Cellobiohydrolase B, a second exo-cellobiohydrolase from the cellulolytic bacterium *Cellulomonas fimi*

Hua SHEN, Neil R. GILKES, Douglas G. KILBURN, Robert C. MILLER, Jr. and R. Antony J. WARREN*

Department of Microbiology and Immunology, University of British Columbia, 6174 University Boulevard, Vancouver, British Columbia, Canada V6T 1Z3

The gene *cbhB* from the cellulolytic bacterium *Cellulomonas fimi* encodes a polypeptide of 1090 amino acids. Cellobiohydrolase B (CbhB) is 1037 amino acids long, with a calculated molecular mass of 109765 Da. The enzyme comprises five domains: an N-terminal catalytic domain of 643 amino acids, three fibronectin type III repeats of 97 amino acids each, and a C-terminal cellulose-binding domain of 104 amino acids. The catalytic domain belongs to family 48 of glycosyl hydrolases. CbhB has a

very low activity on CM-cellulose. Viscometric analysis of CM-cellulose hydrolysis indicates that the enzyme is an exoglucanase. Cellobiose is the major product of hydrolysis of cellulose. In common with two other exoglucanases from *C. fimi*, CbhB has low but detectable endoglucanase activity. CbhB is the second exo-cellobiohydrolase found in *C. fimi*. Therefore, the cellulase system of *C. fimi* resembles those of fungi in comprising multiple endoglucanases and cellobiohydrolases.

INTRODUCTION

Fungal cellulase systems generally contain multiple endo- β -1,4-glucanases (EC 3.2.1.4) and exo- β -1,4-glucanases (EC 3.2.1.91), but there are relatively few reports of even single exoglucanases in bacterial cellulase systems [1–8]. A 95 kDa extracellular cellulose-binding protein (Cbp95) produced by *Cellulomonas fimi* [9] is an exo-cellobiohydrolase, designated CbhA [10], belonging to family 6 of glycosyl hydrolases. A 120 kDa extracellular cellulose-binding protein (Cbp 120) [9] was thought initially to be an inverting endoglucanase, designated CenE, belonging to family 48 of glycosyl hydrolases [11]. This paper reports the sequence of the gene encoding Cbp120 and further characterization of the polypeptide that it encodes. Cbp120 is undoubtedly an exo-cellobiohydrolase, now designated CbhB. Like two other exoglucanases from *C. fimi*, Cex and CbhA, CbhB has weak endoglucanase activity.

EXPERIMENTAL

Materials

Phosphoric acid-swollen (PAS)-cellulose and bacterial microcrystalline cellulose (BMC-cellulose) were prepared as described previously [12,13]. CM-cellulose (Na salt; low-viscosity grade; degree of substitution, 0.7; degree of polymerization, 400), 4-O-methyl-D-glucurono-D-xylan-Remazol Brilliant Blue (RBB-xylan), galactomannan, glucomannan, β -lichenan, and *p*-nitrophenyl β -D-cellobioside, β -D-glucoside, β -D-xyloside and β -D-galactoside (pNPC, pNPG, pNPX and pNPGal respectively) were from Sigma. Birch wood xylan was from Carl Roth KG. barley β -glucan, azo-barley glucan and oat β -glucan were from Megazyme. Cello-oligosaccharides were from Seikagaku America.

Agrobacterium sp. β -glucosidase (Abg), *C. fimi* endo- β -1,4-glucanase A (CenA), *C. fimi* cellobiohydrolase A (CbhA), a mixed-function *C. fimi* exo- β -1,4-glucanase, and β -1,4-xylanase (Cex) and its mutant (Cex E127A) were prepared as previously described [10,14–16].

Strains, vectors and media

The *Escherichia coli* strains used were XL1-Blue and XL1-Blue MRF' from Stratagene, RZ1032 [17] and JM101. Bacteriophage M13K07 was used to produce single-stranded DNA for sequencing and site-directed mutagenesis. The plasmids used were pBluescript SK and KS (–/+) from Stratagene, pTZ18R [18], pTZSH1 [11] and pTug [19].

Cultures were grown at 30° or 37 °C in Luria broth or tryptone/yeast extract/phosphate medium [20], supplemented with 100 μ g/ml ampicillin or 50 μ g/ml kanamycin. Solid media contained 1.5 % agar.

Recombinant DNA techniques

The nucleotide sequences of both strands of *cbhB* were determined by the dideoxy method [21] using single-stranded DNA, T7 DNA polymerase, and 7-deaza-dGTP instead of dGTP. Nucleotide and predicted amino acid sequences were analysed with the program PCGENE from Intelligenetics.

The initial clone of *cbhB* in the plasmid pTZSH1 [11] produced very little CbhB. Therefore, much of the non-coding sequence upstream of *cbhB* was deleted, so that the gene could be expressed from the *lac* promoter in the expression vector p3SN8, a derivative of pTug [19]. A *StuI* site (shown in **bold**) was placed next to the translation initiation codon of *cbhB* by site-directed mutation of pTZSH1 with the oligonucleotide 5'-ACCAGAAA-GGTAAGGCCTATGTCGTCAACG-3', giving pTZSH1s. Plasmid pTZSH1s was digested with *StuI* and *Bam*HI, and the 3.3 kb fragment carrying *cbhB* was recovered. Plasmid p3SN8 was digested with *NcoI*, then treated with mung-bean nuclease, and finally digested with *Bam*HI. It was ligated to the 3.3 kb fragment from pTZSH1 to give pTugSH2. The construct was checked by DNA sequencing. The sequence encoding the CbhB leader peptide in pTugSH1 was replaced with that encoding the Cex leader peptide as follows. A *NheI* site (shown in **bold**) was

Abbreviations used: Abg, *Agrobacterium* sp. β -glucosidase; BMC-cellulose, bacterial microcrystalline cellulose; CBD, cellulose-binding domain; CbhA, *C. fimi* cellobiohydrolase A; CenA, *C. fimi* endo- β -1,4-glucanase A; Cex, a mixed-function *C. fimi* exo- β -1,4-glucanase and β -1,4-xylanase; Cex E127A, a Cex mutant; pNPC, pNPGal, pNPG and pNPX, *p*-nitrophenyl β -D-cellobioside, β -D-galactoside, β -D-glucoside and β -D-xyloside, respectively; PAS-cellulose, phosphoric acid-swollen cellulose; RBB-xylan, 4-O-methyl-D-glucurono-D-xylan-Remazol Brilliant Blue.

