Two new major subunits in the cellulosome of *Clostridium thermocellum*: xyloglucanase Xgh74A and endoxylanase Xyn10D

Vladimir V. Zverlov,^{1,2} Nicolaus Schantz,¹ Philippe Schmitt-Kopplin³ and Wolfgang H. Schwarz¹

Correspondence

Wolfgang H. Schwarz

schwarz@mikro.biologie.tu-muenchen.de

¹Research Group Microbial Biotechnology, Technische Universität München, Am Hochanger 4, D-85350 Freising-Weihenstephan, Germany

²Institute of Molecular Genetics, Russian Academy of Science, Kurchatov Sq., 123182 Moscow, Russia

³GSF - National Research Center for Environment and Health, Institute for Ecological Chemistry, Ingolstädter Landstrasse 1, D-85764 Neuherberg, Germany

The structure and enzymic activity of xyloglucanase Xgh74A and endoxylanase Xyn10D, components in the cellulosomes of cellulose-grown *Clostridium thermocellum*, were determined. Xyn10D is a thermostable endo-1,4- β -xylanase with a module composition identical to Xyn10C (CBM22-GH10-Doc). It hydrolyses xylan and mixed-linkage 1,3-1,4- β -glucan with a temperature optimum of 80 °C. Xyloglucanase Xgh74A contains a catalytic module of GHF74 in addition to a C-terminal dockerin module. It hydrolyses every fourth β -1,4-glucan bond in the xyloglucan backbone, thus producing decorated cellotetraose units. Its low activity on CMC and lack of activity on amorphous cellulose indicates recognition of the xylosidic side chains present in xyloglucan, which is readily hydrolysed (295 U mg⁻¹). The pattern of the hydrolysis products from tamarind xyloglucan resembles that of other GHF74 xyloglucan endoglucanases. The data indicate that Xgh74A and Xyn10D contribute to the *in vivo* degradation of the hemicelluloses xyloglucan and xylan by the cellulosome of *C. thermocellum*. Xgh74A is the first xyloglucanase identified in *C. thermocellum* and the only enzyme in the cellulosome that hydrolyses tamarind xyloglucan.

Received 20 May 2005

Revised 23 June 2005

Accepted 30 June 2005

INTRODUCTION

Lignocellulosic biomass is a recalcitrant substrate for enzymic hydrolysis. It is most effectively hydrolysed by the strictly anaerobic, chemo-organotrophic, thermophilic bacterium *Clostridium thermocellum*, which in contrast to the soluble cellulases and hemicellulases secreted by fungi and aerobic cellulolytic bacteria produces an extracellular multi-enzyme complex, called the cellulosome (Bayer *et al.*, 2000). Although this thermophilic *Clostridium* species is generally described to be specialized in the degradation of crystalline cellulose and indeed cannot readily utilize carbohydrates other than cellodextrins (Demain *et al.*, 2005), it produces a number of hemicellulolytic enzymes, especially xylanases. Of the 71 genes

potentially encoding cellulosomal hydrolytic subunits, 17 code for non-glucanolytic polysaccharide hydrolases (including xylanases), 8 for glycosidases and 4 for esterases (Zverlov *et al.*, 2005). Thus a large fraction of the hydrolytic genes is not directly related to cellulose degradation.

The function of the non-cellulolytic enzymes is presumably the unwrapping of the cellulose crystals from the covering matrix of lignin, pectin and hemicellulose. Whereas lignin apparently cannot be degraded by the anaerobic bacteria, the heterogeneous hemicelluloses like xyloglucan, differently derivatized xylans, galactomannan and related polysaccharides, as well as pectin, are hydrolysed. Xylan has a β -1,4-linked backbone of xylosyl residues which are substituted in the C-2 or C-3 position by α -arabinofuranosyl or glucuronic acid residues (Schwarz *et al.*, 2004). Xyloglucan is the major hemicellulose in dicotyledon type I cell walls and has a backbone of β -1,4-linked cellotetraose units, which are substituted in C-6 by α -D-xylosyl residues at two or three of the first three glucosyl residues (from the reducing end) (York *et al.*, 1993). Some of the xylosyl residues are further substituted by galactosyl, arabinosyl or fucosyl

Abbreviations: CBM, carbohydrate-binding module; CMC, carboxymethylcellulose; ESI/MS, electrospray ionization/mass spectrometry; GH/GHF, glycosyl hydrolase/glycosyl hydrolase family; PASC, phosphoric-acid-swollen cellulose.

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are AJ585344 (*xghA*) and AJ585345 (*xynD*).

residues (Fry *et al.*, 1993). The evident cross-linking between xyloglucan chains is a major factor in the stability of the three-dimensional matrix between adjacent cellulose microfibrils and its removal is a precondition for complete lignocellulosic biomass decomposition (Whitney *et al.*, 1995).

