APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Dec. 1997, p. 4721–4728 Vol. 63, No. 12
0099-2240/97/$04.00+0

Two Cellulases, CelA and CelC, from the Polycentric Anaerobic Fungus *Orpinomyces* Strain PC-2 Contain N-Terminal Docking Domains for a Cellulase-Hemicellulase Complex

XIN-LIANG LI, HUIZHONG CHEN, AND LARS G. LJUNGDAHL*

Department of Biochemistry and Molecular Biology and Center for Biological Resource Recovery, The University of Georgia, Athens, Georgia 30602-7229

Received 8 September 1997/Accepted 22 September 1997

Two cDNAs encoding two cellulases, CelA and CelC, were isolated from a cDNA library of the polycentric anaerobic fungus *Orpinomyces* sp. strain PC-2 constructed in *Escherichia coli*. Nucleotide sequencing revealed that the *celA* cDNA (1,558 bp) and *celC* cDNA (1,628 bp) had open reading frames encoding polypeptides of 459 (CelA) and 449 (CelC) amino acids, respectively. The two cDNAs were 76.9 and 67.7% identical at the nucleotide and amino acid levels, respectively. Analysis of the deduced amino acid sequences showed that starting from the N termini, both CelA and CelC had signal peptides, which were followed by noncatalytic repeated peptide domains (NCRPD) containing two repeated sequences of 33 to 40 amino acid residues functioning as docking domains. The NCRPDs and the catalytic domains were separated by linker sequences. The NCRPDs were homologous to those found in several hydrolases of anaerobic fungi, whereas the catalytic domains were homologous to the catalytic domains of fungal cellobiohydrolases and bacterial endoglucanases. The linker sequence of CelA contained predominantly glutamine and proline residues, while that of CelC contained mainly threonine residues. CelA and CelC did not have a typical cellulose binding domain (CBD). CelA and CelC expressed in *E. coli* rapidly decreased the viscosity of carboxymethyl cellulose (CMC), indicating that there was endoglucanase activity. In addition, they produced cellobiose from CMC, acid-swollen cellulose, and cellotetraose, suggesting that they had cellobiohydrolase activity. The optimal activity conditions with CMC as the substrate were pH 4.3 to 6.8 and 50°C for CelA and pH 4.6 to 7.0 and 40°C for CelC. Despite the lack of a CBD, CelC displayed a high affinity for microcrystalline cellulose, whereas CelA did not.

Microorganisms have evolved diverse strategies for breaking down plant cell wall constituents. Aerobic organisms secrete a number of separate enzymes, which together hydrolyze cellulose. Examples of individual enzyme producers are the fungus *Trichoderma reesei* and the bacteria *Cellulomonas fimi* and *Thermomonospora fusca*. The cellulases of these organisms have catalytic domains joined by linker sequences to cellulose binding domains (CBDs) (21, 41). In contrast, anaerobic bacteria, such as *Clostridium thermocellum* (2, 18) and *Clostridium cellulovorans* (15), produce cellulases and hemicellulases which form high-molecular-weight complexes containing catalytic subunits and scaffolding polypeptides.

Anaerobic fungi discovered by Orpin (35) have highly efficient cellulases (9, 28, 31, 32, 45). On the basis of the morphology of sporangia, mycelia, and zoospores, anaerobic fungi have been classified into two groups, the monocentric and polycentric groups (8, 9). Monocentric fungi have only one sporangium, whereas polycentric isolates have multiple sporangia. Most investigations of anaerobic fungi have focused on monocentric isolates belonging to the genera *Neocallimastix* and *Piromyces* and the polycentric genus *Orpinomyces*. No native endoglucanase or cellobiohydrolase has been purified from anaerobic fungi, but a cellulase-hemicellulase complex with a mass of about 750 kDa has been isolated from *Neocallimastix frontalis* (42). Data from gene cloning and sequencing analyses of polysaccharidases from anaerobic fungi support the view that multiple cellulases and hemicellulases form high-molecular-weight complexes similar to the cellulosome of the clostridia (1, 17, 20, 31, 34, 47). Catalytically active polypeptides lack CBDs but contain noncatalytic repeated peptide domains (NCRPDs) that function as docking domains (17). However, a cellobiohydrolase (CELA) of *Neocallimastix patriciarum*, which lacks an NCRPD but contains a typical fungal CBD, was described recently (14).

