Fn(III)-like modules are common but have not been greatly studied (Little *et al.*, 1994). They can be expressed separately from the cellulosomal multi-module enzymes they are part of as independently folding proteins, but do not show any activity towards cellulose (Watanabe *et al.*, 1994). Some studies support the hypothesis that Fn(III)-like modules are cellulose disruptors that improve the hydrolytic ability of cellulases (Kataeva *et al.*, 2002). Another hypothesis suggests that they are a 'storage form' of the linker peptide that connects other modules of the enzyme that can be extended upon demand or that they may simply function as spacers between the other domains (Jee *et al.*, 2002).

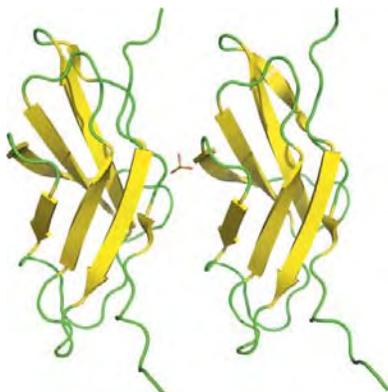
It is our long-term strategy to understand the structure–function relationships of bacterial cellulosomal proteins in order to better understand how they affect cellulose degradation and their specific function in the cellulosomes. With this publication, we open a series of structural and functional studies on the critical cellulosomal domains from *Clostridium thermocellum*.

2. Materials and methods

2.1. Cloning, expression and purification

The Fn(III)-like module studied here is from the *C. thermocellum* cellulosomal cellulase gene (accession No. ZP_00510548). The overall architecture of this gene is GH5-CBM6-[Fn(III)-like]-(type I dockerin). The expressed protein contained an N-terminal methionine and a C-terminal histidine tag with sequence LEHHHHHH.

The *C. thermocellum* Fn(III)-like module was amplified by polymerase chain reaction (PCR) using CAATACCCATGGTTTCCGCTCCCGGTTTC and GGTACCCTCGAGTTTAGGATATTGC-



AGTATCG primers (the sequences in bold indicate the restriction sites) and genomic DNA of *C. thermocellum* as a template. The PCR fragment was inserted into plasmid pET28b (Novagen, Madison, Wisconsin, USA) via *XhoI* to generate the expression plasmid and its sequence was confirmed by DNA sequencing. The clone was over-expressed in *Escherichia coli* BL21 (DE3) strain (Stratagene, La Jolla, California, USA) at 289 K and grown in the presence of 0.3 mM isopropyl β -D-1-thiogalactopyranoside. The expressed protein had 103 residues and a molecular weight of 10 869 Da. The QIAexpress Ni-NTA protein-purification system (Qiagen, Valencia, California, USA) was used for preliminary purification of the fusion protein following the manufacturer's recommended protocol. The crude protein eluate from the histidine-affinity column was first desalted into acetate buffer pH 5 using a HiPrep 26/10 desalting column (GE Life Sciences). Subsequently, the eluate was concentrated and exchanged into acetate buffer pH 5 containing 2 M ammonium sulfate followed by hydrophobic interaction chromatography using a Source 15PHE column (GE Life Sciences).

2.2. Crystallization

The structure was solved using two crystals: a native crystal and an iodine derivative. Native crystals were obtained by sitting-drop vapor diffusion with Crystal Screen (Hampton Research, Aliso Viejo, California, USA) using a 96-well plate. 50 μ l well solution was used with drops containing 1 μ l well solution and 1 μ l protein solution. The crystals were grown at 298 K in 1.6 M ammonium sulfate, 0.1 M MES monohydrate pH 6.5 buffer and 10% (v/v) 1,4-dioxane; the protein was at 37 mg ml⁻¹ in 20 mM acetic acid pH 5 buffer and 100 mM NaCl. Iodine-derivative crystals were grown using the hanging-drop vapor-diffusion method on a 24-well plate with 1 ml well solution and drops containing 1 μ l well solution and 1 μ l protein solution. The crystal grew at 298 K in 0.1 mM MES monohydrate pH 6 buffer and