* To whom correspondence should be addressed.

The nucleotide sequence of *cbhB* has been assigned GenBank accession number L38827.

introduced at the start of the sequence in pTZSH1 encoding native CbhB by site-directed mutation using the oligonucleotide 5'-CCGGTTCGCGCCGCTAGCGACGGCGAGTACGC-3' giving pTZSH1n; this changed the second amino acid of CbhB from valine to serine (codon shown underlined). Plasmid pTZSH1n was digested with *Bam*HI. After filling in the ends using the Klenow fragment of DNA polymerase I, the plasmid was digested with *Nhe*I. The 3.3 kb fragment containing the sequence encoding CbhB was recovered. The vector pTug was digested with *Hind*III; the ends were filled in using Klenow fragment; the linear plasmid was then digested with *Nhe*I. The 3.3 kb fragment was ligated to the linear pTug to give pTugSH3. The construct was checked by DNA sequencing. The Cex leader peptide, which had been substituted for the CbhB leader peptide in the construct, allowed efficient export of CbhB to the periplasm of *E. coli* JM101/pTugSH3.

Purification and sequence analysis of CbhB

A 6-litre culture of *E. coli* JM101/pTugSH3 or *E. coli* JM101/pTugSH2 was grown at 30 °C to an A_{600} of 6; then isopropyl β -D-thiogalactoside was added to a final concentration of 0.1 mM. The cells were recovered by centrifugation 3 h later and resuspended in 1.2 litres of 20 % sucrose/0.1 M Tris/HCl (pH 8.0)/1 mM EDTA/0.02 % NaN_3 . After standing on ice for 5 min, the cells were recovered by centrifugation and resuspended in 1.2 litres of 5 mM MgSO_4 . The cells were removed by centrifugation and the resulting periplasmic extract was passed over a column of CF1 cellulose. The cellulose was washed with 1 M NaCl/50 mM potassium phosphate, pH 7.0, and then with 50 mM potassium phosphate, pH 7.0. Adsorbed protein was eluted with distilled water. The water eluate was concentrated by ultrafiltration through a PM30 membrane (Amicon).

CbhB (4 nmol) was digested with papain (4 m-units) for 10 min at 37 °C in 150 μ l of 100 mM Tris/HCl (pH 8.0)/2 mM EDTA/0.02 % NaN_3 /5 mM cysteine. The digest (15 μ l) was subjected to SDS/PAGE. A proteolytic fragment of CbhB was recovered from the gel and its N-terminal amino acid sequence determined as described previously [9]. The N-terminal sequences of recombinant CbhBs were also determined after SDS/PAGE.

Enzyme and protein assays

Hydrolysis of CM-cellulose, barley β -glucan, oat β -glucan and birch wood xylan was assayed by measuring reducing-sugar production. The rates of CM-cellulose hydrolysis given in Table 1 were calculated from the linear portions of the curves shown in Figure 3(b). Hydrolysis of CM-cellulose was also assayed by viscometry [10]. Hydrolysis of RBB-xylan was assayed as previously described [22]. Hydrolysis of azo-barley glucan was assayed by the procedure of the supplier (Megazyme). Hydrolysis of β -glucomannan, β -galactomannan and β -lichenan was assayed by incubating 2.0 nmol CbhB in 0.5 ml of 50 mM potassium phosphate, pH 7.0, containing 0.10–0.18 % substrate and 0.02 % NaN_3 , at 37 °C for 2 h, and then measuring reducing sugars with dinitrosalicylic acid reagent [23].

Hydrolysis of cellulose by CbhA, CbhB or CenA was assayed by incubating 1.5 nmol of enzyme and 1.5 mg of BMC-cellulose or 15 mg of PAS-cellulose in 1.5 ml of 50 mM potassium phosphate (pH 7.0)/0.02 % NaN_3 for 1 h, then measuring total soluble sugars by the phenol/ H_2SO_4 assay [10]. Soluble cello-oligosaccharides released from BMC-cellulose and PAS-cellulose were analysed by HPLC, as described previously [10].

CM-cellulose hydrolysis was also examined by using Congo Red dye [24]. Enzymes were diluted in 50 mM potassium

phosphate, pH 7.0, to concentrations of 1, 10 and 100 pmol/ μ l, and 20 μ l of each dilution was loaded on a 1.5 % Difco agar plate containing 0.1 % CM-cellulose in 50 mM potassium phosphate, pH 7.0. The plate was incubated overnight at 30 °C, then flooded with 2 mg/ml Congo Red solution. After 15 min, the Congo Red was rinsed off with distilled water and the plate washed with 1 M NaCl. The plate was then photographed.

Absorption coefficients (280 nm; 1 mg/ml; 1 cm) for purified proteins were estimated by the u.v.-absorbance method of Scopes [25]. The coefficient obtained for CbhB was 2.50, in good agreement with the value of 2.46 predicted from the tryptophan and tyrosine contents of the protein [26]. The corresponding values for CbhB with the extra 20 amino acids at the N-terminus were 2.56 and 2.50.

Prediction of leader-peptide processing sites

Probable leader-peptidase cleavage sites in CbhB were predicted as previously described [27] by using the program AnalyzeProk-Signalase (N. Mantei, mantei@neuro.biol.ethz.ch). The program uses a weight matrix calculated from 68 prokaryotic leader peptides, not including peptides processed by signal peptidase II (N. Gilkes, unpublished work).

RESULTS

Nucleotide sequence of *cbhB*

The *cbhB* gene was cloned previously on a 3.8 kb fragment of *C. fimi* DNA [11]. The complete sequence of the gene was obtained from appropriate restriction fragments, which were sub-cloned into the vectors pTZ18R and pBluescript. Both strands were sequenced. The gene encodes a polypeptide of 1090 amino acids (Figure 1). The ATG translation-initiation codon is preceded by a putative ribosome-binding site. The coding sequence of *cbhB* terminates 32 nt before the end of the 3.8 kb fragment of *C. fimi* DNA. There are no repeats within these 32 nt; presumably, the transcription termination signals are located further downstream. There are two overlapping inverted repeats of 11 and 17 nt, 261 to 322 nt upstream of the translation initiation codon for *cbhB*, and 3 to 64 nt downstream of another open reading frame.

