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

Crystallization and preliminary diffraction studies of CBM3b of cellobiohydrolase 9A from *Clostridium thermocellum*

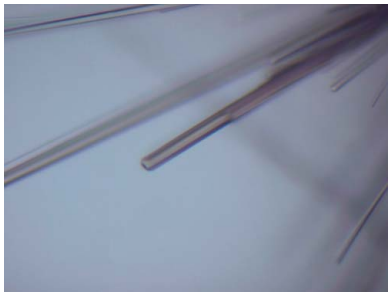
Sadanari Jindou,[a]‡ Svetlana Petkun,[a,b]‡ Linda Shimon,[c] Edward A. Bayer,[d] Raphael Lamed[a,b] and Felix Frolow[a,b]*

[a]Department of Molecular Microbiology and Biotechnology, Tel Aviv University 69978, Israel, [b]The Daniella Rich Institute for Structural Biology, Tel Aviv University 69978, Israel, [c]Department of Chemical Research Support, The Weizmann Institute of Science, Rehovot 76100, Israel, and [d]Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel

‡ These authors made an equal contribution.

Correspondence e-mail: mbfrolow@post.tau.ac.il

Received 2 September 2007
Accepted 30 October 2007



Family 3 carbohydrate-binding modules (CBM3s) are associated with the scaffoldin subunit of the multi-enzyme cellulosome complex and with the family 9 glycoside hydrolases, which are multimodular enzymes that act on plant cell-wall polysaccharides, notably cellulose. Here, the crystallization of CBM3b from cellobiohydrolase 9A is reported. The crystals are tetragonal and belong to space group $P4_1$ or $P4_3$. X-ray diffraction data for CBM3b have been collected to 2.68 Å resolution on beamline ID14-4 at the ESRF.

1. Introduction

Cellulose is a major component of the plant cell wall, lending structural stability and strength to an otherwise flexible and pendulous material. In nature, cellulose is typically found associated with xylan and other hemicelluloses and is both insoluble and chemically unreactive. Numerous enzymes acting synergistically are needed to efficiently degrade the more crystalline forms of cellulose. For this reason, cellulose-degrading organisms produce a remarkable variety of enzymes that either exist in the free state or in association with a multi-enzyme complex known as the cellulosome (Bayer *et al.*, 2004).

Family 3 carbohydrate-binding modules (CBM3s) are the major targeting entity borne by the cellulosomal scaffoldin subunit and are are responsible for delivering the enzyme complex to the crystalline cellulose substrate. In addition, the CBM3s are also found associated with family 9 glycoside hydrolases (GH9), which are multimodular enzymes that act on plant cell-wall polysaccharides, notably cellulose (Bayer, Shimon *et al.*, 1998). There are three major subgroups in the CBM3 family, 3a, 3b and 3c, which are defined on the basis of sequence similarity (Bayer, Morag *et al.*, 1998). Several CBM3a X-ray structures are known and their functions appeared to be clear (Tormo *et al.*, 1996; Shimon *et al.*, 2000). In biochemical studies, CBM3a and CBM3b have been shown to bind strongly to the surface of microcrystalline cellulose and were proposed to promote the cellulolytic reaction by concentrating the enzyme near the cellulose surface (Gilad *et al.*, 2003; Tormo *et al.*, 1996). In contrast, members of subfamily CBM3c failed to bind to crystalline cellulose substrates; instead, the CBM3c is intimately associated with certain GH9 catalytic modules and has been shown to alter GH9 function from the standard endo-acting mode to a processive endoglucanase (Sakon *et al.*, 1997; Irwin *et al.*, 1998; Zverlov *et al.*, 2003). Members of subfamilies CBM3a and CBM3b contain well conserved cellulose-binding residues that differ from those of CBM3c (Fig. 1). Recently, however, the commonly assumed hypothesis that CBM3 functions primarily as a cellulose-binding module has been questioned because of the finding that certain CBM3s, notably CBM3b′ and CBM3c′, lack some of the highly conserved residues (of the 'classic' CBM3s) in their cellulose-binding sites and also lack the binding function to microcrystalline cellulose (Jindou *et al.*, 2006).