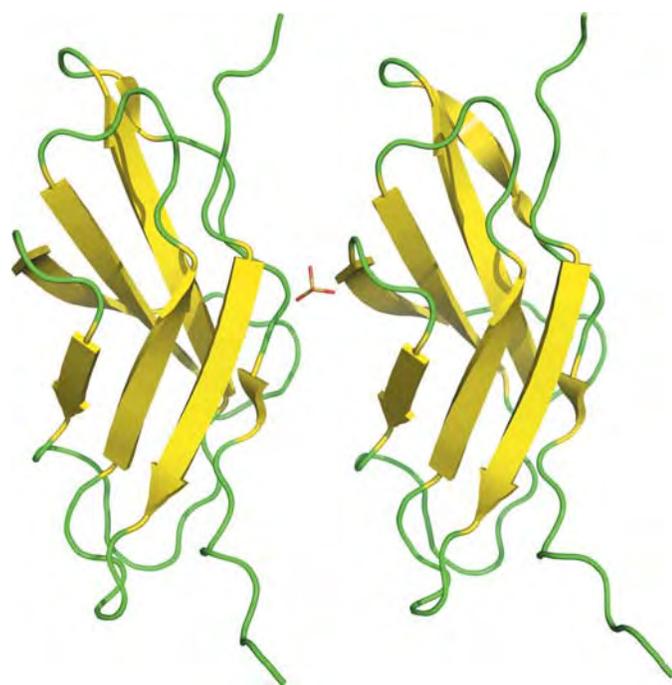


Figure 1

The two molecules in the asymmetric unit. β -Strands are shown in yellow, loops are shown in green and the sulfate ion is shown in stick representation with a yellow S atom and red O atoms.

Table 1

X-ray data-collection and refinement statistics.

Values in parentheses are for the highest resolution bin.

	Native	Iodine derivative
Data collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (\AA , $^\circ$)	$a = 35.43$, $b = 45.73$, $c = 107.72$, $\alpha = \beta = \gamma = 90.0$	$a = 35.69$, $b = 46.64$, $c = 107.35$, $\alpha = \beta = \gamma = 90.0$
Wavelength (\AA)	1.54178	1.54178
Temperature (K)	100	100
Resolution (\AA)	25–1.6 (1.65–1.6)	35–2.0 (2.1–2.0)
Unique reflections	23168	23146
Observed reflections	118620	169892
R_{merge}^\dagger	0.043 (0.141)	0.052 (0.199)
Redundancy	5.12 (2.36)	7.34 (3.75)
$I/\sigma(I)$	24.87 (6.26)	27.6 (5.64)
Completeness (%)	96.8 (94.8)	99.3 (99.0)
Refinement		
R/R_{free}	0.166 (0.224)/0.196 (0.298)	
Protein atoms	1479	
Water molecules	223	
Other atoms	5	
R.m.s.d. bond lengths (\AA)	0.023	
R.m.s.d. bond angles ($^\circ$)	1.980	
Average B factors (\AA^2)		
Protein atoms	11.844	
Water molecules	23.468	
Ramachandran plot statistics (%)		
Allowed	100	
Favored	97.87	
Outliers	0	

$^\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an individual reflection and $\langle I(hkl) \rangle$ is the mean intensity of a group of equivalent reflections.

2.4 M ammonium sulfate; the protein was at 37 mg ml⁻¹ in 20 mM acetic acid pH 5 and 100 mM NaCl.

2.3. Data collection and processing

Prior to data collection, the iodine-derivative crystal was transferred into 5 μ l well solution containing 100 mM potassium iodide and incubated for 30 min in a sealed hanging-drop setup. Both the native and the iodine-treated crystals were soaked for 1 min in a 5 μ l drop of well solution containing 20% ethylene glycol and flash-frozen in a cold nitrogen-gas stream at 100 K. The data collections were performed using a Bruker X8 MicroStar X-ray generator with Helios mirrors and a Bruker PLATINUM¹³⁵ CCD detector. Data were indexed and processed with the *Bruker Suite* of programs v.2008.1-0 (Bruker AXS, Madison, Wisconsin, USA).

2.4. Structure solution and refinement

Intensities were converted into structure factors and 5% of the reflections were flagged for R_{free} calculations using the programs *F2MTZ*, *TRUNCATE*, *CAD* and *UNIQUE* from the *CCP4* package of programs (Collaborative Computational Project, Number 4, 1994). The iodine-site coordinates were identified by SIRAS (single isomorphous replacement with anomalous signal; Dauter *et al.*, 2000) using the *SHELXC/D/E* suite of programs (Sheldrick, 2008). Diffraction data in the resolution range 25–2.3 \AA were used for substructure determination. Nine iodine sites were identified with occupancies ranging from 1.00 to 0.365 and a CC (correlation coefficient) of 30.34% for all reflections (19.81% for weak reflections). The resulting set of phases obtained by *SHELXE* yielded a map with a CC of 80.5%, a contrast of 0.869 and a connectivity of 0.933 at 1.6 \AA resolution. *ARP/wARP* (Cohen *et al.*, 2008) v.7.0 was used to build the initial model using the previously calculated phases. Further

refinement and manual correction was performed using *REFMAC5* v.5.5.0109 (Murshudov *et al.*, 1997) and *Coot* v.0.6 (Emsley & Cowtan, 2004). Noncrystallographic symmetry restraints between the two molecules in the asymmetric unit were used in *REFMAC5* in all refinement cycles. The *MolProbity* method (Chen *et al.*, 2010) was used to analyze the Ramachandran plot. The data-collection and refinement statistics are shown in Table 1.

