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

Preliminary X-ray analysis of cellobiohydrolase Cel7B from *Melanocarpus albomyces*

Tarja Parkkinen,[a]* Anu Koivula,[b] Jari Vehmaanperä[c] and Juha Rouvinen[a]

[a]Department of Chemistry, University of Joensuu, PO Box 111, 80101 Joensuu, Finland, [b]VTT Technical Research Centre of Finland, PO Box 1000, 02044 VTT, Finland, and [c]ROAL Oy, PO Box 57, 05201 Rajamäki, Finland

Correspondence e-mail: tarja.parkkinen@joensuu.fi

Received 18 April 2007
Accepted 14 July 2007

Cellobiohydrolases are enzymes that cleave off cellobiose units from cellulose chains in a processive manner. *Melanocarpus albomyces* Cel7B is a thermostable single-module cellobiohydrolase that has relatively low activity on small soluble substrates at room temperature. It belongs to glycoside hydrolase family 7, which includes endo-β-1,4-glucanases and cellobiohydrolases. Cel7B was crystallized using the hanging-drop vapour-diffusion method and streak-seeding. The crystals belonged to space group $P2_1$, with unit-cell parameters a = 50.9, b = 94.5, c = 189.8 Å, β = 90.0° and four monomers in the asymmetric unit. Analysis of the intensity statistics showed that the crystals were pseudo-merohedrally twinned, with a twinning fraction of 0.37. X-ray diffraction data were collected at 1.6 Å resolution using synchrotron radiation.

1. Introduction

Cellulose is found in plant cell walls as a structural material and is the most abundant polysaccharide on Earth. Cellulose is composed of β-D-glucopyranose units linked together by 1,4-glycosidic bonds. The linear cellulose chains form a crystalline polymer through hydrogen bonding and van der Waals forces. The hydrolysis of cellulose requires the synergistic action of three types of cellulolytic enzymes: endo-β-1,4-glucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) and β-glucosidases (EC 3.2.1.21) (Bayer *et al.*, 1998; Lynd *et al.*, 2002). Endoglucanases cut the cellulose chain at random positions, producing oligosaccharides of various lengths and consequently new chain ends. Cellobiohydrolases are exo-acting enzymes that produce cellobiose from the ends of cellulose chains in a sequential manner. β-Glucosidases are needed to hydrolyze cellobiose to glucose. Many cellulases have a modular structure composed of a catalytic module linked to a carbohydrate-binding module (CBM), which helps the enzymatic action on the solid substrates (Bayer *et al.*, 1998). Both the catalytic modules and the CBMs can be divided into different families on the basis of their sequences and structures (http://www.cazy.org/; Henrissat & Davies, 1997). The glycoside hydrolase (GH) family of the catalytic module is also related to the stereochemistry of the O-glycosidic bond cleavage, in which C1 stereochemistry is retained or inverted by an acid hydrolysis mechanism (Yip & Withers, 2004).

GH family 7 consists of endoglucanases and cellobiohydrolases. The GH7 cellobiohydrolases are of fungal origin and seem to be particularly active in crystalline cellulose degradation. It has been suggested hydrolysis proceeds from the reducing end of a cellulose chain and follows a retaining mechanism (Divne *et al.*, 1998; Ståhlberg *et al.*, 1996). Three-dimensional GH7 cellobiohydrolase structures have been reported from three different organisms: *Trichoderma reesei* (Divne *et al.*, 1994), *Phanerochaete chrysosporium* (Muñoz *et al.*, 2001) and *Talaromyces emersonii* (Grassick *et al.*, 2004). The overall structure is a β-sandwich formed by two antiparallel sheets packed face to face. A long cellulose-binding tunnel is formed by loops that extend from the sandwich.

Three neutral cellulases from the thermophilic ascomycete fungus *Melanocarpus albomyces* have been purified and characterized; they have also been cloned and expressed at high levels in the industrial production host *Trichoderma reesei* (Haakana *et al.*, 2004; Miettinen-Oinonen *et al.*, 2004). Two of the cellulases, Cel7A and Cel45A, were

© 2007 International Union of Crystallography
All rights reserved

endoglucanases belonging to GH families 7 and 45, respectively. The third cellulase, Cel7B, was a family 7 cellobiohydrolase. Cel7B is a single-module enzyme composed of 430 amino acids and has a theoretical molecular weight of 47.5 kDa. Its sequence identity with the *T. reesei* GH7 cellobiohydrolase catalytic module is 50%. *M. albomyces* Cel7B demonstrates an ~5 K improved unfolding temperature and a threefold lower catalytic rate for soluble substrates at room temperature compared with *T. reesei* Cel7A. In addition, cellobiose is a strong competitive inhibitor of *M. albomyces* Cel7B (Voutilainen *et al.*, 2007, and manuscript in preparation). *M. albomyces* Cel7B has also been expressed in *Saccharomyces cerevisiae* and the thermostability of the enzyme has been further improved by random mutagenesis (Voutilainen *et al.*, 2007). In this article, we present the crystallization and X-ray data analysis of the *M. albomyces* Cel7B.