Non-cellulase polysaccharide hydrolases are usually active in an endo-mode, opening the molecule strands anywhere in the molecule and producing a new reducing end with each cut. This enzymic activity reduces the local viscosity of the hemicellulose matrix. If eventually two cuts are near enough to each other, a short oligosaccharide is produced which is soluble and can diffuse away, thus removing the non-cellulosic matrix from the surface of the cellulose crystals and allowing access of the cellulases. It is therefore no surprise that the cellulosome of *C. thermocellum* contains a number of non-cellulolytic enzymes, namely xylanases as well as 1,3-1,4- β -glucanases, a mannanase and a chitinase, which make up about half of the cellulosomal protein (Hayashi *et al.*, 1997, 1999; Hallstead *et al.*, 1999; Blum *et al.*, 2000; Kurokawa *et al.*, 2001; Zverlov *et al.*, 1994, 2002, 2005).

The exceptional efficiency of the extracellular hydrolytic machinery present in anaerobic bacteria is due to the combination of all of those activities together with the cellulases in the cellulosome (Shoham *et al.*, 1999). About 30 genes involved in cellulosome formation have hitherto been isolated by screening of genomic libraries (Bayer *et al.*, 2000; Schwarz, 2001). The following enzymes have recently been identified among the most abundant subunits in the cellulosome and therefore might have a major function in cellulolysis: CelA, CelG, CelK, CelN, CelR, CelS, CbhA, ChiA, XynC and XynZ (Zverlov *et al.*, 2005). In addition two new presumably hemicellulolytic subunits were identified: the xylanase XynD and the putative xyloglucanase XghA. Both were among the most prominent hemicellulolytic proteins in the cellulosome. The xylanases XynC and XynZ have been described earlier (Grépinet *et al.*, 1988; Hayashi *et al.*, 1997). A xyloglucanase has not hitherto been described as a subunit of the cellulosome.

The aim of the paper was to evaluate the gene sequences and the biochemical characteristics of the newly identified cellulosomal hemicellulolytic subunits Xyn10D and Xgh74A, which belong to the glycosyl hydrolase families GHF10 and GHF74 respectively and represent a new endoxylanase and the first xyloglucanase in the cellulosome.

METHODS

Strains and media. Strain F7 of *Clostridium thermocellum* was obtained from the All-Russian Collection of Microorganisms (VKMB 2203). It was grown at 60 °C in rubber-stoppered glass bottles with pre-reduced, anaerobic GS-2 medium containing 0.5% (w/v) microcrystalline cellulose (Avicel) (Johnson *et al.*, 1982). Recombinant *Escherichia coli* M15 (Qiagen) or F10 (Invitrogen) was shaken at 37 °C in Luria-Bertani broth (LB) containing ampicillin (100 μ g ml⁻¹) or kanamycin (25 μ g ml⁻¹), respectively.

Recombinant DNA techniques. Preparation of chromosomal and plasmid DNA, endonuclease digestion, ligation and transformation were carried out by standard procedures or according to supplier protocols. Plasmid DNA was prepared with the QIAprep Spin Miniprep Kit (Qiagen). Restriction digests of DNA were done as recommended by the manufacturer (MBI Fermentas). *E. coli* cells were transformed with plasmid DNA by electroporation (Bio-Rad Gene Pulser) as suggested by the supplier.

Vectors from Invitrogen (pCR-XL-TOPO) or Qiagen (pQE32) and the *E. coli* host strains supplied with the kits were used for cloning. PCR was carried out using the synthetic oligonucleotide primers h14f (3'-TAGGGGAGGGGTTGTTTAATG-5') and h14r (3'-TTAACGT-TTTAAAGGCAATTCATC-5') for pCR-XL-TOPO-XynD, or h28f (3'-CGATTTGCATGCAGGCTGTAACCAGCGTG-5') and h28r (3'-TCTGAAGCTTGTTCCGCGTAAACAATACCT-5') for pQE32-XghA, with chromosomal DNA from *C. thermocellum* F7 as template and the Expand High Fidelity PCR System (Roche Diagnostics).

The PCR amplicons were cloned and subsequently sequenced from supercoiled double-stranded plasmid DNA on both strands using a LICOR automated sequencer (MWG Biotech) with the Thermo Sequenase fluorescent-labelling primer sequencing kit (Amersham Pharmacia Biotech). Sequence data were analysed, edited and compared with the DNASIS/PROSIS for Windows package (Hitachi Software Engineering).