In this paper we describe the cloning and sequencing of cDNAs encoding two similar cellulases (CelA and CelC) of *Orpinomyces* sp. strain PC-2 and the characterization of these cellulases obtained by expression in *Escherichia coli*. CelA and CelC have catalytic domains homologous to those of *N. patriciarum* CELA and other family 6 glycosyl hydrolases (24) but contain NCRPDs instead of CBDs. CelA and CelC have both endoglucanase and cellobiohydrolase activities.

MATERIALS AND METHODS

Microbial strains and vectors. The polycentric anaerobic fungus *Orpinomyces* sp. strain PC-2 isolated from a bovine rumen (9) was cultured as described by Barichievich and Calza (3). *E. coli* XL-1Blue and λZAPII pBluescript SK(−) were products of Stratagene Cloning Systems (La Jolla, Calif.).

Cloning and sequencing of cellulase cDNAs. Extraction of total RNA from *Orpinomyces* mycelia grown in liquid media containing 0.2% (wt/vol) Avicel, purification of mRNA, and construction of a cDNA library in λZAPII have been described previously (10). To isolate cellulase clones, λ plaques were formed after *E. coli* cells were infected with λ phages in standard NZY medium (Stratagene) containing 5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and either 0.2% remazol brilliant blue (RBB)–carboxymethyl cellulose (CMC) or lichenan. Positive plaques were recognized by the formation of clear haloes on a blue background in the case of RBB-CMC or by the formation of light yellow zones on a red background after staining with 1 mg of Congo red per ml and destaining with 1 M NaCl (5) in the case of lichenan. Positive pure λ plaques were obtained after a secondary screening with a lower density of plaques and then converted into pBluescript plasmids by in vivo excision (Stratagene). *E. coli* XL-1Blue cells harboring the pBluescript recombinants were grown overnight at 37°C in Luria-Bertani medium containing 100 μg of ampicillin per ml. Plasmids were purified

* Corresponding author. Mailing address: Department of Biochemistry and Molecular Biology, A214 Life Sciences Building, The University of Georgia, Athens, GA 30602-7229. Phone: (706) 542-7640. Fax: (706) 542-2222. E-mail: Ljungdah@bscr.uga.edu.

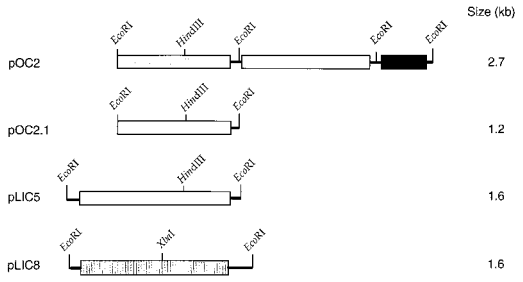

FIG. 1. Restriction maps of cellulase-positive cDNAs isolated by using RBB-CMC or lichenan as the substrate. The position at which the 5′ ends of pOC2 and pOC2.1 start in pLIC5 is shown in Fig. 2. pOC2, pOC2.1, and pLIC5 have ORFs encoding CelA, whereas the ORF in pLIC8 encodes CelC. The shaded boxes and horizontal lines represent ORFs and untranslated regions, respectively. The open and solid boxes of pOC2 represent ORFs encoding sequences homologous to yeast amino peptidase and H4 histone, respectively.

by using a spin column miniprep kit (Qiagen) or a Maxiprep kit (Promega, Madison, Wis.). Plasmids from different primary clones were subjected to restriction digestion with various enzymes, and DNA fragments were separated on agarose gels (37). DNA samples were further cleaned by rinsing them in Centricon-100 tubes (Amicon) before they were subjected to sequencing with an automatic DNA sequencer (Applied Biosystems). Both strands of the cDNA inserts were sequenced by walking from the ends of the inserts by using plasmid- and insert-specific primers. Sequence data analyses, data bank searches, and multiple sequence alignment were performed by using the Genetics Computer Group package (University of Wisconsin Biotechnology Center, Madison, Wis.) and the VAX/VMS system of the BioScience Computing Resource of The University of Georgia.