The CBM3b from cellobiohydrolase 9A of *Clostridium thermocellum* diverges from the standard CBM3b and lacks several of the conserved residues from its putative cellulose-binding surface (Fig. 1).

Table 1
Crystal and diffraction data.

Values in parentheses are for the highest resolution cell.

No. of crystals	1
Total rotation range (°)	125
Temperature (K)	100
Unit-cell parameters	
$a = b$ (Å)	95.63
c (Å)	81.21
Space group	$P4_1$ or $P4_3$
V (Å^3)	742673.32
No. of molecules in ASU	4 or 5
Resolution range (Å)	28.6–2.68 (2.73–2.68)
Total No. of reflections	105458
Unique reflections	20871
Mosaicity	0.729
Redundancy	5.1
Completeness (%)	99.5 (96.1)
Mean $I/\sigma(I)$	29.27 (2.2)
$R_{\rm merge}$† (%)	5.9 (49.0)
Overall B factor from Wilson plot (Å^2)	77.4

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \overline{I(hkl)}|/\sum_{hkl}\sum_i I_i(hkl)$.

Moreover, it was inconsistent in its cellulose-binding ability, failing to bind crystalline cellulose in our hands (Jindou *et al.*, 2006), although others (Kataeva *et al.*, 2004) have reported that this CBM3b does bind cellulose. Thus, X-ray structure analysis was initiated to explore a structural explanation for this inconsistency. Here, we report the cloning, overexpression, purification, crystallization and preliminary X-ray analysis of CBM3b from cellobiohydrolase 9A of *C. thermocellum*.

2. Cloning, expression and purification

A DNA fragment encoding CBM3b from cellobiohydrolase 9A (GenBank accession code X80993) was amplified by PCR from *C. thermocellum* genomic DNA using two specific primers: 5′-CCATGGGCGATGTAAAAGTACAGTATTTGTGC-3′ and 5′-CTCGAGCGGCGGCGTTCCCCAAAC-3′. The PCR products were purified, cleaved with the restriction enzymes *Nco*I and *Xho*I and inserted into the pET-28a(+) expression vector (Novagen, Madison, Wisconsin, USA) together with a C-terminal hexahistidine tag, yielding pET-CBM3b.

Escherichia coli strain BL21(DE3) harbouring pET-CBM3b was aerated at 310 K in 3 l Terrific broth supplemented with 25 µg ml^{-1} kanamycin. After 3 h, the culture reached an A_{600} of 0.6; 0.1 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce gene expression and cultivation was continued at 310 K for an additional 12 h. Cells were harvested by centrifugation (5000*g* for 10 min) at 277 K and were subsequently resuspended in 50 m*M* NaH_2PO_4 pH 8.0 containing 300 m*M* NaCl at a ratio of 1 g wet pellet to 4 ml buffer solution. A few micrograms of DNase powder were added prior to the sonication procedure. The suspension was kept on ice during sonication, after which it was centrifuged (20 000*g* at 277 K for 20 min) and the supernatant collected. The expressed His-tagged protein was isolated by metal-chelate affinity chromatography using Ni-IDA resin (Rimon Biotech, Israel). No attempts to remove the His tag were made. Fast protein liquid chromatography (FPLC) was performed using a Superdex 75pg column and ÄKTA Prime system (GE Healthcare). The protein was concentrated to 25 mg ml^{-1} using Centriprep YM-3 centrifugal filter devices (Amicon Bioseparation, Millipore). The protein concentration was determined by measuring the absorbance at 280 nm.