3. Results and discussion

We have solved the X-ray structure of a fibronectin III-like module from *C. thermocellum* at 1.6 Å resolution. It has two molecules in the asymmetric unit with a sulfate ion bound between them (Fig. 1). There are no significant differences between the two molecules (root-mean-square deviation of 0.65 Å for all C^α atoms). The overall fold is a canonical Fn(III)-like fold with three-stranded and four-stranded antiparallel β-sheets. Because of weak electron density, two residues at the N-terminal end (methionine and valine) and five histidines at the C-terminal end were not modelled. This structure has been deposited in the Protein Data Bank with code 3mpc.

Structural comparison using the *DALI* search tool (Holm *et al.*, 2008) found 850 structures with a *Z* score of 5 or higher. From the highest scoring hits we were able to find two bacterial Fn(III)-like modules. These were the modules in chitinase A1 from *Bacillus circulans* WL-12 (PDB code 1k85; Jee *et al.*, 2002) and in β-*N*-acetylglucosaminidase from *C. perfringens* (PDB code 2w1n; Ficko-Blean *et al.*, 2009). Closer inspection of these structures shows that they have practically identical folds, with root-mean-square deviations of 1.81 Å for 1k85 and 1.65 Å for 2w1n (all C^α atoms were used). Specifically, aromatic residues inside the protein, such as Trp35 and Tyr26, are identical in all three structures.

The 850 similar structures found by the *DALI* search show how common this fold is in various proteins. The hits included fibronectins, immunoglobulin-like domains, carbohydrate-binding modules and many other proteins with the classic jelly-roll fold. At the same time, the sequence identities of these structurally similar proteins were very low: the highest score we observed in our *DALI* search was 27%. The high structural similarity combined with low sequence identity

makes the classification of these modules very challenging. It seems that this highly abundant fold is used for many different purposes. For example, it has been proposed to be a ligand-binding module, a compact form of peptide linkers or spacers between other domains (Jee *et al.*, 2002), a cellulose-disrupting module (Kataeva *et al.*, 2002) or a protein that assists large enzyme complexes to remain soluble. Further studies are necessary in order to understand the function of this Fn(III)-like module from *C. thermocellum*.

This work was supported by the DOE Office of Science, Office of Biological and Environmental Research through the BioEnergy Science Center (BESC), a DOE Bioenergy Research Center.

References

- Bayer, E. A., Lamed, R., White, B. A. & Flint, H. J. (2008). *Chem. Rec.* **8**, 364–377.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Cohen, S. X., Ben Jelloul, M., Long, F., Vagin, A., Knipscheer, P., Lebbink, J., Sixma, T. K., Lamzin, V. S., Murshudov, G. N. & Perrakis, A. (2008). *Acta Cryst.* **D64**, 49–60.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* **D56**, 232–237.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Ficko-Blean, E., Gregg, K. J., Adams, J. J., Hehemann, J. H., Czjzek, M., Smith, S. P. & Boraston, A. B. (2009). *J. Biol. Chem.* **284**, 9876–9884.
- Holm, L., Kaariainen, S., Rosenstrom, P. & Schenkel, A. (2008). *Bioinformatics*, **24**, 2780–2781.
- Jee, J.-G., Ikegami, T., Hashimoto, M., Kawabata, T., Ikeguchi, M., Watanabe, T. & Shirakawa, M. (2002). *J. Biol. Chem.* **277**, 1388–1397.
- Kataeva, I. A., Seidel, R. D. III, Shah, A., West, L. T., Li, X.-L. & Ljungdahl, L. G. (2002). *Appl. Environ. Microbiol.* **68**, 4292–4300.
- Little, E., Bork, P. & Doolittle, R. F. (1994). *J. Mol. Evol.* **39**, 631–643.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240–255.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- Watanabe, T., Ito, Y., Yamada, T., Hashimoto, M., Sekine, S. & Tanaka, H. (1994). *J. Bacteriol.* **176**, 4465–4472.