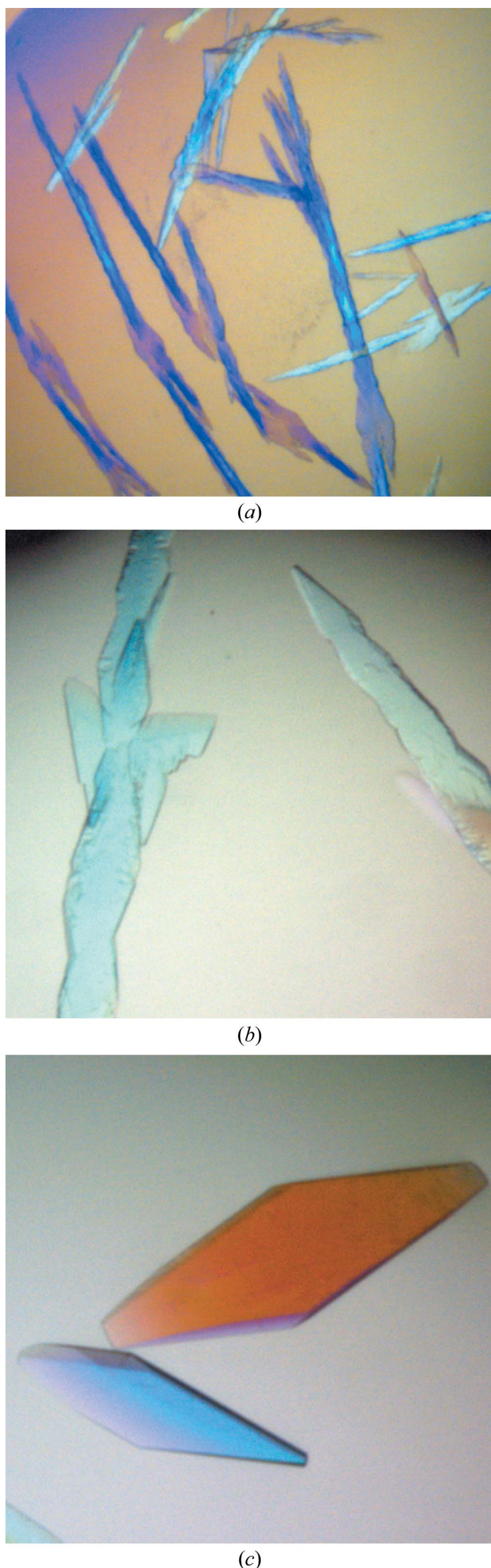

Figure 1
Cel7B crystals. (*a*) Crystals from initial screens. (*b*) Crystals from optimized conditions. (*c*) Streak-seeded crystals.

2. Materials and methods

2.1. Protein purification and dynamic light scattering

The Cel7B cellobiohydrolase was purified from *M. albomyces* as described previously (Miettinen-Oinonen *et al.*, 1997). The Cel7B protein sample was in 25 m*M* PIPES buffer pH 6.0 containing 0.25 m*M* EDTA and 5% ammonium sulfate. The sample had an absorption of 3.1 at 280 nm, corresponding to a Cel7B concentration of 1.7 mg ml^{-1} (ε = 84 860 cm^{-1} M^{-1}). Prior to the dynamic light-scattering (DLS) experiment, the protein sample was filtered through 0.1 µm Anodisc 13 membrane filters (Whatman) to eliminate any gas bubbles or large aggregates. DLS studies on Cel7B were performed using a DynaPro99 DLS Instrument (Protein Solutions Inc.). *DYNAMICS* v.6.3.18 software was used in data collection and analysis. The number of acquisition scans was 20 and the acquisition time was 10 s per scan at 298 K. DLS experiments of Cel7B showed a radius of 3.2 nm and a polydispersity of 40%. The relatively high polydispersity indicated that the Cel7B sample was less likely to crystallize than well behaved samples. However, the size distribution was still monomodal and crystallization trials were worth carrying out. A monomodal fit of the DLS data predicted a molecular weight of 50 kDa, indicating that Cel7B is a monomer in solution.