Purification of enzymes. Cellulosomes were prepared from Avicel-grown *C. thermocellum* F7 cultures by a modified affinity digestion method (Morag *et al.*, 1992) as described by Zverlov *et al.* (2005). Recombinant Xyn10D protein was purified from 400 ml *E. coli* TOP10F' (pCR-XL-TOPO-XynD) cultures. The cells were harvested after 15 h aerated growth at 37 °C, washed with 10 mM Tris/HCl, pH 7.6, and suspended in 10 ml of the wash buffer. After sonication the extract was heated for 15 min to 60 °C and cleared by centrifugation (20 000 g, 20 min, 4 °C). The cell-free extract was applied to a HiTrapQ column (1 ml, Amersham Biosciences) and eluted with a linear NaCl gradient (0–0.5 M) in wash buffer. Fractions with activity on xylan were pooled and concentrated by ultrafiltration (VIVAspin 2 ml concentrator, 30 000 MWCO PES, VIVASCIENCE).

Recombinant Xgh74A proteins were purified from 400 ml *E. coli* (pQE32-XghA) culture according to the procedures described in the Qiagen handbook with 3 ml Ni-NTA Superflow Agarose columns (Qiagen).

Enzyme assays. Enzyme aliquots in standard assays were incubated in MES buffer (50 mM) at the optimum pH and temperature. The substrate concentration was 1% for soluble and 2% (w/v) for insoluble polysaccharides. Reducing sugars released from polymeric substrates were quantitatively detected by the 3,5-dinitrosalicylic acid method (Wood & Bhat, 1988) assuming that one unit of enzyme liberates 1 μ mol glucose equivalent min⁻¹ (mg protein)⁻¹. Specific activities were determined in the linear range of the reaction. All determinations were performed at least in triplicate.

The optimum pH was determined by measuring the specific activity of the enzyme at a given pH. Buffers were MES, Tris/HCl and potassium phosphate. Xylan was used as substrate for XynD and tamarind seed xyloglucan for XghA. The optimum temperature was the temperature with the highest activity of the enzyme during 30 min incubation at optimum pH. Protein concentration was determined with Coomassie brilliant blue G-250 (Protein Assay, Bio-Rad). Substrates were pachyman, tamarind seed xyloglucan and barley β -glucan from Megazyme International, and birch wood glucuronoxylan (xylan) and low-viscosity carboxymethylcellulose sodium salt (CMC) from

Sigma-Aldrich. Phosphoric-acid-swollen cellulose (PASC) was prepared from Avicel CF1 (Sigma-Aldrich) according to Wood & Bhat (1988).

Zymograms of denatured cellosomal proteins were prepared by running denaturing SDS-PAGE, renaturing the enzymes within the gel slabs by three successive washes in 50 mM phosphate/citrate buffer (pH 6.5) with and without 2-propanol as described previously (Schwarz *et al.*, 1987) and soaking the gel slabs for 1 h with 0.05 % tamarind xyloglucan solution. After incubation at 60 °C in enzyme assay buffer, the gels were stained with Congo red and destained with 1 M NaCl to make bands of hydrolytic activity visible against a red background.

Thin-layer chromatography (TLC). Hydrolysis products from oligo- and polysaccharides were separated on 0.2 mm silica gel 60 aluminium plates (VWR) with acetonitrile/water (80:20, v/v) as eluent. Sugars were detected by spraying the plates with a freshly prepared mixture of 10 ml stock solution (1 g diphenylamine + 1 ml aniline dissolved in 100 ml acetone) with 1 ml orthophosphoric acid, followed by heating the plates at 120 °C until colour developed.

Identification of glucan hydrolysis products by ESI/MS. Infusion experiments were conducted using a homemade coaxial microelectrospray ionization interface coupled to a Finnigan LCQ Duo ion-trap mass spectrometer (ThermoQuest) as described by Frommberger *et al.* (2004). The sample flux was set to 0.35 $\mu\text{l min}^{-1}$ and the sheath liquid was methanol/water/acetic acid (90:10:1, by vol.) pumped at 5 $\mu\text{l min}^{-1}$; electrospray ionization was conducted in the positive and the negative ionization mode and the heated capillary was set to 150 °C for optimum signals (Schmitt-Kopplin & Kettrup, 2003).

Nucleotide sequence accession numbers. Nucleotide sequences are deposited under GenBank/EMBL/DDJB accession no. AJ585344 (*xghA*) and AJ585345 (*xynD*). The draft genomic sequence from *C. thermocellum* ATCC 27405 with GenBank ID AABG000000000 was used for primer design (Whole Genome Shotgun project WGS by DOE).

RESULTS

The proteins XghA and XynD were previously detected by MALDI-TOF of the prominent protein spots in 2-D gels of the *C. thermocellum* cellosomal components (Zverlov *et al.*, 2005). Their putative genes, called *xghA* and *xynD* (ORF Chte02002261 and Chte0200364, respectively), were cloned by targeted DNA amplification from *C. thermocellum* DNA using oligonucleotide primers specifically designed on the basis of the open reading frames identified in the genomic sequence. The amplified genes were sequenced to confirm the integrity of the cloned fragments.