Amino acid sequence of CbhB

Residues 1–22 of the predicted amino acid sequence of the polypeptide encoded by *cbhB* match the N-terminal amino acid sequence determined previously for Cbp120 [9], except for three residues. The protein from *C. fimi* was re-sequenced [11]; the sequence of the first 25 amino acids corresponded exactly to residues 1–25 of the predicted sequence (Figure 1). The N-terminal sequence of a fragment of recombinant CbhB released by papain also corresponded exactly to residues 403–408 of the predicted amino acid sequence (Figure 1). CbhB is 1037 amino acids long, with a calculated molecular mass of 109765 Da.

The N-terminus determined for CbhB from *C. fimi* indicates that the enzyme has an unusually long leader peptide of 53 amino acid residues. The most probable leader-peptide processing site predicted by the algorithm of von Heijne [27] was between Ala –21 and Ala –20 (Figure 1), i.e. the processing site determined for CbhB from *E. coli* JM101/pTugSH2 (see below); the Ala –1/Ala +1 site seen for *C. fimi* received the sixth highest score (results not shown).

The domain organization of CbhB is similar to that of CbhA, CenB and CenD from *C. fimi*. Amino acids 934–1037 at the C-terminus comprise a type II CBD, a feature characteristic of most β -1,4-glycanases from *C. fimi* [28]. The CBD is preceded by three fibronectin type III repeats, similar to those present in

CstAvilI	XSDDPYKQRFLLDLWDDLHDPSSNGYFSXH-GIPYHAV	35
CfiCbhB	AVDGEYAQRFLAQYDKIKDPANGYFSA_QGIPYHAVETLMVEAPDYQHETTSEAYSYWLWLEALYQQVTQDWA	72
CsaCelA	SGLGKYGQRFMWLNKIHDPASGYFNQD_GIPYHSVETLICEAPDYGHLLTSEAFSYVWLEAVYGLKLTGDWS	72
CthCelS	KDGTSYKDLFLELYGKIKDPKNGYFSPDEGIPYHSIETLIVEAPDYGHVTTSEAFSYVWLEAMYGNLTGNWS	80
CfiCbhB	PLNHAWDTMEKYMIPQSVDPQNTSFYNPNPATYAPFNHPPSSYPSQLN_SGISGQTDPIGAELKATYGNADV	144
CsaCelA	KFKTAWDTLEKYMIPSAEDQP_MRSYDPNKPATYAGEWETPDKYSPLEFN_VPVGKDLPHNELVSTYGSTLM	143
CthCelS	GVETAWKVMEDWIIPDSTEQPGMSSYNPNPATYADEYEDPSYYPSELKFDTVRVGSDPVHNDLVSAYGPN_M	152
CfiCbhB	YQMHWLADVDNIYQFGATPGAGCTLGPTATGTSFINTFQRCQESVWETVPQPSCEEFKYGGKNGYLDLFTKD	217
CsaCelA	YGMHWLMDVDNWYGYG_____KRGDGVSRASFINTFQRCPEESVWETVPHPSWEEFKWGGPNGFLDLFIKD	209
CthCelS	YLMHWLMDVDNWYGF_____TG_____TRATFINTFQRCQESTWETIPHPSIEEFKYGGPNGFLDLFTKD	214
CfiCbhB	ASYAKQWKYTSASDADARAVEAVYWANQWATEQKKAADVAATVAKAAKMGDYLRYTLFDKYFKKIGCTSPTCA	290
CsaCelA	QNYSKQWRYTNAPDADARA IQATYWAKVWAKEQGFNEISSYVGKAAKMGDYLRYAMFDKYFKPLGCQDKNAA	282
CthCelS	RSYAKQWRYTNAPDAEGRAIQAVYWANKWAKEQKGSAVASVSKAAKMGDFLRNDMFDKYFMKIGAQDKTPA	285
CfiCbhB	AGQGREAAHYLLSWYMAWGGATDTSSGAWRIGSSHAHFGYQNPPLAAWALSTDPKLTPKSP TAKADWAASMQR	363
CsaCelA	GGTGYDSAHYLLSWYAWGGALDGA_WSWKIGCSHAHFGYQNPMAAWALANDSDMKPKSPNGASDWAKSLKR	353
CthCelS	___TGYSADHYLMAWYTAWGGGIGAS_WAWKIGCSHAHFGYQNPFGQWVSATQSDFAPKSSNGKRDWTTSYKR	356
CfiCbhB	QLEFYTWLQASNGGIAGGATNSWDGAYAQQPAGTPTFYGMGYTEAPVYVDPSPSNRWFGMQAWGVQORVAELYA	436
CsaCelA	QIEFYRWLQSAEGA IAGGATNSWNGRYEKYPAGTATFYGMAYEPNPVYRDPGSNTWFGFQAWSMQORVAEYV	426
CthCelS	QLEFYQWLQSAEGGIAGGATNSWNGRYEKYPAGTSTFYGMAYVPHPVYADPGSNQWFGFQAWSMQORVMEYYLE	429
CceICCF	EFYQWLQSAEGA IAGGATNSWNGRYEAVPSGTSTFYGMGYVENPVYADPGSNWFGMQVWSMQORVAELYK	71
CjosORF1	EFYQWLQSSEGA IAGGATNSWNGRYESI PSGTSTFYGMGYVENPVYADPGSNWFGMQVWSMQORVAELYK	71
CfiCbhB	SGNAQAKKILDKWVPWVANI__STDGASWKVPSELKWTGKPDWNAAP_TGNPGLTVEVTSYGQDVGVAAAD	506
CsaCelA	TGDKDAGTLLKVVSWIKSVVVLNSDG_TFAIPSTLDWSGQPDWTNGT__YTGNPNLHVKVVDYGTDLGITAS	496
CthCelS	TGDSSVKNLIKWVDWVMSEIKLYDDG_TFAIPSDLEWSGQPDWTGT__YTGNPNLHVRVTSYGTDLGVAGS	499
CceICCF	TGDARAKLLDKWAKWINGEIKFNADG_TFQIPSTIDWEGQPDWTNPQTQGYTGANLHVKVVNNGTDLGCASS	143
CjosORF1	TGDTRAKNLLDKWAKWVNSEIKFNADG_TFQIPGLDWEQPDWTNPQTQGYTGANLHVKVVNNTDLGCASS	143
CfiCbhB	TARALLFYAAKSG_____DTASRDKAKALLDAIWAN_NQDPLGVSAVETRGDYKRFDDTYVANGDGIYIPS	571
CsaCelA	LANALLYYSAGTRKY_GVFDEEAKNLAKELDRMWKLY_RDEKGLSAPEKRADYKRFFEQEV_____YIPA	560
CthCelS	LANALATYAAATERWEGKLDTKARDMAAELVNRAWYNFYCSEGKGVVTEEARADYKRFFEQEV_____YVPA	560
CceICCF	LANTLTYYAAKSG_____DETSRQNAQKLLDAMWNNY_SDSKGISTVEQRGDYHRFLDQEV_____FVPA	212
CjosORF1	LANTLTYYAAKSG_____DTTSKENAKKLLDAMWNNY_SDSKGISTIEQRGDYHRFLDQEV_____YVPA	212
CfiCbhB	GWTGTPNGDVIKPGVVSFLDIRSFYKKDPNWSKVQTFLDGGAEPQFRYHRFWAQTAVALADYARLF...	639
CsaCelA	GWTGKMPNGDVIKSGVKFIDIRSKYQDPDWPKEAAYKSGQVPEFRYHRFWAQCDIAIVNATYEILF...	628
CthCelS	GWSGTPNGDKIQPGIKFIDIRTKYRQDPYDIDVYQAYLRGEAPVLNYHRFWHEVDLAVAMGVLATYF...	629
CceICCF	GWTGKMPNGDVIKSGVKFIDIRSKYQDPEWQTMVAALQAGQVPTQRLHRFWAQSEFAVANGVYAILF...	270
CjosORF1	GWTGKMPNGDVIKSGVKFIDIRSKYQDPEWQTMVAALQAGQVPTQRLHRFWAQSEFAVANGVYAILF...	270

