Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

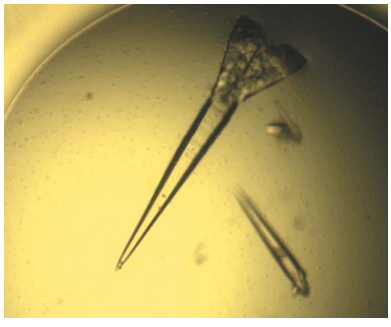
Influence of the N-terminal peptide on the cocrystallization of a thermophilic endo-β-1,4-glucanase with polysaccharide substrates

Baisong Zheng,[a] Wen Yang,[b] Yuguo Wang,[a] Zhiyong Lou[b]* and Yan Feng[a,c]*

[a]Key Laboratory for Molecular Enzymology and Engineering of Ministry of Education, Jilin University, Changchun 130023, People's Republic of China, [b]Laboratory of Structural Biology, School of Medicine, Tsinghua University, Beijing 100084, People's Republic of China, and [c]State Key Laboratory of Microbial Metabolism, College of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China

Correspondence e-mail: louzy@xtal.tsinghua.edu.cn, yfeng2009@mail.sjtu.edu.cn

Received 20 June 2011
Accepted 17 August 2011



It is well known that protein cocrystallization is affected by several parameters such as the ratio of the protein to the ligand, the reservoir solution, the pH and the temperature. Previously, spatial blocking by the N-terminus was observed in the active site in the crystal structure of the native protein of a thermostable endoglucanase from the thermophilic bacterium *Fervidobacterium nodosum* Rt17-B1 (FnCel5A). It was speculated that the N-terminal α-helix might form interactions with the substrate-binding residues and it was believed that this spatial block is special to some extent. In order to confirm the effect on cocrystallization, two N-terminally truncated variants of FnCel5A were constructed, purified and cocrystallized at 291 K. A crystal of FnCel5AND_12–343 in complex with cellobiose was obtained using PEG 8000 as a precipitant. A 2.2 Å resolution data set was collected. This crystal form (space group $P4_12_12$, unit-cell parameters $a = b = 47.3$, $c = 271.4$ Å) differed from that of the native protein. One molecule is assumed to be present per asymmetric unit, which gives a Matthews coefficient of 2.05 Å^3 Da^{-1}.

1. Introduction

Cellulose from dedicated energy crops and agricultural waste is a potential carbon source for the next generation of liquid transportation fuels, including cellulosic ethanol and more complex advanced biofuels (Ragauskas *et al.*, 2006; Zhang *et al.*, 2006; Chang, 2007; Rubin, 2008; Lou *et al.*, 2010; Wang *et al.*, 2010; Zhang *et al.*, 2010; Liu *et al.*, 2010). However, this biopolymer is recalcitrant to degradation into the glucose monomers required for fermentation. Even after an energy-intensive pre-treatment step in which the noncellulosic cell-wall components are stripped away from the cellulose, the bonds between glucose monomers must be hydrolyzed. One commonly used method is the enzymatic hydrolysis of cellulose, which has been studied for several decades (Kim *et al.*, 1987). Enzymatic hydrolysis of cellulose into glucose monomers requires the synergistic activity of several enzymes. Endoglucanases cleave random glycosidic bonds within the cellulose polymer, specific exoglucanases cleave cellobiose molecules from the reducing or nonreducing ends of the cellulose polymer and β-glucosidases hydrolyze cellobiose into glucose. The synergism of these enzymes has been reviewed elsewhere (Zhang *et al.*, 2006).

Although enzymes from different glycosidase families have little or no overall sequence homology, their catalytic domains are expected to share the same $(\beta/\alpha)_8$-barrel topology as observed in other enzymes belonging to these glycosidase families. Crystal structures have been reported for members of glycosidase families 1, 2, 5, 10 and 17. The active site is located at the carboxyl-terminal end of the β-strands, which is characteristic of all known family 5 members. Two strictly conserved glutamic acids located at the carboxyl-termini of β-strands 4 and 7 have been identified as the proton donor and the nucleophile, respectively, and play an important role in catalysis. Several aromatic and polar groups form the surface of a deep extended substrate-binding cleft that can accommodate at least four D-glucosyl subsites: two on each side of the labile glycosidic bond.