Enzyme preparation. Single colonies of *E. coli* harboring plasmid pOC2.1, pLIC5, or pLIC8 (Fig. 1) containing DNA encoding truncated CelA (ΔCelA), CelA, or CelC, grown on solid Luria-Bertani medium containing 100 μg of ampicillin per ml, were inoculated into flasks containing 50 ml of liquid medium plus ampicillin. The cultures were shaken (280 rpm) at 37°C and grown to an optical density at 600 nm of 0.5. IPTG (1 mM) was added, and the cultures were grown while they were being shaken for an additional 4 h. Cells were harvested by centrifugation (7,000 × *g*, 30 min), washed with 20 ml of buffer containing 50 mM sodium phosphate (pH 6.0) and 2 mM EDTA, and resuspended in 10 ml of the same buffer. The cells were then disrupted by sonication (Sonifier 450; Branson Ultrasonics Corp., Danbery, Conn.). The release of cytoplasmic and periplasmic proteins was monitored by following the increase in protein concentration. Cell debris was removed by centrifugation (15,000 × *g*, 30 min).

Enzyme assays and analytical methods. Unless otherwise stated, the activities of enzymes with various substrates were determined at 39°C in 0.4 ml of 50 mM sodium phosphate buffer (SPB) (pH 6.0) containing 0.7% (wt/vol) soluble or insoluble substrate. Phosphoric-acid-swollen cellulose (ASC) was prepared as described by Wood (44). Reducing sugars were quantified by the dinitrosalicylic acid (DNS) procedure (33). Before the absorption values were measured with a spectrophotometer (Hewlett-Packard), residual insoluble substrates were removed by centrifugation at 8,000 × *g* for 20 min. Glucose was used as a standard for assays of reducing sugars. The stabilities of enzyme preparations at various pH values and temperatures were determined by assaying samples preincubated under specified conditions by using the standard procedures and CMC as the substrate. The buffers used to study the pH range were 0.1 M sodium acetate (pH 2.8 to 5.4), sodium phosphate (pH 5.8 to 7.8), and sodium borate (pH 8.2 to 9.6).

The activities of enzyme preparations with *p*-nitrophenol (PNP)-linked substrates were determined in 0.2 ml of SPB containing 2 mM substrate at 39°C. Reactions were terminated after 15 min by the addition of 1 ml of 1 M Na_2CO_3. The release of PNP was measured spectrophotometrically at 405 nm. PNP was used as the standard, and 1 U of activity was defined as the amount of enzyme required to release 1 μmol of glucose equivalent or PNP per min.

The hydrolysis of CMC was determined by measuring the decrease in viscosity in addition to using the DNS method for assaying reducing sugars. A solution of 0.5% (wt/vol) high-viscosity CMC (Sigma) in 5 ml of SPB was added to a Kimax viscometer (10 ml) placed in a 40°C water bath. After 5 min of temperature equilibration, the viscosity was measured before and then at different times after the addition of 100 μl (1 mg) of *E. coli* cell lysate proteins. The viscosity of buffer without CMC was measured and used for calculating relative viscosity. Protein was determined with the MicroBCA reagent (Pierce) by using bovine serum albumin as the standard.

The products formed by hydrolysis of cellodextrins (G2 to G5), CMC, and ASC were analyzed by thin-layer chromatography (TLC). Solutions containing 200 μg of *E. coli* cell lysate proteins and 1 mM cellodextrins, 0.2% (wt/vol) CMC, or ASC in 50 mM SPB (pH 6.0) were incubated at 39°C for 5 h. The reactions were terminated by boiling the preparations for 5 min. Hydrolysis products were separated by TLC on silica gel plates (Analtech, Inc., Neward, Del.) by using a mixture of chloroform, glacial acetic acid, and water (6:7:1, vol/vol/vol) as the solvent (26). A mixture of glucose, cellobiose, cellotriose, and cellotetraose (Sigma) was used as a standard for the identification of hydrolysis products. After partition, the plates were sprayed with a reagent containing aniline (2 ml), diphenylamine (2 g), acetone (100 ml), and 85% H_3PO_4 (15 ml), and then the sugars were visualized after the plates were heated at 105°C in an oven for 15 min (22).

Nucleotide sequence accession numbers. The nucleotide sequences of CelA and CelC of *Orpinomyces* sp. strain PC-2 have been assigned accession no. U63837 and U63838, respectively, in the GenBank database.