Figure 2 Alignment of the amino acid sequences of the catalytic domains of family-48 β -1,4-glucanases

Numbers refer to the amino acid at the end of each line; sequences are numbered from the first amino acid of the mature protein. Hyphens show gaps introduced to improve the alignment. CfiCbhB is *C. fimi* CbhB (GenBank accession number L38827); CthCelS is *Clostridium thermocellum* CelS (GenBank accession number L06942); CsaCelA is *Caldocellum saccharolyticum* CelA (GenBank accession number L32742); CceICCF is *Clostridium cellulosyticum* CelCCF (GenBank accession number M87018); CjoORF1 is from *Clostridium josui* (GenBank accession number D16670); CstAvilI is the N-terminal amino acid sequence of Avicelase II from *Clostridium stercorarium* [1]. Amino acid residues that are conserved in all known family-48 enzymes are shown in **bold** in the sequence of CfiCbhB.

Table 1 Comparison of the activities of CbhB, CbhA and CenA on various cellulosic substrates

Enzyme	Substrate		
	CM-cellulose*	BMC-cellulose†	PAS-cellulose‡
CbhB	0.7	0.7	0.4
CbhA‡	3.7	1.4	1.2
CenA	8130	0.3	10.6

* Activity expressed as mol of reducing-sugar groups generated/min per mol of enzyme.

† Activity expressed as g of soluble sugar released/min per mmol of enzyme.

‡ Data from [10].

CenB, CenD and CbhA [9,10,29]. The N-terminal catalytic domain is 643 amino acids long. As in CenD and CbhA, there are no obvious linkers between the domains.

Relationship of CbhB to other β -1,4-glucanases

Alignment of the N-terminal amino acid sequence of CbhB (previously called CenE) with those of three β -1,4-glucanases from *Clostridium* spp. showed that the enzymes were related [11]. On the basis of other reported sequence similarities in the literature, we proposed that the catalytic domains of *C. fimi* CbhB, *Clostridium thermocellum* CelS [7], *C. thermocellum* S8 [30], *Clostridium stercorarium* Avicelase II [1], *Clostridium cellulolyticum* CelCCF [31], *Clostridium cellulovorans* P70 [32], and the C-terminal catalytic domain of *Caldocellum saccharolyticum* CelA [33], comprise a new family of β -1,4-glucanases called family L [11]. An open reading frame (ORF1) from *Clostridium josui* encodes a further, related, polypeptide [34]. Family L was subsequently designated family 48 in the general classification of glycosyl hydrolases ([35,36], and B. Henrissat, personal communication). The sequences of the catalytic domains of CbhB, CelS and CelA are 47% identical; only partial amino acid sequences are currently available for CelCCF and the *C. josui* polypeptide (Figure 2). The similarities reported for Avicelase II and S8 are based on N-terminal amino acid sequence data. An amino acid sequence for P70 is not yet available.

Preparation of CbhB from *E. coli*

CbhB was produced initially in *E. coli* JM101/pTugSH2, from wild-type *cbhB*. The N-terminal amino acid sequence of the CbhB obtained was Ala-Ala-Gly-Ala-Gln-Pro-Thr..., indicating processing between Ala -21 and Ala -20 (Figure 1). This is the most probable leader-peptide processing site according to the algorithm of von Heijne [27], but is different from the site observed in *C. fimi*. It was decided to engineer *cbhB* so that the enzyme produced by *E. coli* was the same length as that produced by *C. fimi*. Accordingly, the sequence encoding the CbhB leader peptide was replaced with that encoding the leader peptide of *C. fimi* Cex, which functions efficiently in *E. coli* [37]. The resulting plasmid was designated pTugSH3.