3. Crystallization

Initially, the protein sample, which contained 20 mg ml^{-1} protein solution in 1.2 m*M* Tris–HCl pH 7.5, 1.5 m*M* NaCl, 0.025% sodium azide, was screened for crystallization conditions using the micro-batch method (Chayen *et al.*, 1990) under a 1:1 mixture of silicon and paraffin oils using 288 pre-formulated crystallization solutions from

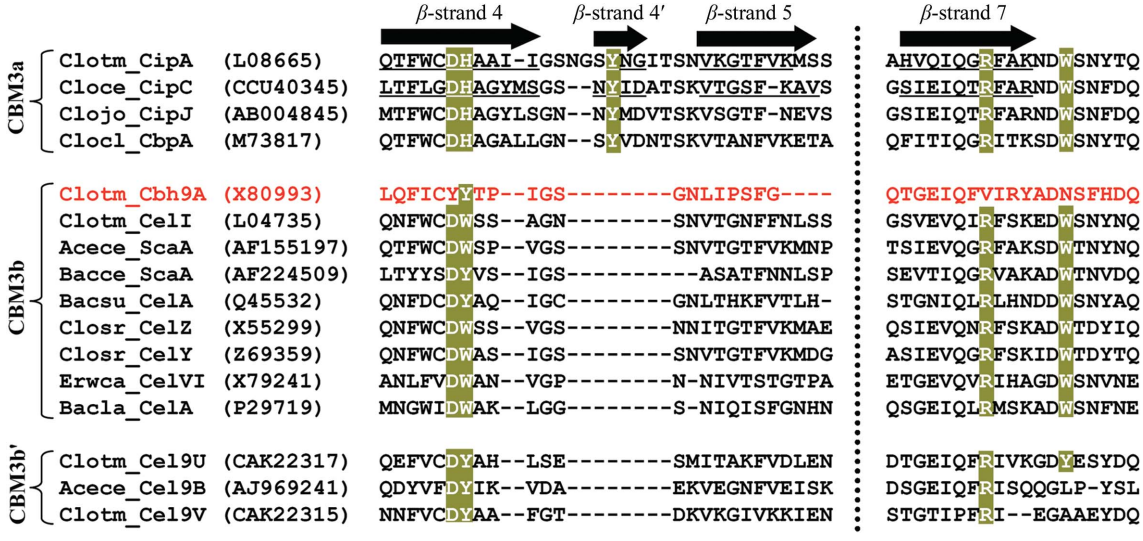


Figure 1
Multiple sequence alignment (the relevant part is shown) of selected CBM3a, CBM3b and CBM3b′ modules. A planar strip (five and four residues for CBM3a and CBM3b, respectively) considered to participate in binding to the crystalline cellulose surface is highlighted in olive. CBM3b from cellobiohydrolase 9A of *C. thermocellum* (the subject of the present study) is coloured red. The appropriate predicted β-strands, confirmed by X-ray analyses of CBM3a from *C. thermocellum* CipA (PDB code 1nbc) and CBM3a from *C. cellulolyticum* CipC (PDB code 1g43), are indicated by arrows. Sequence alignments were generated using *ClustalW* (http://www2.ebi.ac.uk/clustalw/) based on the representative GenBank sequences (accession codes in parentheses). CBM3b lacks 6–8 amino acids between β-strands 4 and 5, eliminating a small $\beta4'$ strand that contains one of the putative cellulose-binding residues that characterizes CBM3a. Clotm, *C. thermocellum*; Cloce, *C. cellulolyticum*; Clojo, *C. josui*; Clocl, *C. cellulovorans*; Acece, *Acetivibrio cellulolyticus*; Bacce, *Bacteroides cellulosolvans*; Bacau, *Bacillus subtilis*; Closr, *C. stercorarium*; Erwca, *Pectobacterium carotovorum* (*Erwinia carotovora*); Bacla, *Bacillus lautus*.

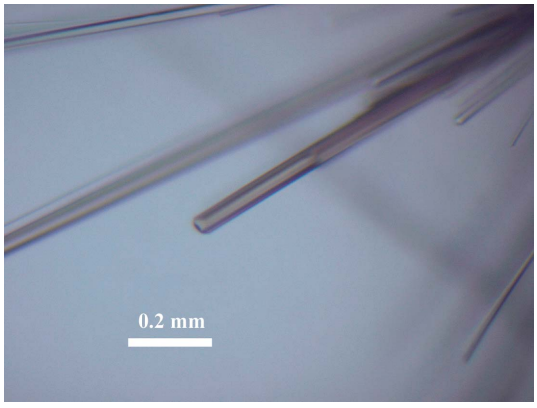


Figure 2
Crystals of CBM3b from cellobiohydrolase 9A of *C. thermocellum* grown by the microbatch-under-oil method in 16.25% PEG 8000, 0.05 *M* sodium acetate pH 4.5, 0.12 *M* Li_2SO_4.