RESULTS

Cloning, expression, and sequencing of *celA* and *celC* of *Orpinomyces* sp. strain PC-2. In a previous publication (31), we described the screening of a cDNA library constructed in λZAPII by using mRNA extracted from *Orpinomyces* sp. strain PC-2 (10) for plaques hydrolyzing RBB-CMC. Two different clones with insert sizes of 2.7 kb (pOC2, *celA*) (Fig. 1) and 1.8 kb (pOC1, *celB*) were obtained. As reported previously (31), the insert cDNA (*celB*) of pOC1 contained 1,825 bp and a complete open reading frame (ORF) which encoded a polypeptide (CelB) of 471 amino acids. Sequencing of the insert cDNA of pOC2, as reported here, revealed that this plasmid possessed cDNAs of three unrelated genes, resulting from ligation at the *Eco*RI sites of the adapters of unrelated cDNA sequences (Fig. 1). A 1.2-kb sequence in the 5′ region consisted of an incomplete ORF (Fig. 2) encoding a polypeptide homologous to fungal and bacterial cellulases. This sequence was followed by two sequences coding for polypeptides with homology to a yeast amino peptidase and an H4 histone protein (data not shown). The incomplete ORF encoding the cellulase in pOC2 was fused in frame to the *lacZ* gene. Thus, the cellulase was synthesized as a fusion protein. Subcloning of the 1.2-kb fragment into pBluescript in the same orientation yielded pOC2.1, which had the same level of activity on CMC as pOC2 did (Fig. 1).

The λZAPII library was also screened for plaques that hydrolyzed lichenan, a glucan with alternating linkages of β-1,3 and β-1,4 bonds. Twenty positive plaques were isolated from 2.5×10^6 PFU. Restriction analyses revealed that these plaques represented cDNAs of four distinct genes (data not shown). Sequencing of the inserts in the plasmids, converted from plaques by in vivo excision, revealed that pLIC5 (Fig. 1) contained 1,558 bp (*celA*) with a complete ORF encoding a polypeptide (CelA) of 459 amino acids with a mass of 50,580 Da (Fig. 2). The difference between pOC2.1 and pLIC5 was that pLIC5 contained a 5′ noncoding region and a region encoding 115 N-terminal amino acid residues that were missing in pOC2.1 (Fig. 2). The sequences of these two clones encoding the carboxy-terminal 345 amino acids and 3′ noncoding ends were identical (Fig. 2). These results suggest that the 115 amino acids in the amino region of CelA were not required for catalysis. A second lichenan-hydrolyzing plaque, pLIC8 (Fig. 1), had an insert of 1,628 bp (*celC*) (Fig. 3) with a complete ORF coding for a polypeptide (CelC) of 449 amino acids with a mass of 49,389 Da.

The assignment of translation start codons for *celA* and *celC* was based on the following findings: (i) both ORFs had stop codons preceding the ATG codon; (ii) the next codon was a codon for a Lys residue, which was followed by a peptide sequence rich in hydrophobic amino acids which are typical of secretion signal peptides for extracellular enzymes (29); and (iii) very high-A+T-content regions preceded the putative ATG codons, as found previously for cDNAs encoding a cyclophilin (10), an enolase (16), and other hydrolases (17, 31) of anaerobic fungi. In genes isolated from *Orpinomyces* sp. strain