Fervidobacterium nodosum Rt17-B1, a hyperthermophilic hot-spring bacterium with an optimum temperature of 333 K, was first isolated from a hot spring in New Zealand. In 2007, genome sequencing of *F. nodosum* Rt17-B1 was completed by the US DOE

Joint Genome Institute and several endoglucanases have been identified in the genome sequence (accession No. NC_009718). One of them, FnCel5A, is remarkably active and stable at high temperatures. Furthermore, it has activity towards a broad range of substrates containing β-1,4-linked polysaccharides, including barley β-glucan, carboxymethyl cellulose, xyloglucan, lichenin, galactomannan and glucomannan (Wang *et al.*, 2010). These properties make FnCel5A an attractive potential target for protein engineering to improve its activity towards cellulose.

Structural information for FnCel5A will be valuable for the bioengineering of cellulases using structure-based experiments in order to produce the more thermally resistant and/or catalytically efficient enzymes required for biofuel production. We have previously determined the native structure of full-length FnCel5A (unpublished work). All attempts to soak or cocrystallize full-length FnCel5A with polysaccharide substrates in order to solve the structure of the enzyme–substrate complex have so far been unsuccessful. Since wild-type FnCel5A is predicted to have a flexible N-terminal end which may hinder cocrystallization with substrate, two N-terminal deletion variants were designed. The two variants are termed FnCel5AND_7–343 and FnCel5AND_12–343 as they delete the N-terminal six and 11 amino acids of FnCel5A, respectively. In this report, we describe the cloning, expression, purification and preliminary crystallographic analysis of these two N-terminally truncated variants. We found that only the FnCel5AND_12–343 mutant could be crystallized with substrates in the reservoir solution (23% PEG 8000, 0.2 *M* ammonium sulfate, 0.1 *M* sodium citrate pH 5.6). A complete diffraction data set for FnCel5AND_12–343 cocrystallized with cellobiose has been collected for structure determination and refinement.

2. Cloning, expression and purification

Two truncated mutants, FnCel5AND_7–343 and FnCel5AND_12–343, were generated by PCR using plasmid pET15b with the FnCel5A gene as a template in our laboratory. The oligonucleotides used as the forward and reverse primers to obtain the mutants were 5′-CAGCGC**CATATG**GTTGATAAAAGT-3′ (forward for FnCel5AND_7–343), 5′-CAGCGC**CATATG**GCTTTTGAATAT-3′ (forward for FnCel5AND_12–343) and 5′-GACGTC**GGATCC**TTATTTTCCAAG-3′ (reverse for both). The primers included *Nde*I and *Bam*HI restriction sites (shown in bold). The amplification reactions consisted of 30 cycles of 367 K for 1 min, 328 K for 1 min and 345 K for 2 min. After amplification, the purified PCR products were digested with *Nde*I and *Bam*HI and then inserted into the expression vector pET11a with the same digestion sites. After transformation into *Escherichia coli* strain XL1-Blue, the cloned fragments were completely sequenced. The recombinant plasmid was transformed into *E. coli* strain BL21 (DE3). The transformed cells were then cultured at 310 K in LB medium containing 50 mg ml^{-1} ampicillin. When the culture density reached an A_{600} of 0.6–0.8, induction with 1 m*M* IPTG was performed and cell growth continued for 4 h at 310 K. Cells were harvested by centrifugation, resuspended in 20 m*M* Tris–HCl pH 8.0 buffer containing 10 m*M* NaCl and then homogenized by sonication. Crude bacterial extracts were subjected to heat incubation at 338 K for 30 min and centrifuged at 20 000*g* for 30 min to remove heat-aggregated proteins and cell debris. The supernatant obtained by centrifugation for 30 min at 20 000*g* was applied onto a HiTrap Q-Sepharose column. Contaminant protein was thoroughly washed off with at least ten bed volumes of wash buffer (20 m*M* Tris–HCl pH 8.0, 10 m*M* NaCl) and the target protein was eluted with a linear salt-gradient buffer (0–1 *M* NaCl) at a flow rate of 60 ml h^{-1}. The target fractions were collected and desalted. Resource Q anion-exchange chromatography (GE Healthcare, USA) was subsequently applied using a 0–1 *M* gradient of NaCl in 20 m*M* Tris–HCl pH 8.0 buffer. The target protein was finally eluted with approximately 0.2 *M* NaCl. The concentrated target protein was finally loaded onto a Superdex 200 10/300 GL column and eluted with 20 m*M* Tris–HCl buffer pH 8.0 containing 10 m*M* NaCl. The purity of the protein was estimated to be greater than 99% by SDS–PAGE analysis (Fig. 1). To calculate the protein concentration, we used the formula $[E] = (1.5\,\mathrm{OD}_{280} - 0.75\,\mathrm{OD}_{260})$/extinction coefficient. The approximate extinction coefficient for both FnCel5AND_7–343 and FnCel5AND_12–343 was 96 955 M^{-1} cm^{-1}.