Hampton Research HT screens (SaltRx, Index HT and Crystal Screen HT; http://www.hamptonresearch.com/) and 96 conditions from the Wizard I and II sparse-matrix crystallization screens from Emerald BioSystems (http://www.emeraldbiosystems.com/). Samples were dispensed using an Oryx 6 crystallization robot from Douglas Instruments (http://www.douglas.co.uk/). A 1 µl aliquot of the protein solution and 1 µl of the selected crystallization condition were dispensed into each well. Crystallization was performed at 293 K in a temperature-controlled room. After 2–3 d, small crystals appeared in several conditions, with the most promising being Wizard I condition No. 17 [30%(*v*/*v*) PEG 8000, 0.1 *M* sodium acetate pH 4.5, 0.2 *M* Li_2SO_4]. These initial crystallization conditions were optimized using the microbatch and hanging-drop methods. The best needle-like crystals (Fig. 2) were obtained using the microbatch method by setting up 2 µl crystallization drops containing 16.25% PEG 8000, 0.05 *M* sodium acetate pH 4.5, 0.12 *M* Li_2SO_4 and 9.47 mg ml^{-1} protein.

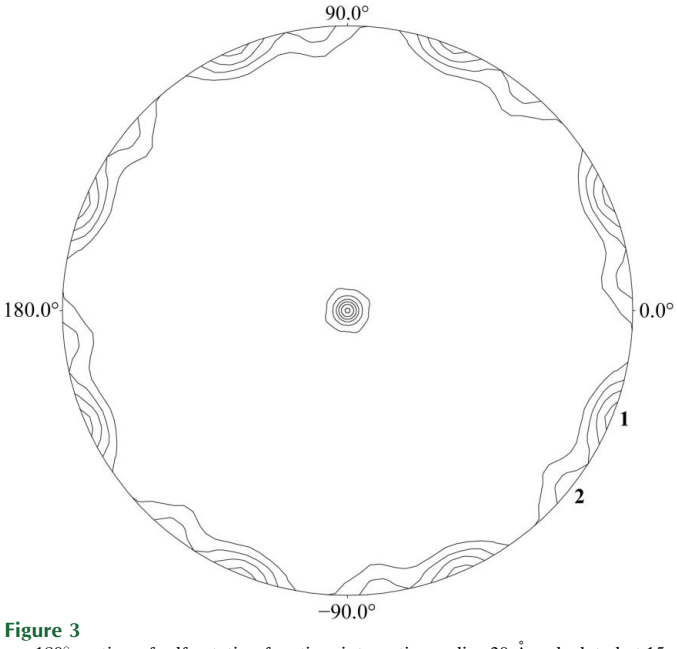


Figure 3
$\kappa = 180^\circ$ section of self-rotation function, integration radius 30 Å, calculated at 15–3 Å resolution. The contour levels start at 1 r.m.s., with increments of 0.5 r.m.s. Two independent noncrystallographic twofold axes are observed and are numbered 1 and 2.

4. X-ray diffraction

Crystals were harvested from the crystallization drop using a nylon loop (Hampton Research) and transferred for less than a minute into a solution mimicking the mother liquor and supplemented with 25% ethylene glycol. For data collection, crystals were mounted on the MiTeGen stiff micro-mount (http://www.mitegen.com/) made of polyimide and flash-cooled in a nitrogen stream at a temperature of 100 K produced by an Oxford Cryostream low-temperature generator (Cosier & Glazer, 1986).