Analysis of the sequences revealed complex module structures similar to those found in other cellosomal genes. In both genes the dockerin type I module for binding to the cohesin modules of the cellosomal scaffoldin is located C-terminally. The modules are separated by short linker sequences rich in hydroxy amino acids (PTS boxes). Xyn10D contained, in addition to a central catalytic GHF10 module, an N-terminal substrate-binding module of family CBM22 known to bind xylan (Charnock *et al.*, 2000).

Xgh74A consists of an N-terminal catalytic module of family GHF74 and the dockerin module.

To characterize the proteins biochemically, the genes were expressed in *E. coli*. Both proteins had the expected mass in denaturing SDS-PAGE (not shown) as was calculated from the sequence, and were enzymically active (Table 1). This corroborated the correct cloning procedure and the predicted extent of the reading frames.

The recombinant Xyn10D protein was purified by ion-exchange chromatography with an apparent molecular mass of 80 kDa. The temperature with the highest activity was 80 °C using the substrate oat-spelt xylan (Table 1). The cumulative activity after 30 min at 90 °C was 15 % of that at 80 °C; this indicates rather slow inactivation at 90 °C and high thermostability. As a typical GHF10 xylanase it had activity on mixed-linkage β -glucan (barley 1,3-1,4- β -D-glucan). TLC analysis showed the production of larger xylo-oligosaccharides from xylan at the onset of the reaction. With longer incubation times mainly xylobiose and minor amounts of xylose and xylotriose appeared (Fig. 1a, C–F). Xylotriose and xylotetraose were readily degraded to roughly equimolar amounts of xylose and xylobiose in both cases, indicating statistical attack of all bonds, at least with xylotetraose (Fig. 1a, A, B). Xylobiose did not seem to be processed further. This reflects the typical hydrolysis pattern of an endo- β -1,4-xylanase.

Recombinant Xgh74A protein was purified by His-tag affinity chromatography. It had a pH optimum at pH 6.4 on tamarind seed xyloglucan, with 50 % activity between pH 5.7 and 7.8, and a temperature optimum at 75 °C (Table 1). Cumulative activity after 30 min incubation at 90 °C was 25 % of the optimum value at 75 °C, indicating great heat stability at least up to 75 °C. Despite a high specific activity on barley β -glucan as determined by release of reducing sugars (Table 1), the hydrolysis products of the

Table 1. Hydrolytic activity of the new components on different polysaccharides

Specific activity of purified recombinant enzymes was determined at optimal pH and temperature.

Substrate	Specific activity (U mg ⁻¹)	
	Xyn10D	Xgh74A
Glucuronoxylan	28.1 ± 1.3	1.2 ± 0.1
Tamarind xyloglucan	0.1 ± 0.01	295 ± 4
Pachyman	0	0
Barley β -glucan	11.1 ± 0.6	238 ± 6
CMC	0	3.0 ± 0.2
pH _{opt}	pH 6.4	pH 6.4
T _{opt}	80 °C	75 °C
Amino acid residues	639 aa	743 aa
Molecular mass	71.7 kDa	81.8 kDa

Fig. 1. TLC of the reaction products from enzymic activity. Substrates were hydrolysed in incubation buffer and subjected to TLC to make the reaction products visible. Markers are (from top down): M^x, xylose, xylobiose, xylotriose and xylotetraose; M^c, glucose, cellobiose, cellotriose and cellotetraose. (a) Xyn10D was incubated with xylotriose (lane A), xylotetraose (lane B) and birch wood xylan for 0, 2, 4 and 18 h (lanes C–F). (b) Xgh74A was incubated for 0, 2, 4, 8 and 18 h with tamarind xyloglucan (lanes A–E).

mixed-linkage glucan did not separate well in TLC: relatively large oligosaccharides had been produced which were not further degraded (data not shown). Marginal activity was observed on CMC and on glucuronoxylan, and no activity on the 1,3- β -glucan pachyman. This substrate pattern suggests a specificity for the hydrolysis of β -1,4-glucosidic bonds. However, non-derivatized, pure 1,4- β -glucans like amorphous cellulose (PASC) were not hydrolysed (not shown).

The enzyme depolymerized xyloglucan completely and released reducing sugars from tamarind xyloglucan quickly. Large oligosaccharides with more than six sugar residues were produced which could not be separated sufficiently by TLC (Fig. 1b, A–E). No smaller molecules were visible as hydrolysis products even after extended incubation with excess enzyme, indicating an extremely low or no activity on the oligosaccharides (Fig. 1b, E).