CbhB was purified from the periplasm of *E. coli* JM101/pTugSH3 and JM101/pTugSH2 by osmotic shock followed by affinity chromatography on cellulose (see the Experimental section). Approx. 190 mg of purified CbhB were obtained from a 6-litre culture in both cases. The N-terminal amino acid sequence of the protein produced with pTugSH3 matched that of the protein obtained from *C. fimi*, except for the second residue; this was a serine, not a valine, because of the procedure used to

Figure 3 Hydrolysis of CM-cellulose by CbhB, CbhA and CenA

(a) Relationship between specific fluidity (ϕ) and production of reducing-sugar groups during hydrolysis. (b) Release of reducing sugars versus time. The reaction mixture contained 4 μ M CbhB (■), 3.6 nM CenA (▲) or 500 nM CbhA (●), and 3.2% (w/v) CM-cellulose in 50 mM sodium citrate, pH 7. Data for CenA and CbhB are from [10].

subclone *cbhB* into pTug. CbhB produced with pTugSH3 had the same substrate specificity as that produced with pTugSH2.

Hydrolytic activities of CbhB

CbhB did not hydrolyse the following substrates: barley β -glucan, oat β -glucan, azo-barley- β -glucan, RBB-xylan, birch wood xylan, β -glucomannan, β -galactomannan, β -lichenan, pNPC, pNPG, pNPX and pNPGal.

CbhB was known to hydrolyse cellobiose to a mixture of cellotetraose, cellotriose and cellobiose, with only a trace of glucose [11]. It hydrolysed cellopentaose to cellotriose and cellobiose, and cellotetraose to cellobiose, but it did not hydrolyse cellotriose (results not shown).

The activities of CbhB on cellulosic substrates were compared with those of CenA, a *C. fimi* endoglucanase [38], and CbhA, a *C. fimi* exo-cellobiohydrolase [10] (Table 1). CbhB was more active than CenA, but less active than CbhA, on BMC-cellulose.

Figure 4 Soluble sugars released from cellulose by CbhB

Soluble sugars produced by hydrolysis of BMC-cellulose (a) or PAS-cellulose (b) for 20 h were resolved by HPLC. Glucose (1), cellobiose (2) and cellotriose (3) were resolved as single peaks.

In 1 h, the enzymes solubilized 3%, 1% and 9% of the BMC-cellulose, respectively. The apparent K_m and V_{max} for CbhB, determined from a Hanes plot (correlation coefficient $r^2 = 0.996$) were 0.5 mg of BMC-cellulose/ml and 1.0 g of soluble sugar released/min per mmol of enzyme, respectively (results not shown). The apparent K_m and V_{max} values previously determined for CbhA were 0.3 mg of BMC-cellulose/ml and 1.8 g of soluble sugar released/min per mmol of enzyme (correlation coefficient of Hanes plot = 0.991) [10,39]. CbhB was the least active, and CenA the most active, of the three enzymes on CM-cellulose or PAS-cellulose (Table 1). The rates for CenA on CM-cellulose was approx. 1.2×10^4 and 2×10^3 times those for CbhB and CbhA respectively.

The CM-cellulose used in this study had a nominal degree of polymerization of 400 and a nominal degree of substitution of 0.7. The calculated average molecular mass was 86400 Da. Incubation of $4 \mu\text{M}$ CbhB for 26 h with 3.2% (i.e. $370 \mu\text{M}$) CM-cellulose released approx. $200 \mu\text{M}$ reducing sugar (Figure 3b), equivalent to 0.55 scission per CM-cellulose molecule on average. The plot of specific fluidity (ϕ) versus reducing-sugar production during the hydrolysis of CM-cellulose is an indicator of exoglucanase or endoglucanase activity, because the random action of an endoglucanase results in a greater increase in specific fluidity than obtained by the restricted action of an exoglucanase [10,40]. The slopes for CbhB, CbhA and CenA were 20, 20 and $96 \phi \cdot \text{ml} \cdot \text{mmol}^{-1}$ respectively (Figure 3a). Therefore, CbhB behaved like an exoglucanase, not an endoglucanase.

The major soluble product released from BMC-cellulose by CbhB was cellobiose, with traces of glucose and cellotriose (Figure 4a). Cellobiose was also the major soluble product released from PAS-cellulose, but there was more cellotriose than

Figure 5 Endoglucanase activity of exoglycanases

Enzymes were loaded on a plate containing CM-cellulose and agar (see the Experimental section). After incubation at 30°C overnight, the plate was stained with Congo Red. The enzymes were endoglucanase A (CenA), cellobiohydrolases A and B (CbhA and CbhB), the mixed-function xylanase–glucanase (Cex), and an inactive mutant of Cex (E127A), all from *C. fimi*, and Abg, a β -glucosidase from an *Agrobacterium* sp. Loads of 20, 200 or 2000 pmol of CbhA, CbhB, Cex and Abg were used; only one loading of CenA (20 pmol) was used.

from BMC-cellulose (Figure 4b). Cellotetraose did not accumulate in either case. The products were consistent with exo-cellobiohydrolase activity. In contrast, *C. fimi* endoglucanases CenA, CenB and CenD produce significant levels of glucose and cellotriose [39].

Endoglucanase activity of exoglycanases from *C. fimi*

The interaction of Congo Red with CM-cellulose is commonly used to differentiate endo- and exo-glucanases. Congo Red forms a red complex with oligosaccharides containing a minimum of five monosaccharides linked by β -1,3 or β -1,4 bonds [41]. When CM-cellulose is incorporated into agar, hydrolysis of the CM-cellulose to products with fewer than five glucose units is revealed by staining with Congo Red. Endoglucanases produce a yellow 'clear' zone against a red background. Exoglucanases appear to lack hydrolytic activity, because they cannot decrease the sizes of the glucan chains sufficiently to prevent interaction with Congo Red [24,40,41]. CbhB was first thought to be an endoglucanase, because it produced a zone of clearing [11], but its activity in the viscosity assay clearly showed it is an exo-cellobiohydrolase. This raised the possibility that it has some intrinsic endoglucanase activity.

Two other enzymes with exoglycanase activity have been characterized from *C. fimi*: an exoglucanase, Cex, which is most active on xylan, but which also hydrolyses CM-cellulose slowly [40,42], and CbhA [10]. CbhA, CbhB and Cex hydrolyse CM-cellulose in the manner characteristic of exoglucanases ([10] and Figure 3), giving relatively low slopes when specific fluidity is plotted against reducing-sugar release, but only CbhB appeared to give clearing on CM-cellulose plates [11]. The clearing assays, however, used relatively low concentrations of enzyme.