2.2. Crystallization

Initial crystallization screening was performed using the hanging-drop vapour-diffusion method against Crystal Screen I (Hampton Research) at room temperature, using the same protein solution as in the DLS experiments. The drop size was 2 + 2 µl and the reservoir volume was 500 µl in all experiments. From the conditions screened, two hits were found from solution Nos. 18 (20% PEG 8000, 0.1 *M* sodium cacodylate pH 6.5, 0.2 *M* magnesium acetate) and 46 (18% PEG 8000, 0.1 *M* sodium cacodylate pH 6.5, 0.2 *M* calcium acetate). The initial crystals (Fig. 1*a*) appeared after 1–2 d. New screenings, which varied the pH of the buffer, the concentration of the precipitant and the concentration and composition of the salt, were performed but did not yield crystals suitable for X-ray analysis. The crystals (Fig. 1*b*) were still intergrown, forming large plates. After streak-seeding experiments, we were able to grow crystals (Fig. 1*c*) that could be used for data collection. The optimized crystallization conditions contained 12% PEG 8000, 0.1 *M* sodium cacodylate pH 6.5 and 0.1 *M* calcium chloride. In streak-seeding experiments, the existing crystal was touched with a dog hair, which was subsequently used to streak a line across a new drop. The crystallization drops were allowed to equilibrate for 6 h before seeding. Crystals grew to dimensions of 0.4 × 0.3 × 0.1 mm in 4–6 d.

2.3. Data collection

Prior to data collection, the crystals were briefly soaked in a cryoprotectant solution consisting of the crystallization solution with an additional 20%(*v*/*v*) glycerol. Crystals were transferred from the crystallization drop to the cryoprotectant and finally mounted in cryoloops for X-ray data collection. After mounting, the crystals were cooled by plunging them into liquid nitrogen and then stored prior to transport to the synchrotron. Diffraction data were collected using synchrotron radiation on beamline X12 at EMBL Hamburg, Germany; the wavelength was 0.9 Å. Diffraction images were collected at 100 K with 0.5° oscillations using a MAR Mosaic CCD detector. Diffraction intensities were integrated using *XDS* and scaled and merged using *XSCALE* (Kabsch, 1993). Data-collection and processing parameters are given in Table 1.

2.4. Detection of crystal twinning

Twinning was first suggested from the intensity distributions provided by *XDS* (Kabsch, 1993). Detection and analysis of twinning was then performed using the program *TRUNCATE* from *CCP*4 (Collaborative Computational Project, Number 4, 1994). The twinning fraction was estimated from the cumulative distribution of *H* and the Britton plot calculated using the program *DETWIN* from *CCP*4.

3. Results and discussion

The *M. albomyces* Cel7B crystals diffracted to 1.6 Å resolution. The crystals belonged to space group $P2_1$, with unit-cell parameters $a = 50.9$, $b = 94.5$, $c = 189.8$ Å, $\beta = 90.0°$. Since the $P2_1$ cell has a β angle close to 90°, we considered the possibility of pseudo-merohedral crystal twinning. Analysis of the intensity statistics calculated by *XDS* showed the first indication of twinning. The second intensity moment was 0.8 for Cel7B crystals. The value is low compared with the value of 1.0 for untwinned data. Further tests for twinning were performed using intensity statistics and distributions (Dauter, 2003; Rees, 1980; Stanley, 1972). The cumulative intensity distribution $N(Z)$, which shows a clearly sigmoidal shape compared with the expected distribution, is shown in Fig. 2(*a*). An analysis of the second intensity moment of the acentric data ($\langle I^2\rangle/\langle I\rangle^2$; Fig. 2*b*) gave a value of 1.6 for Cel7B data; the expected values are 2.0 and 1.5 for untwinned and perfectly twinned data, respectively. These tests clearly indicated the presence of twinning. In order to estimate the twinning fraction α, the cumulative intensity distribution of H (Yeates, 1988, 1997) and Britton plot (Britton, 1972; Fisher & Sweet, 1980) were calculated using the twinning operator $h, -k, -l$. In the H test, the distribution of experimental acentric reflections is plotted against theoretical expectations (Fig. 2*c*). This graph and the Britton plot (Fig. 2*d*) indicate a twinning fraction of 0.37.

Estimation of the content of the asymmetric unit (Matthews, 1968) suggested that four Cel7B monomers was most likely, giving a solvent content of 46% and a Matthews coefficient of 2.3 $Å^3\ Da^{-1}$. Based on the sequence homologies of available structures of GH family 7 cellulases, the structure of cellobiohydrolase I from *T. reesei* (PDB code 1cel) was chosen as an initial model for structure solution.