The oligosaccharides produced from the xyloglucan were further investigated by infusion ESI/MS. The MS data fit well with the occurrence of the hepta- to nonaglycosides XXXG, XLXG (or XXLG) and XLLG as major products according to the nomenclature of Fry *et al.* (1993) with m/z values of 1080, 1242 and 1404 respectively (Fig. 2). Minor products can be identified as XXG, XXX and XXGG. Each oligosaccharide appears as a cluster of peaks from protonated molecules as well as different salt adducts. Acetyl esters from XXXG as well as from other oligosaccharides ($m/z + 42$) were found in traces. The majority of the products were substituted cellotetraose units of xyloglucan.

Cellulosomes were purified from a cellulose-grown *C. thermocellum* culture and incubated with tamarind xyloglucan. The specific activity of the cellulosomal preparation on tamarind xyloglucan was 4.7 U mg^{-1} (± 0.4). The

Fig. 2. Product analysis from tamarind xyloglucan hydrolysis by Xgh74A with infusion ESI/MS. The m/z scale and the numbers on the major peaks indicate the molecular mass. The probable composition of the xyloglucan oligosaccharides is given above the respective peaks (nomenclature according to Fry *et al.*, 1993).

degradation pattern was indistinguishable in TLC from that obtained with purified recombinant Xgh74A (data not shown). This indicates (1) that xyloglucanase is active in native cellulosomes, and (2) that the decorated cellotetraose units produced are not further degraded by either Xgh74A, or any β -glucosidase or other glycosidase eventually present in the cellulosome.

DISCUSSION

The *C. thermocellum* cellulosome is the most effective cellulose-degrading enzyme complex known so far (Demain *et al.*, 2005). But even if cellulases are maximally expressed by growing the bacterium on cellulose, about half of the proteins in the cellulosomes are hemicellulolytic enzymes (Zverlov *et al.*, 2005). This emphasizes the importance of the hemicellulases in the degradation of natural cellulosic biomass. Two hitherto unknown major hemicellulolytic subunits have been identified by analysing denatured cellulosome components: Xgh74A and Xyn10D, encoded by the genes *xghA* and *xynD*, respectively. They were cloned and characterized as a xyloglucanase and an endo- β -xylanase respectively.

The cellulosomal subunit Xyn10D cleaves xylan in endo-mode and with a hydrolysis pattern known from other endo-xylanases of GHF10. It is at present not clear why the bacterium produces at least three different GHF10 endo-xylanases for xylan hydrolysis: Xyn10C, Xyn10D and Xyn10Z. Xyn10D is the third most prevalent xylanase in cellulose-grown cellulosomes, with a molar abundance of 0.3 (about one Xyn10D in each third cellulosomal particle) (Zverlov *et al.*, 2005). The similarities in product formation between Xyn10D and Xyn10C (Hayashi *et al.*, 1997) were conspicuous. But nevertheless and despite an identical

module architecture the sequence homology between Xyn10C and Xyn10D is low (21 % identity in the catalytic module), and the presence of these two similar genes in the genome does not seem to be the result of a gene duplication. However, another non-cellulosomal xylanase gene with much higher homology (56 % aa identity) was identified in the genome of *C. thermocellum*: ORF Chte02002923 (zp_00311867), which has an additional N-terminal module of unknown function and lacks a binding and a dockerin module. This ORF has not yet been investigated. It is not known if it is expressed or functional. The characterized xylanase with the highest homology to Xyn10D is XynA of *Thermoanaerobacterium thermosulfurigenes*, with 34 % identity (Matuschek *et al.*, 1996).

An explanation for the presence of multiple xylanases could be the resulting improvement in substrate decomposition, e.g. (1) in the competition for binding to different cohesin sites on CipA (and thus occupying more of the limited cohesin positions with xylanases), (2) in differences in substrate binding or (3) in the cleavage of different sites in the heterogeneous substrate, which is difficult to test. An example is Xyn10Z, which has an additional carbohydrate esterase that helps to make native xylan accessible for degradation, and another type of binding module (CBM6), which is described to bind to cellulose.

One type of hemicellulolytic enzyme which has not previously been addressed in the context of the cellulosome is the xyloglucanase hydrolysing the plant cell wall polysaccharide xyloglucan. This 1,4- β -glucan consists of heavily decorated cellotetraose subunits. Its hydrolysis greatly stimulates the cellulase activity in xyloglucan-cellulose copolymers (Irwin *et al.*, 2003) and may play a crucial role in the degradation of biomass at least from dicotyledonous plants. Xyloglucanase Xgh74A represents the first xyloglucanase identified in *C. thermocellum* and the first active xyloglucanase in a cellulosome. Only one ORF for a cellulosomal GHF74 protein has been identified so far: in the genome of *C. acetobutylicum* ATCC 824. It was annotated as sialidase, but was not characterized further and is probably a xyloglucanase (accession no. AE007608; Nolling *et al.*, 2001; Doi & Kosugi, 2004). This gene encodes the GHF74 protein in the databases that is the most homologous to Xgh74A (55 % identity).