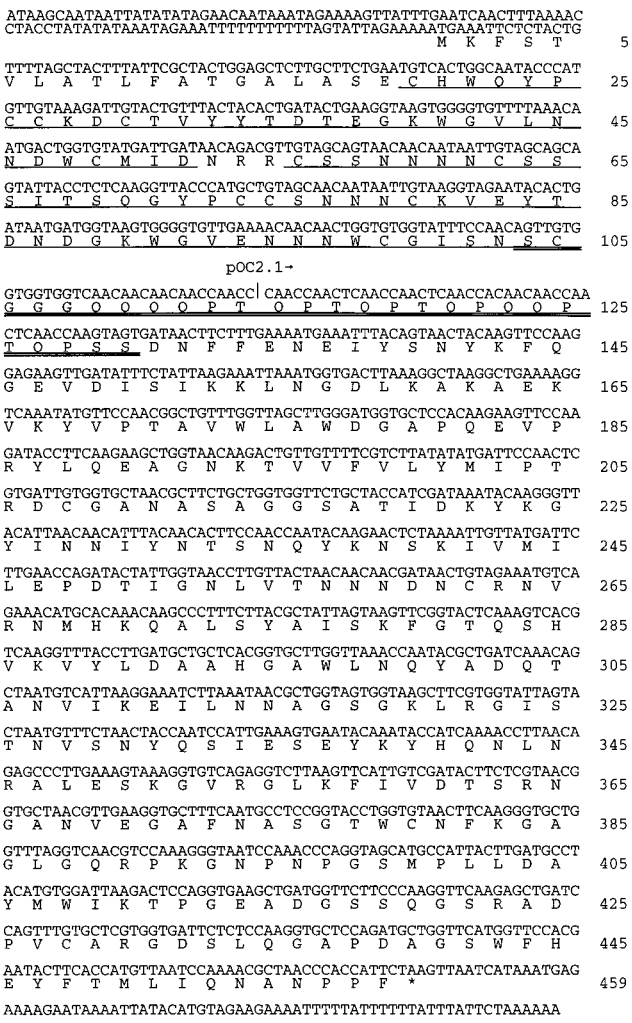

FIG. 2. Nucleotide sequence and deduced amino acid sequence of *Orpinomyces celA* cDNA (pLIC5). The repeated peptides of the NCRPD and the linker sequence are underlined and double underlined, respectively; the 5′ end of pOC2.1 located in the linker region is indicated. The asterisk indicates the stop codon.

PC-2 so far, including *celA* and *celC*, the wobble position is severely biased toward A or T, and G is rarely used. Codons such as GGG, GCG, AGG, TCG, CGG, CGA, CAG, CTG, and CCG and translation stop codons containing G (TGA and TAG) are not used. High-A+T-content genes and extremely A+T-rich noncoding regions are prominent in anaerobic fungi (14, 16, 17, 31, 47), indicating that monocentric or polycentric anaerobic fungi have similar nucleotide compositions.

Domains of CelA and CelC. The nucleotide and deduced amino acid sequences of *celA* and *celC* were compared with each other, the CelB sequence of *Orpinmyces* sp. strain PC-2 (31), and homologous sequences obtained from the SwissProt and GP data banks. The amino acid sequences of CelA and CelC are 67.6% identical, and the level of identity at the nucleotide level is even higher (76.9%). Notably, there are regions of one, three, and five amino acids present in CelA which are not found in CelC. CelA and CelC have no significant level of identity with CelB except that the N-terminal regions (amino acids 20 to 100) of mature CelA and CelC are highly homologous to the C-terminal region of CelB (amino acids 390 to 470). This region (Fig. 4) is also present in XynA of *Orpinomyces* sp. strain PC-2 and in several polysaccharide hydrolases of monocentric anaerobic fungal species belonging to the genera *Neocallimastix* and *Piromyces* (17, 20, 47). It has been identified as an NCRPD. It has been postulated that NCRPDs function like the dockerin domain of catalytic subunits of clostridial cellulosomes (17, 31). The occurrence of this domain at N termini is unusual. This domain has always been found attached by a linker region at the carboxy ends of catalytic domains or between two catalytic domains.

The catalytic domains of *Orpinomyces* CelA and CelC involve the C-terminal regions starting with amino acid residues 128 and 127, respectively. These domains are highly homologous with the catalytic domain of *N. patriciarum* CELA (14), with which they exhibit 71.9 and 70.3% identity, respectively, at the nucleotide level and 65.0 and 60.5% identity, respectively, at the amino acid level. In addition, the catalytic domains of CelA and CelC display significant levels of homology with fungal cellobiohydrolases and bacterial endoglucanases (Fig. 5) belonging to family B glycanases (6, 19, 21, 23) or family 6 glycosyl hydrolases (24). Thus, CelA and CelC should be