3. Crystallization

The purified FnCel5AND_7–343 and FnCel5AND_12–343 proteins were concentrated to ∼20 mg ml^{-1} in 20 m*M* Tris–HCl pH 8.0 and 10 m*M* NaCl. The molar ratio of protein to polysaccharide substrate (cellobiose or carboxymethyl cellulose) was 1:1.5. Cocrystallization was performed by the hanging-drop vapour-diffusion method at 291 K in 16-well plates. Each drop consisted of 1 µl protein solution and 1 µl reservoir solution, with 200 µl reservoir solution in the well. Screening was carried out with Hampton Research Crystal Screen kits and positive hits were optimized. Initial crystals were obtained from condition No. 15 of Crystal Screen (30% PEG 8000, 0.2 *M* ammonium sulfate, 0.1 *M* sodium cacodylate trihydrate pH 6.5). Further crystallization optimization was performed by carefully

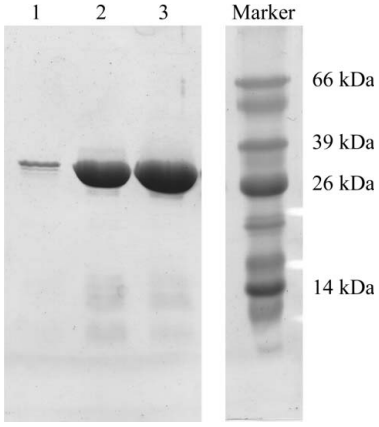


Figure 1
SDS–PAGE of purified FnCel5AND_12–343.

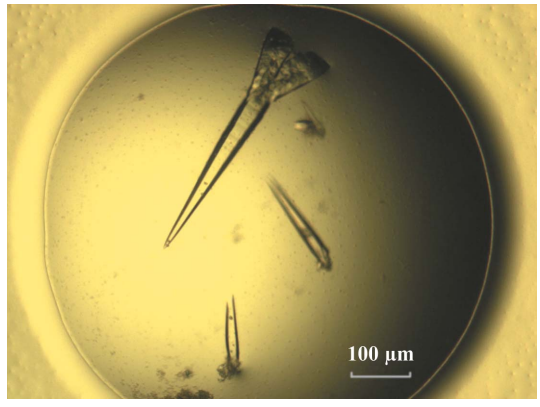


Figure 2
Typical crystals of FnCel5AND_12–343.

Table 1
Data-collection and refinement statistics for FnCel5AND_12–343.

Values in parentheses are for the highest resolution shell.