The hydrolysis pattern of Xgh74A as one of the prominent subunits in the *C. thermocellum* cellulosome was investigated (Fig. 2). Few enzymes of GHF74 have been characterized to date; most of them are fungal β -glucanases (Yaoi & Mitsuishi, 2002; Hasper *et al.*, 2002; Irwin *et al.*, 2003). The GHF74 xyloglucanases show a variety of substrate specificities, mostly on the β -glucan backbone substituted with xylose. Xgh74A hydrolysed xyloglucan primarily into hepta- to nonasaccharides based on cellotetraose, seemingly XXXG, XLXG (or XXLG) and XLG using the known composition of the substrate (Fry *et al.*, 1993). The *Thermobifida fusca* enzyme Xeg74 shows a similar product pattern and size distribution (Irwin *et al.*, 2003).

This suggests that Xgh74A is a depolymerizing 1,4- β -glucanase like Xeg74 from *T. fusca* and cuts the xyloglucan backbone at the 1'-end of the unsubstituted glucose residue. Xeg74 was also reported to be an inverting enzyme (Irwin *et al.*, 2003), which implies that Xgh74A also inverts the anomeric carbon on hydrolysis.

The high activity of Xgh74A on mixed-linkage barley β -glucan is unusual amongst xyloglucanases. However, the *Thermotoga maritima* xyloglucanase Cel74 has its highest activity on barley β -glucan (Chhabra & Kelly, 2002). It shows 38 % sequence identity in the catalytic module. The only 3-D structure of a GHF74 enzyme has been obtained from the OXG-RCBH xyloglucanase of *Geotrichum* sp. M128 (Yaoi *et al.*, 2004). Although this fungal enzyme is rather an oligoxyloglucan reducing-end-specific cellobiohydrolase (Yaoi & Mitsuishi, 2002, 2004), its basic structure as a tandem repeat of two similar seven-bladed β -propeller subunits should be preserved in the homologous Xgh74A protein (Yaoi *et al.*, 2004). The proposed catalytic amino acids Asp-35 and Asp-465 in the active centre of OXG-RCBH are located in highly conserved regions and are preserved in Xgh74A (D70 and D480, respectively). The extra loop aa 375–485 in OXG-RCBH is missing in Xgh74A as it is in all other xyloglucanases (Yaoi *et al.*, 2004).

The removal of the xyloglucan in plant cell walls is obviously a precondition for efficient cellulose hydrolysis, and Xgh74A may be able to perform this task in *C. thermocellum*, as was demonstrated by incubating xyloglucan with purified cellulosomes. The reading frame of *xghA* is the only cellulosomal gene in the draft genome of *C. thermocellum* to which a xyloglucanase activity can be ascribed, and the only ORF with homology to GHF74 modules in the genome.

Zymogram staining of denaturing gel electrophoresis slabs from purified cellulosomes show only one band of activity (data not shown), indicating that no other β -glucanase is able to degrade tamarind xyloglucan sufficiently to be detected with the Congo red staining assay. Xgh74A is thus the only xyloglucanase in the cellulosome, opening the possibility to calculate the prevalence of Xgh74A in the cellulosome. Assuming an average mass of 900 kDa for a single cellulosome unit (containing one scaffoldin molecule and nine of the major catalytic components) and applying the formula $(\text{SpA}_{\text{Xgh74A}}/\text{SpA}_{\text{cellulosome}}) \times (\text{M}_{\text{Xgh74A}}/\text{M}_{\text{cellulosome}})$, where SpA is the specific activity and M is the molecular mass, it can be concluded that one Xgh74A molecule is present in $\sim 5.7 (\pm 0.7)$ cellulosomal units (Zverlov *et al.*, 2002). This calculated prevalence is in good agreement with the abundance of 0.2 per particle obtained from protein determinations for XghA in 2-D electrophoresis gels (Zverlov *et al.*, 2005).

Xgh74A can be ascribed a key role in the degradation of natural cellulosic material by the cellulosome of *C. thermocellum*, considering the protective role that xyloglucans have on cellulose fibres. This was shown by Irwin *et al.* (2003) with copolymers of bacterial cellulose and

xyloglucan: cellulose was completely protected from *T. fusca* cellulase degradation as long as the xyloglucan was present. Only removal of the xyloglucan by *T. fusca* xyloglucanase Xeg74 brought about the complete digestion of the cellulose fibres. Xgh74A should have a similar function in dismantling the cellulose fibres in biomass. It is thus no surprise that it is one of the most prominent hemicellulolytic enzymes in the extracellular cellulose-degrading enzyme complex of *C. thermocellum* (Zverlov *et al.*, 2005).