When higher concentrations of enzyme were tested, all three enzymes produced zones of clearing in proportion to the amount of enzyme deposited on the plate (Figure 5). The three exoglucanases clearly had weak endoglucanase activity. Although

the catalytic activity of CbhB and the other exoglucanases on CM-cellulose is low under the conditions of the viscometric assay (see above), the conditions used in the CM-cellulose–Congo Red plate assay are different. In order to see clearing, it was necessary to apply a 20 μ l sample of 100 μ M CbhA, CbhB or Cex to the plate. There is very little penetration of the agar by the dye during staining, so the assay monitors only the hydrolysis of a small volume of 0.1% CM-cellulose near the plate surface. Presumably, it is these conditions that allow detection of the endoglucanase activities of the enzymes.

Under the same assay conditions, Cex E127A, an inactive mutant of Cex in which the acid/base catalyst in the active site is mutated to alanine [16], did not cause clearing. The endoglucanase CenA was at least 100 times more active than the exoglucanases in clearing CM-cellulose. Abg is a β -glucosidase from an *Agrobacterium* sp. that removes glucosyl residues from the non-reducing end of cellotriose and longer cello-oligosaccharides [43]. It is also active on CM-cellulose and gives a low slope (13 $\phi \cdot \text{ml} \cdot \text{mmol}^{-1}$) characteristic of an exoglucanase in the viscometric assay [44]. Abg did not cause clearing in the CM-cellulose–Congo Red plate assay, showing that this exoglucanase does not have endoglucanase activity.

DISCUSSION

The cellulase system of *C. fimi* contains multiple endoglucanases [9,38,40,45] and exo-cellobiohydrolases ([10]; the present paper). In this respect, the *C. fimi* cellulase system resembles systems from fungi such as *T. reesei*. The cellulosome of the bacterium *C. thermocellum* was thought initially to be devoid of exoglucanase; simultaneous ‘multicutting’ of the cellulose chain by endoglucanases was thus proposed to explain cellulose hydrolysis by this organism [46]. However, exoglucanases are present in cellulolytic bacteria, including *Clostridia*, indicating that bacteria and fungi probably use a similar strategy to hydrolyse cellulose.

The cellulase systems of both groups of organisms are proving to be complex. For example, the *T. reesei* system was thought to comprise two endoglucanases and two exo-cellobiohydrolases, but at least two other endoglucanases are now known to be present [47,48]. In *Phanerochaete chrysosporium*, a white-rot fungus, typical endoglucanases appear to be absent, and it is suggested that cellobiohydrolase variants provide endoglucanase activity in this organism [49]. The interactions between the individual components of cellulase systems are not yet well understood in any of the bacteria or fungi.

All known fungal exo-cellobiohydrolases belong to either family 6 or family 7 of glycosyl hydrolases [35,36]. CbhA from *C. fimi* belongs to family 6, but there are no known bacterial enzymes in family 7. All of the known enzymes in family 48 are bacterial. At least three of these are exo-cellobiohydrolases: S8, and its analogue CelS, from *C. thermocellum* [6,30,50], Avicelase II from *Clostridium stercoarium* [1], and CbhB. Perhaps family-48 exo-cellobiohydrolases are the functional equivalents in bacteria of the fungal enzymes in family 7. It has proved difficult to detect activity in some family-48 enzymes, such as the C-terminal catalytic domain of CelA from *C. saccharolyticum* [33] and P70 from *C. cellulovorans* [32], perhaps because of the lack of a convenient and sensitive assay for exo-cellobiohydrolase activity. Hydrolysis of pNPC has been considered diagnostic of exo-cellobiohydrolase activity [51], but neither CbhA nor CbhB hydrolyse this substrate.

The sequences of four endoglucanase genes, *cenA*, *cenB*, *cenC* and *cenD*, two exo-cellobiohydrolase genes, *cbhA* and *cbhB*, and two xylanase genes, *xynD* and *cex* (also called *xynB*), from *C. fimi* are known ([9,10,29,38,52–54]; the present paper). The

catalytic domains of the enzymes that they encode come from six different families of glycoside hydrolases: family 5 (CenD); family 6 (CenA and CbhA); family 9 (CenB and CenC); family 10 (Cex); family 11 (XynD); family 48 (CbhB). All of the enzymes are modular [28]. They fall into four groups in terms of module composition: CenA and Cex comprise a catalytic domain and a type II cellulose-binding domain (CBD); CenB, CenD, CbhA and CbhB comprise a catalytic domain, a type II CBD and two or three fibronectin type III repeats; CenC comprises a catalytic domain, two tandem type IV CBDs, and two tandem sequence-related modules of unknown nature; and XynD comprises a catalytic domain, a type II CBD, a second sequence related to type II CBDs, and sequence with substantial sequence similarity to *Rhizobium* NodB proteins [53]. Finally, CenA, CenB, CenC, CbhA and CbhB hydrolyse glucosidic bonds with inversion of anomeric configuration, whereas CenD, XynD and Cex are retaining enzymes. Overall, the cellulase system of *C. fimi* is complex on several levels: the number of enzymes, which may be more than the eight characterized to date, the different modular arrangements, the different specificities, the widely different specific activities, and the different families of catalytic domains represented.

The CM-cellulose–Congo Red assay shows that Cex, CbhA and CbhB have detectable endoglucanase activity, although it is very much less than the activity of CenA. Therefore, Cex, CbhA and CbhB are predominantly, but not exclusively, exohydrolytic. CBH I and CBH II from *T. reesei* were reported to have endoglucanase activity [55], but the enzymes were purified from culture supernatants which also contained endoglucanases. Contamination of CBH I or CBH II with < 0.2% of endoglucanase could account for the level of endoglucanase activity observed. The *C. fimi* enzymes were produced from cloned genes expressed in *E. coli*, which does not encode β -1,4-glycanases. Therefore, the endoglucanase activities of Cex, CbhA and CbhB could not result from contaminating endoglucanases. Their properties support the claim that CBH I and CBH II also have endoglucanase activity.

Exoglucanases are thought to have limited action on CM-cellulose because of the carboxymethyl substituents. Exoglucanases can remove the few unsubstituted cellobiosyl units from the ends of CM-cellulose chains, but soon encounter a glucosyl unit with a substituent, which then blocks further progress of the enzyme along the chain. This means that the effective substrate concentration in a CM-cellulose solution at the start of the reaction is much lower for an exoglucanase than for an endoglucanase and quickly approaches zero.