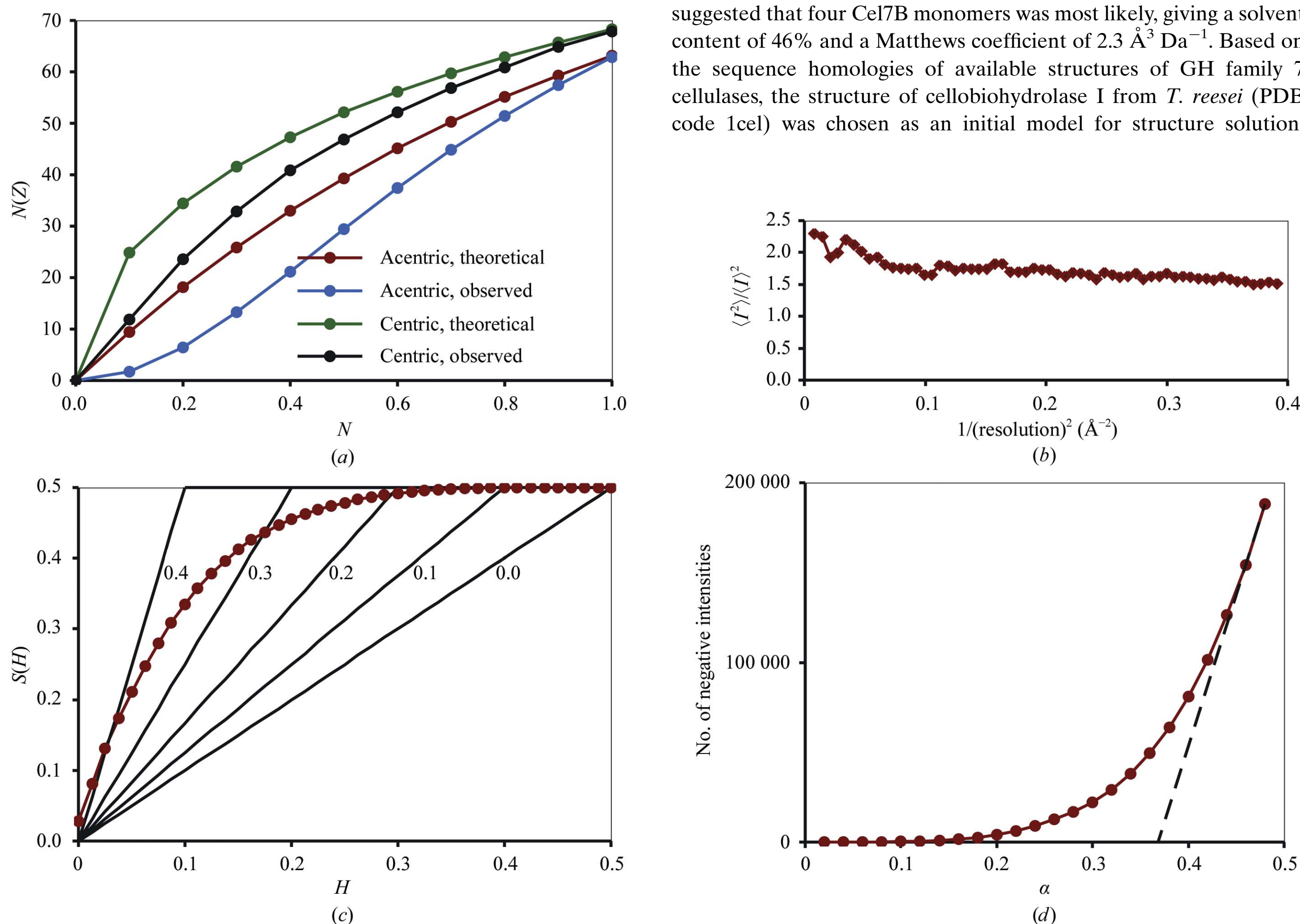

Figure 2
Detection and analysis of twinning. (*a*) The cumulative intensity distribution $N(Z)$ for acentric and centric data. The sigmoidal shape of the distribution indicates twinning. (*b*) Plot of the second moment of I ($\langle I^2\rangle/\langle I\rangle^2$). The expected value for the untwinned data is 2.0 and that for the perfect twin is 1.5. (*c*) The cumulative distribution of H calculated using twinning operator $h, -k, -l$. The red line is the experimental slope for Cel7B data. The black lines illustrate the theoretically expected slopes for various twinning fractions. (*d*) Britton plot. The line is a linear fit of the data with $\alpha \geq 0.46$ and intersects the x axis at $\alpha = 0.37$.