We have described two hemicellulolytic subunits in the cellulosome of *C. thermocellum*, XghA and Xyn10D, representing the only xyloglucanase in the cellulosome and a typical endo- β -xylanase, respectively. Both enzymes degrade hemicellulose components protecting cellulose in plant fibres from enzymic degradation. They thus contribute to the cellulose degradation by the cellulosomes of *C. thermocellum*.

ACKNOWLEDGEMENTS

This work was supported by grants from the Deutsche Forschungsgemeinschaft DFG and the Leonhard-Lorenz-Foundation to W. H. S., and from the A.-v.-Humboldt Foundation to V. V. Z. We are very grateful to W. L. Staudenbauer for countless stimulating discussions. M. Frommberger is thanked for his assistance with ESI-MS. The genomic sequence data were produced by the US Department of Energy Joint Genome Institute (<http://www.jgi.doe.gov/>).

REFERENCES

- Bayer, E. A., Shoham, Y. & Lamed, R. (2000). Cellulose-decomposing bacteria and their enzyme systems. In *The Prokaryotes: an Evolving Electronic Resource for the Microbiological Community*, 3rd edn. Edited by M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer & E. Stackebrandt. New York: Springer.
- Blum, D. L., Kataeva, I. A., Li, X. L. & Ljungdahl, L. G. (2000). Feruloyl esterase activity of the *Clostridium thermocellum* cellulosome can be attributed to previously unknown domains of XynY and XynZ. *J Bacteriol* **182**, 1346–1351.
- Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M., Davies, G. J. & Fontes, C. M. (2000). The X6 “thermostabilizing” domains of xylanases are carbohydrate-binding modules: structure and biochemistry of the *Clostridium thermocellum* X6b domain. *Biochemistry* **39**, 5013–5021.
- Chhabra, S. R. & Kelly, R. M. (2002). Biochemical characterization of *Thermotoga maritima* endoglucanase Cel74 with and without a carbohydrate binding module (CBM). *FEBS Lett* **531**, 375–380.
- Demain, A. L., Newcomb, M. & Wu, J. H. D. (2005). Cellulase, clostridia, and ethanol. *Microbiol Mol Biol Rev* **69**, 124–154.
- Doi, R. H. & Kosugi, A. (2004). Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nat Rev Microbiol* **2**, 541–551.
- Frommberger, M., Schmitt-Kopplin, P., Ping, G., Frisch, H., Schmid, M., Zhang, Y., Hartmann, A. & Kettrup, A. (2004). A simple and robust setup for on-column sample preconcentration – nano liquid chromatography-electrospray ionization mass spectrometry for the analysis of N-acetylhomoserine lactones. *Anal Bioanal Chem* **378**, 1014–1020.
- Fry, S. C., York, W. S., Albersheim, P. & 13 other authors (1993). An unambiguous nomenclature for xyloglucan-derived oligosaccharides. *Physiol Plant* **89**, 1–3.
- Grépinet, O., Chebrou, M. C. & Béguin, P. (1988). Purification of *Clostridium thermocellum* xylanase Z expressed in *Escherichia coli* and identification of the corresponding product in the culture medium of *C. thermocellum*. *J Bacteriol* **170**, 4576–4581.
- Hallstead, J. R., Vercoe, P. E., Gilbert, H. J., Davidson, K. & Hazlewood, G. P. (1999). A family 26 mannanase produced by *Clostridium thermocellum* as a component of the cellulosome contains a domain which is conserved in mannanases from anaerobic fungi. *Microbiology* **145**, 3101–3108.
- Hasper, A. A., Dekkers, E., van Mil, M., van de Vondervoort, P. J. I. & de Graaf, L. H. (2002). EglC, a new endoglucanase from *Aspergillus niger* with major activity towards xyloglucan. *Appl Environ Microbiol* **68**, 1556–1560.
- Hayashi, H., Takagi, K.-I., Fukumura, M., Kimura, T., Karita, S., Sakka, K. & Ohmiya, K. (1997). Sequence of *xynC* and properties of XynC, a major component of the *Clostridium thermocellum* cellulosome. *J Bacteriol* **179**, 4246–4253.
- Hayashi, H., Takehara, M., Hattori, T., Kimura, T., Karita, S., Sakka, K. & Ohmiya, K. (1999). Nucleotide sequences of two contiguous and highly homologous xylanases genes *xynA* and *xynB* and characterization of XynA from *Clostridium thermocellum*. *Appl Microbiol Biotechnol* **51**, 348–357.
- Irwin, D. C., Cheng, M., Xiang, B., Rose, J. K. & Wilson, D. B. (2003). Cloning, expression and characterization of a family-74 xyloglucanase from *Thermobifida fusca*. *Eur J Biochem* **270**, 3083–3091.
- Johnson, E. A., Madia, A. & Demain, A. L. (1982). Chemically defined minimal medium for growth of the anaerobic cellulolytic thermophile *Clostridium thermocellum*. *Appl Environ Microbiol* **41**, 1060–1062.
- Kurokawa, J., Hemjinda, E., Arai, T., Karita, S., Kimura, T., Sakka, K. & Ohmiya, K. (2001). Sequence of the *Clostridium thermocellum* mannanase gene *man26B* and characterization of the translated product. *Biosci Biotechnol Biochem* **65**, 548–554.
- Matuschek, M., Sahm, K., Zibat, A. & Bahl, H. (1996). Characterization of genes from *Thermoanaerobacterium thermosulfurigenes* EM1 that encode two glycosyl hydrolases with conserved S-layer-like domains. *Mol Gen Genet* **252**, 493–496.
- Morag, E., Bayer, E. A. & Lamed, R. (1992). Affinity digestion for the near-total recovery of purified cellulosome from *Clostridium thermocellum*. *Enzyme Microb Technol* **14**, 289–292.
- Nolling, J., Breton, G., Omelchenko, M. V. & 16 other authors (2001). Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J Bacteriol* **183**, 4823–4838.
- Schmitt-Kopplin, P. & Kettrup, A. (2003). Capillary electrophoresis–electrospray ionization–mass spectrometry for the characterization of natural organic matter: an evaluation with free flow electrophoresis–off-line flow injection electrospray ionization–mass spectrometry. *Electrophoresis* **24**, 3057–3066.
- Schwarz, W. H. (2001). The cellulosome and cellulose degradation by anaerobic bacteria. *Appl Microbiol Biotechnol* **56**, 634–649.
- Schwarz, W. H., Bronnenmeier, K., Gräbnitz, F. & Staudenbauer, W. L. (1987). Activity staining of cellulases in polyacrylamide gels containing mixed linkage β -glucans. *Anal Biochem* **164**, 72–77.
- Schwarz, W. H., Zverlov, V. V. & Bahl, H. (2004). Extracellular glycosyl hydrolases from clostridia. *Adv Appl Microbiol* **56**, 215–261.
- Shoham, Y., Lamed, R. & Bayer, E. A. (1999). The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. *Trends Microbiol* **7**, 275–281.