Indirect support for this hypothesis comes from the structures of two cellulases which belong to family 6 of related amino acid sequences: endoglucanase E2 from *Thermomonospora fusca* [56], and exo-cellobiohydrolase II (CBH II) from *T. reesei* [57]. The three-dimensional structures of the two enzymes are virtually superimposable, but there is a striking difference. In E2, the active site is a cleft on the surface of the molecule; in CBH II it is in the corresponding location, but it is tunnel-shaped because two surface loops, which are not present in E2, fold over the active-site cleft. CenA and CbhA from *C. fimi* are also in family 6; CbhA, but not CenA, has stretches of amino acids corresponding to the surface loops of CBH II [10]. The active site of exo-cellobiohydrolase I (CBH I) from *T. reesei*, which belongs to family 7 of β -1,4-glucanases, is also tunnel-shaped [58].

It cannot be assumed, however, that all exoglucanases have tunnel-shaped active sites into which cellulose molecules thread from one end. The active site of Cex is an open cleft [59], as is that of another xylanase in family 10 [60]. The active site of a xylanase from *Bacillus circulans* is partially occluded by a loop

which moves aside in the binding of substrate (R. L. Campbell and D. Rose, personal communication). The endoglucanase activity exhibited by CbhA, CBH I and CBH II would require that the loops occluding their active sites have enough flexibility to allow a cellulose chain occasionally to enter the active site through the open roof rather than from one end of the tunnel.

We thank Carrie Hirsch, Emily Kwan, Alasdair MacLeod, Ned Mantei and Peter Tomme for advice and assistance. This work was supported by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