Unit-cell parameters (Å)	a = 47.3, b = 47.3, c = 271.4
Space group	$P4_12_12$
Wavelength (Å)	1.5418
Resolution (Å)	50.0–2.2 (2.28–2.20)
Total No. of reflections	117210 (9629)
No. of unique reflections	16724 (1632)
Completeness (%)	98.9 (99.5)
Average $I/\sigma(I)$	19.2 (2.7)
$R_{\rm merge}$† (%)	9.5 (47.9)
Molecules per asymmetric unit	1
$V_{\rm M}$ (Å^3 Da^{-1})	2.05
Sovent content (%)	40

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean intensity of the observations $I_i(hkl)$ of reflection hkl.

adjusting the concentration of PEG 8000 and the buffer pH value together with the protein concentration. Single crystals with dimensions of 0.4 × 0.1 × 0.05 mm (Fig. 2) were obtained from an optimized reservoir solution (23% PEG 8000, 0.2 *M* ammonium sulfate, 0.1 *M* sodium citrate pH 5.6) within 2 d and these crystals diffracted to 2.2 Å resolution on an in-house Rigaku MicroMax-007 X-ray source.

4. Data collection and processing

Native diffraction data were collected on a MAR345dtb (MAR Research, Hamburg) image-plate detector at 100 K using a Rigaku MicroMax-007 rotating-anode home X-ray generator operated at 40 kV and 20 mA (λ = 1.5418 Å). The crystal was mounted on a nylon loop and flash-cooled in a nitrogen-gas cryostream at 100 K using an Oxford Cryosystem. Crystals were cryoprotected by the addition of 10%(*v*/*v*) glycerol to the crystallization conditions. A total of 260 frames of data were collected with a 0.5° oscillation range. All intensity data were indexed, integrated and scaled with the *HKL*-2000 package (Otwinowski & Minor, 1997). The crystal belonged to space group $P4_12_12$, with unit-cell parameters a = 47.3, b = 47.3, c = 271.4 Å. We assumed the presence of one molecule per asymmetric unit, which gives a Matthews coefficient of 2.05 Å^3 Da^{-1} with 40% solvent content (Matthews, 1968). Complete data-collection statistics are given in Table 1. Molecular replacement was performed with the crystal structure of wild-type FnCel5A as the initial search model. This procedure was performed using *CNS* v.1.2 (Brünger *et al.*, 1998) and *Phaser* (McCoy *et al.*, 2007). Molecular-replacement results confirmed the space group to be $P4_12_12$. Structure refinement is ongoing and attempts at cocrystallization with other substrates are being made. The final statistics for data collection and processing are summarized in Table 1. The results suggest that our cocrystallization was successful because the initial substrate cellobiose was cleaved to form glucose in the active site.

This work was supported by the National Basic Research Program of China (973 Program), the Natural Science Foundation of China (31000359), the High Technology Project of China (863 Program) and the Science and Technology Commission of Shanghai Municipality.

References

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Chang, M. C. Y. (2007). *Curr. Opin. Chem. Biol.* **11**, 677–684.
Kim, J.-M., Kong, I.-S. & Yu, J.-H. (1987). *Appl. Environ. Microbiol.* **53**, 2656–2659.
Liu, C., Yang, G., Tian, G., Zhang, Z. & Feng, Y. (2010). *Protein Cell*, **2**, 497–506.
Lou, Z., Li, M., Sun, Y., Liu, Y., Liu, Z., Wu, W. & Rao, Z. (2010). *Protein Cell*, **1**, 760–770.
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.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., Eckert, C. A., Frederick, W. J., Hallett, J. P., Leak, D. J., Liotta, C. L., Mielenz, J. R., Murphy, R., Templer, R. & Tschaplinski, T. (2006). *Science*, **311**, 484–489.
Rubin, E. M. (2008). *Nature (London)*, **454**, 841–845.
Wang, Y., Wang, X., Tang, R., Yu, S., Zheng, B. & Feng, Y. (2010). *J. Mol. Catal. B Enzym.* **66**, 294–301.
Wang, J., Yan, X., Zhou, C. & Xu, J. (2010). *Acta Biophys. Sin.* **26**, 472–480.
Zhang, S., Du, Y., Wang, Y. & Liu, P. (2010). *Acta Biophys. Sin.* **26**, 97–105.
Zhang, Y.-H. P., Himmel, M. E. & Mielenz, J. R. (2006). *Biotechnol. Adv.* **24**, 452–481.