- Whitney, S. E. C., Brigham, J. E., Darke, A. H., Reid, J. S. G. & Gidley, M. J. (1995). In vitro assembly of cellulose/xyloglucan networks: ultrastructure and molecular aspects. *Plant J* **8**, 491–504.
- Wood, T. M. & Bhat, K. M. (1988). Methods for measuring cellulase activities. *Methods Enzymol* **160**, 87–112.
- Yaoi, Y. & Mitsuishi, Y. (2002). Purification, characterization, cloning, and expression of a novel xyloglucan-specific glycosidase, oligoxyloglucan reducing-end-specific cellobiohydrolase. *J Biol Chem* **277**, 48276–48281.
- Yaoi, K. & Mitsuishi, Y. (2004). Purification, characterization, cDNA cloning, and expression of a xyloglucan endoglucanase from *Geotrichum* sp. M128. *FEBS Lett* **560**, 45–50.
- Yaoi, K., Kondo, H., Noro, N., Suzuki, M., Tsuda, S. & Mitsuishi, Y. (2004). Tandem repeat of a seven-bladed β -propeller domain in oligoxyloglucan reducing-end-specific cellobiohydrolase. *Structure* **12**, 1209–1217.
- York, W. S., Harvey, L. K., Guillen, R., Albersheim, P. & Darvill, A. G. (1993). Structural analysis of tamarind seed xyloglucan oligosaccharides using β -galactosidase digestion and spectroscopic methods. *Carbohydr Res* **248**, 285–301.
- Zverlov, V. V., Fuchs, K. P., Schwarz, W. H. & Velikodvorskaya, G. A. (1994). Purification and cellulosomal localization of *Clostridium thermocellum* mixed linkage β -glucanase LicB (1,3-1,4- β -D-glucanase). *Biotechnol Lett* **16**, 29–34.
- Zverlov, V. V., Fuchs, K.-P. & Schwarz, W. H. (2002). Chi18A, the endochitinase in the cellulosome of the thermophilic, cellulolytic bacterium *Clostridium thermocellum*. *Appl Environ Microbiol* **68**, 3176–3179.
- Zverlov, V. V., Kellermann, J. & Schwarz, W. H. (2005). Functional subgenomics of *Clostridium thermocellum* cellulosomal genes: identification of the major catalytic components in the extracellular complex and detection of three new enzymes. *Proteomics* **5** (in press). doi:10.1002/pmic.200401199