Table 1
Data-collection statistics for *M. albomyces* Cel7B.

Values in parentheses are for the highest resolution shell.

Resolution (Å)	25–1.6 (1.7–1.6)
No. of observations	882338 (146682)
No. of unique reflections	235645 (39005)
Redundancy	3.7 (3.8)
Completeness (%)	99.6 (99.4)
$I/\sigma(I)$	13.6 (4.8)
R_{sym} (%)	7.2 (30.2)

Molecular replacement was performed with the program *MOLREP* (Vagin & Teplyakov, 1997) from *CCP*4. The rotation and translation functions yielded four monomers per asymmetric unit. Model building and refinement are in progress using the programs *O* (Jones *et al.*, 1991) and *SHELX* (Sheldrick & Schneider, 1997), respectively.

We wish to thank Taina Simoinen (VTT) and Reetta Kallio-Ratilainen (UJ) for their excellent technical assistance. We gratefully acknowledge access to the EMBL beamline X12 at the DORIS storage ring, Hamburg, Germany. Support from the European Community Research Infrastructure Action under the FP6 'Structuring the European Research Area' programme, contract No. RII3-CT-2004-5060008, is acknowledged.

References

Bayer, E. A., Chanzy, H., Lamed, R. & Shoham, Y. (1998). *Curr. Opin. Struct. Biol.* **8**, 548–557.
Britton, D. (1972). *Acta Cryst.* A**28**, 296–297.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dauter, Z. (2003). *Acta Cryst.* D**59**, 2004–2016.
Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K. C., Teeri, T. T. & Jones, T. A. (1994). *Science*, **265**, 524–528.
Divne, C., Ståhlberg, J., Teeri, T. T. & Jones, T. A. (1998). *J. Mol. Biol.* **275**, 309–325.
Fisher, R. G. & Sweet, R. M. (1980). *Acta Cryst.* A**36**, 755–760.
Grassick, A., Murray, P. G., Thompson, R., Collins, C. M., Byrnes, L., Birrane, G., Higgins, T. M. & Tuohy, M. G. (2004). *Eur. J. Biochem.* **271**, 4495–4506.
Haakana, H., Miettinen-Oinonen, A., Joutsjoki, V., Mäntylä, A., Suominen, P. & Vehmaanperä, J. (2004). *Enzyme Microb. Technol.* **34**, 159–167.
Henrissat, B. & Davies, G. (1997). *Curr. Opin. Struct. Biol.* **7**, 637–644.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
Lynd, L. R., Weimer, P. J., van Zyl, W. H. & Pretorius, I. S. (2002). *Microbiol. Mol. Biol. Rev.* **66**, 506–577.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Miettinen-Oinonen, A., Londesborough, J., Joutsjoki, V., Lantto, R. & Vehmaanperä, J. (2004). *Enzyme Microb. Technol.* **34**, 332–341.
Miettinen-Oinonen, A., Londesborough, J., Vehmaanperä, J., Haakana, H., Mäntylä, A., Lantto, R., Elovainio, M., Joutsjoki, V., Paloheimo, M. & Suominen, P. (1997). Patent application WO9714804.
Muñoz, I. G., Ubhayasekera, W., Henriksson, H., Szabó, I., Pettersson, G., Johansson, G., Mowbray, S. L. & Ståhlberg, J. (2001). *J. Mol. Biol.* **314**, 1097–1111.
Rees, D. C. (1980). *Acta Cryst.* A**36**, 578–581.
Sheldrick, G. & Schneider, T. (1997). *Methods Enzymol.* **277**, 319–343.
Ståhlberg, J., Divne, C., Koivula, A., Piens, K., Claeyssens, M., Teeri, T. T. & Jones, T. A. (1996). *J. Mol. Biol.* **264**, 337–349.
Stanley, E. (1972). *J. Appl. Cryst.* **5**, 191–194.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Voutilainen, S., Boer, H., Linder, M. B., Puranen, T., Rouvinen, J., Vehmaanperä, J. & Koivula, A. (2007). *Enzyme Microb. Technol.* **41**, 234–243.
Yeates, T. O. (1988). *Acta Cryst.* A**44**, 142–144.
Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.
Yip, V. L. & Withers, S. G. (2004). *Org. Biomol. Chem.* **2**, 2707–2713.