Diffraction data from the CBM3b crystals were measured using the ID14-4 beamline at ESRF, Grenoble, France. An ADSC Q315 detector and X-ray radiation of wavelength 0.976 Å were used and a total of 250 0.5° oscillation frames were collected. Data were processed with *DENZO* and scaled with *SCALEPACK* as implemented in *HKL*-2000 (Otwinowski & Minor, 1997). The crystals diffracted to 2.6 Å and belonged to the primitive tetragonal space group $P4$, with unit-cell parameters $a = b = 95.63$, $c = 81.21$ Å, $\alpha = \beta = \gamma = 90^\circ$. Systematic absences suggested two possible space groups, $P4_1$ or $P4_3$, with either four or five monomers in the asymmetric unit. With Matthews densities V_M (Matthews, 1968) of 2.86 and 2.29 Å^3 Da^{-1}, these correspond to solvent contents of 56.98 and 46.23%, respectively. The X-ray data-analysis statistics are presented in Table 1.

5. Comments

The self-rotation function of the CBM3b with two noncrystallographic twofold axes is shown in Fig. 3. A pseudotranslation was not detected on the Patterson function.

Molecular replacement was attempted using the X-ray structures of CBM3a from the cellulosomal CipA scaffoldin subunit of the *C. thermocellum* (Tormo *et al.*, 1996; PDB code 1nbc; 32% amino-acid sequence identity) and of CipC from the *C. cellulolyticum* cellulosome (Shimon *et al.*, 2000; PDB code 1g43; 29% identity) as search models in *Phaser* (McCoy *et al.*, 2005). *Phaser* searches led to a marginal solution (Z score of 4.6 in the translation function after placing the last molecule) with four molecules in the asymmetric unit. Rebuilding and structural refinement is in progress.

We thank the ESRF, Grenoble for the use of the macromolecular crystallographic data-collection facilities and the ID14eh cluster staff for their assistance.

References

Bayer, E. A., Belaich, J. P., Shoham, Y. & Lamed, R. (2004). *Annu. Rev. Microbiol.* **58**, 521–554.
Bayer, E. A., Morag, E., Lamed, R., Yaron, S. & Shoham, Y. (1998). *Carbohydrases From Trichoderma reesei and Other Microorganisms*, edited by M. Claeyssens, W. Nerinckx & K. Piens, pp. 39–65. London: Royal Society of Chemistry.
Bayer, E. A., Shimon, L. J., Shoham, Y. & Lamed, R. (1998). *J. Struct. Biol.* **124**, 221–234.
Chayen, N. E., Shaw Stewart, P. D., Maeder, D. L. & Blow, D. M. (1990). *J. Appl. Cryst.* **23**, 297–302.
Cosier, J. & Glazer, A. M. (1986). *J. Appl. Cryst.* **19**, 105–107.
Gilad, R., Rabinovich, L., Yaron, S., Bayer, E. A., Lamed, R., Gilbert, H. J. & Shoham, Y. (2003). *J. Bacteriol.* **185**, 391–398.
Irwin, D., Shin, D. H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A. & Wilson, D. B. (1998). *J. Bacteriol.* **180**, 1709–1714.

Jindou, S., Xu, Q., Kenig, R., Shulman, M., Shoham, Y., Bayer, E. A. & Lamed, R. (2006). *FEMS Microbiol. Lett.* **254**, 308–316.
Kataeva, I. A., Uversky, V. N., Brewer, J. M., Schubot, F., Rose, J. P., Wang, B.-C. & Ljungdahl, L. G. (2004). *Protein Eng. Des.* **17**, 759–769.
McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Sakon, J., Irwin, D., Wilson, D. B. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 810–818.
Shimon, L. J. W., Pagès, S., Belaich, A., Belaich, J.-P., Bayer, E. A., Lamed, R., Shoham, Y. & Frolow, F. (2000). *Acta Cryst.* D**56**, 1560–1568.
Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y. & Steitz, T. A. (1996). *EMBO J.* **15**, 5739–5751.
Zverlov, V. V., Velikodvorskaya, G. A. & Schwarz, W. H. (2003). *Microbiology*, **149**, 515–524.