- Bronnenmeier, K., Rücknagel, K. P. and Staudenbauer, W. L. (1991) *Eur. J. Biochem.* **200**, 379–385.
- Creuzet, N., Berenger, J.-F. and Frixon, C. (1983) *FEMS Microbiol. Lett.* **20**, 347–350.
- Gardner, R. M., Doerner, K. C. and White, B. A. (1987) *J. Bacteriol.* **169**, 4581–4588.
- Hu, P., Kahrs, S. K., Chase, T. and Eveleigh, D. E. (1992) *J. Ind. Microbiol.* **10**, 103–110.
- MacKenzie, C. R., Bilous, D. and Johnson, K. G. (1984) *Can. J. Microbiol.* **30**, 1171–1178.
- Morag, E., Halevy, I., Bayer, E. A. and Lamed, R. (1991) *J. Bacteriol.* **173**, 4155–4162.
- Wang, W. K., Kruus, K. and Wu, J. H. D. (1993) *J. Bacteriol.* **175**, 1293–1302.
- Zhang, S., Lao, G. and Wilson, D. B. (1995) *Biochemistry* **34**, 3386–3395.
- Meinke, A., Gilkes, N. R., Kilburn, D. G., Miller, R. C., Jr. and Warren, R. A. J. (1993) *J. Bacteriol.* **175**, 1910–1918.
- Meinke, A., Gilkes, N. R., Kwan, E., Kilburn, D. G., Warren, R. A. J. and Miller, R. C., Jr. (1994) *Mol. Microbiol.* **12**, 413–422.
- Shen, H., Tomme, P., Meinke, A., Gilkes, N. R., Kilburn, D. G., Warren, R. A. J. and Miller, R. C. J. (1994) *Biochem. Biophys. Res. Commun.* **199**, 1223–1228.
- Coutinho, J. B., Gilkes, N. R., Warren, R. A. J., Kilburn, D. G. and Miller, R. C., Jr. (1992) *Mol. Microbiol.* **6**, 1243–1252.
- Gilkes, N. R., Jervis, E., Henrissat, B., Tekant, B., Miller, R. C., Jr., Warren, R. A. J. and Kilburn, D. G. (1992) *J. Biol. Chem.* **267**, 6743–6749.
- Wakarchuk, W. W., Kilburn, D. G., Miller, R. C., Jr. and Warren, R. A. J. (1986) *Mol. Gen. Genet.* **205**, 146–152.
- Gilkes, N. R., Warren, R. A. J., Miller, R. C., Jr. and Kilburn, D. G. (1988) *J. Biol. Chem.* **263**, 10401–10407.
- MacLeod, A. M., Lindhorst, T., Withers, S. G. and Warren, R. A. J. (1994) *Biochemistry* **33**, 6371–6376.
- Kunkel, T. A., Roberts, J. D. and Zakour, J. (1983) *Methods Enzymol.* **154**, 367–382.
- Mead, D. A., Szczesna-Skorupa, E. and Kemper, B. (1986) *Protein Eng.* **1**, 67–74.
- Graham, R. W., Atkinson, T., Kilburn, D. G., Miller, R. C., Jr. and Warren, R. A. J. (1994) *Nucleic Acids Res.* **21**, 4923–4928.
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*, Cold Spring Harbour Laboratory Press, New York.
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467.
- Biely, P., Mislovicová, D. and Toman, R. (1988) *Methods Enzymol.* **160**, 536–541.
- Miller, G. L., Blum, R., Glennon, W. E. and Burton, A. L. (1960) *Anal. Biochem.* **1**, 127–132.
- Teather, R. M. and Wood, P. J. (1982) *Appl. Environ. Microbiol.* **43**, 777–780.
- Scopes, R. K. (1974) *Anal. Biochem.* **59**, 277–282.
- Cantor, C. R. and Schimmel, P. R. (1980) *Biophysical Chemistry*, vol. 2, W. H. Freeman, San Francisco.
- von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4683–4690.
- Tomme, P., Warren, R. A. J. and Gilkes, N. R. (1995) *Adv. Microbial Physiol.* **37**, 1–81.
- Meinke, A., Gilkes, N. R., Kilburn, D. G., Miller, R. C., Jr. and Warren, R. A. J. (1991) *J. Bacteriol.* **173**, 7126–7135.
- Morag, E., Bayer, E. A., Hazlewood, G. P., Gilbert, H. J. and Lamed, R. (1993) *Appl. Biochem. Biotechnol.* **43**, 147–151.
- Belaich, J.-P., Gaudin, C., Belaich, A., Bagnara-Tardif, C., Fierobe, H.-P. and Reverbel, C. (1993) in *Genetics, Biochemistry and Ecology of Lignocellulose Degradation* (Shimada, K., Ohmiya, K., Kobayashi, Y., Hoshino, S., Sakka, K. and Karita, S., eds.), pp. 53–62, Uni Publishers, Tokyo.
- Doi, R. H., Goldstein, M. A., Park, J.-S., Lui, C.-C., Matano, Y., Takagi, M., Hashida, S., Foong, F. C.-F., Hamamoto, T., Segel, I. and Shoseyov, O. (1993) in *Genetics, Biochemistry and Ecology of Lignocellulose Degradation* (Shimada, K., Ohmiya, K., Kobayashi, Y., Hoshino, S., Sakka, K. and Karita, S., eds.), pp. 43–52, Uni Publishers, Tokyo.
- Bergquist, P. L., Gibbs, M. D., Saul, D. J., T'eo, V. S. J., Dwivedi, P. P. and Morris, D. (1993) in *Genetics, Biochemistry and Ecology of Lignocellulose Degradation* (Shimada, K., Ohmiya, K., Kobayashi, Y., Hoshino, S., Sakka, K. and Karita, S., eds.), pp. 276–285, Uni Publishers, Tokyo.
- Fujino, T., Karita, S. and Ohmiya, K. (1993) *J. Ferment. Bioeng.* **76**, 243–250.
- Henrissat, B. (1991) *Biochem. J.* **280**, 309–316.
- Henrissat, B. and Bairoch, A. (1993) *Biochem. J.* **293**, 781–788.
- Ong, E., Gilkes, N. R., Miller, R. C., Jr., Warren, R. A. J. and Kilburn, D. G. (1993) *Biotechnol. Bioeng.* **42**, 401–409.
- Wong, W. K. R., Gerhard, B., Guo, Z. M., Kilburn, D. G., Warren, R. A. J. and Miller, R. C., Jr. (1986) *Gene* **44**, 315–324.
- Meinke, A., Gilkes, N. R., Kilburn, D. G., Warren, R. A. J. and Miller, R. C., Jr. (1993) in *Genetics, Biochemistry and Ecology of Lignocellulose Degradation* (Shimada, K., Ohmiya, K., Kobayashi, Y., Hoshino, S., Sakka, K. and Karita, S., eds.), pp. 286–297, Uni Publishers, Tokyo.
- Gilkes, N. R., Langsford, M. L., Kilburn, D. G., Miller, R. C., Jr. and Warren, R. A. J. (1984) *J. Biol. Chem.* **259**, 10455–10459.
- Wood, P. J. (1980) *Carbohydr. Res.* **85**, 271–287.
- Gilkes, N. R., Claessens, M., Aebersold, R., Henrissat, B., Meinke, A., Morrison, H. D., Kilburn, D. G., Warren, R. A. J. and Miller, R. C., Jr. (1991) *Eur. J. Biochem.* **202**, 367–377.
- Day, A. G. and Withers, S. W. (1986) *Biochem. Cell. Biol.* **64**, 914–922.
- Meinke, A., Damude, H. G., Tomme, P., Kwan, E., Kilburn, D. G., Miller, R. C., Jr., Warren, R. A. J. and Gilkes, N. R. (1995) *J. Biol. Chem.* **270**, 4383–4386.
- Moser, B., Gilkes, N. R., Kilburn, D. G., Warren, R. A. J. and Miller, R. C., Jr. (1989) *Appl. Environ. Microbiol.* **55**, 2480–2487.
- Mayer, F., Coughlan, M. P., Mori, Y. and Ljungdahl, L. G. (1987) *Appl. Environ. Microbiol.* **53**, 2785–2792.
- Salohelmo, A., Hoffer, A.-M., Teleman, O., Henrissat, B. and Penttilä, M. (1994) *Mol. Microbiol.* **13**, 219–228.
- Ward, M., Wu, S., Dauberman, J., Weiss, G., Larenas, E., Bower, B., Rey, M., Clarkson, K. and Bott, R. (1993) in *Trichoderma reesei Cellulases and Other Hydrolases* (Suominen, P. and Reinikainen, T., eds.), pp. 153–158, Foundation for Biotechnical and Industrial Fermentation Research, Helsinki.
- Sims, P. F. G., Soares-Felipe, M. S., Wang, Q., Gent, M. E., Tempelaars, C. and Broda, P. (1994) *Mol. Microbiol.* **12**, 209–216.
- Kruus, K., Wang, W. K., Ching, J. and Wu, J. H. D. (1995) *J. Bacteriol.* **177**, 1641–1644.
- Despande, M., Erickson, K.-E. and Pettersson, L. G. (1984) *Anal. Biochem.* **138**, 481–487.
- Coutinho, J. B., Moser, B., Kilburn, D. G., Warren, R. A. J. and Miller, R. C., Jr. (1991) *Mol. Microbiol.* **5**, 1221–1233.
- Millward-Sadler, S. J., Poole, D. M., Henrissat, B., Hazlewood, G. P., Clarke, J. H. and Gilbert, H. J. (1994) *Mol. Microbiol.* **11**, 375–382.
- O'Neill, G., Guo, S. H., Warren, R. A. J., Kilburn, D. G. and Miller, R. C., Jr. (1986) *Gene* **44**, 325–330.
- Ståhlberg, J., Johansson, G. and Pettersson, G. (1993) *Biochim. Biophys. Acta* **1157**, 107–113.
- Spezio, M., Wilson, D. B. and Karplus, P. A. (1993) *Biochemistry* **32**, 9906–9916.
- Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C. and Jones, T. A. (1990) *Science* **249**, 380–386.
- Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K. C., Teeri, T. and Jones, T. A. (1994) *Science* **265**, 524–528.
- White, A., Withers, S. G., Gilkes, N. R. and Rose, D. R. (1994) *Biochemistry* **33**, 12546–12552.
- Derewenda, U., Swenson, L., Green, R., Wei, Y., Morosoli, R., Shareck, F., Kluepfel, D. and Derewenda, Z. S. (1994) *J. Biol. Chem.* **269**, 20811–20814.