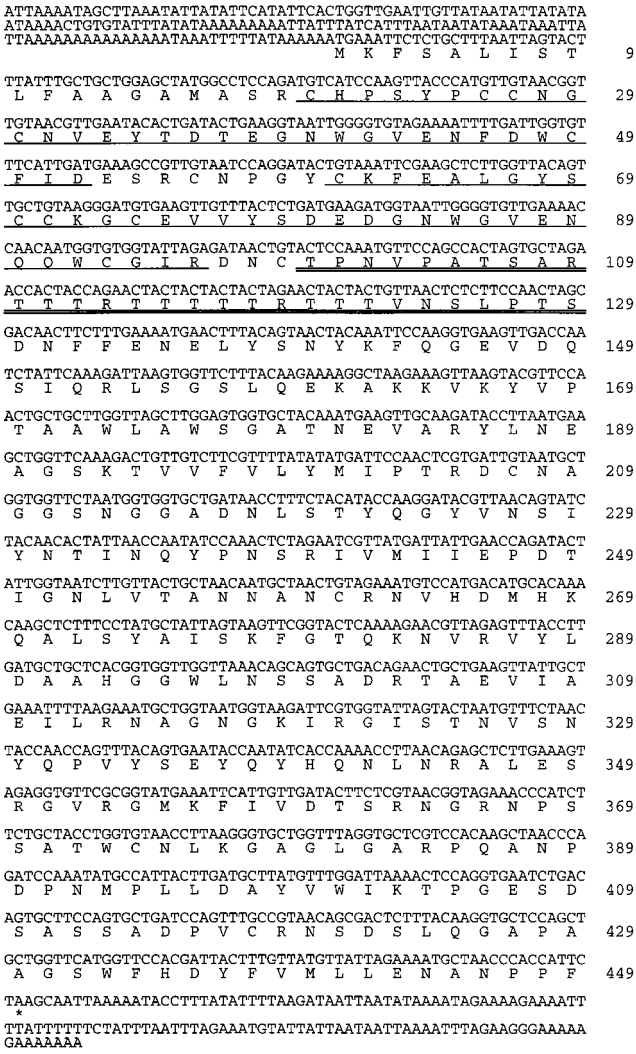

FIG. 3. Nucleotide sequence and deduced amino acid sequence of *Orpinomyces celC* cDNA (pLIC8). The repeated peptides of the NCRPD and the linker sequence are underlined and double underlined, respectively. The asterisk indicates the stop codon.

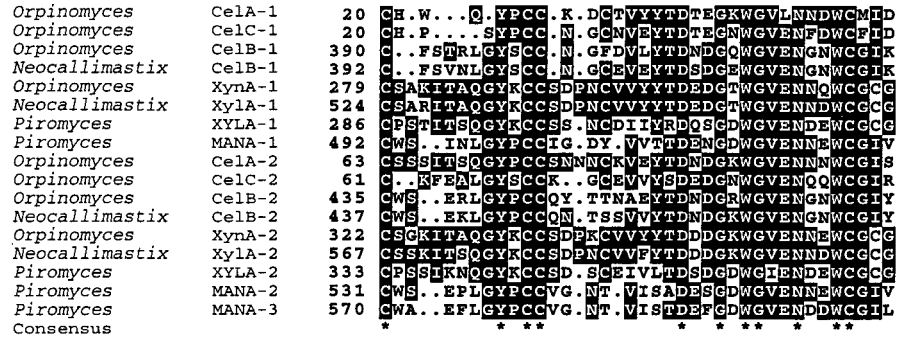

FIG. 4. Alignment of the amino acids of the repeated peptides (peptides 1, 2, and 3) of the NCRPDs of polysaccharide hydrolases of anaerobic fungi. The enzymes included are CelA and CelC (this study) and CelB and XynA (30) of *Orpinomyces* sp. CelB (46) and XylA (20) of *N. patriciarum*, and XYLA and MANA of *Piromyces* sp. (17). The positions of the first amino acid residues of the repeated peptides in the enzyme precursors are shown. Residues conserved in the repeated peptides are indicated with asterisks.

placed in family 6. However, they do not have a CBD, which is present in all other cellulases of this family.

The linker sequence (amino acid residues 103 to 130) of CelA is comprised mainly of Gln and Pro residues (Fig. 2). Thr and Ser residues are also present. In contrast, the CelC linker region (amino acid residues 100 to 129) contains predominantly Thr (Fig. 3). Thr is also the most abundant residue in the linkers of *Orpinomyces* CelB and XynA. The linker sequence between the CBD and the catalytic domain of *N. patriciarum* CELA is long and consists almost entirely of Asn residues (14). A possible noncatalytic 455-residue linker sequence has been identified in XYLB of *N. patriciarum*; it has repeats of an octapeptide containing Ser, Thr, and Pro residues (7).

Enzymatic properties. Cell extracts of *E. coli* expressing *Orpinomyces* CelA, CelA with the NCRPD truncated (ΔCelA), and CelC were prepared, and the activities of these extracts with various substrates were determined (Table 1). The two cellulases rapidly hydrolyzed CMC, ASC, lichenan, and barley β-glucan. Low-level but detectable hydrolysis of Avicel by CelA, ΔCelA, and CelC was observed. In addition, CelC hydrolyzed other polymeric substrates containing β-1,4-, β-1,3, and β-1,6 glucoside bonds. CelA and ΔCelA had almost identical substrate specificities, demonstrating that the NCRPD is not involved in catalysis or substrate binding.

The end products of hydrolysis of CMC, ASC, and cellodextrins by the two cellulases were determined by TLC (Fig. 6). Hydrolysis of CMC and ASC by the two enzymes yielded

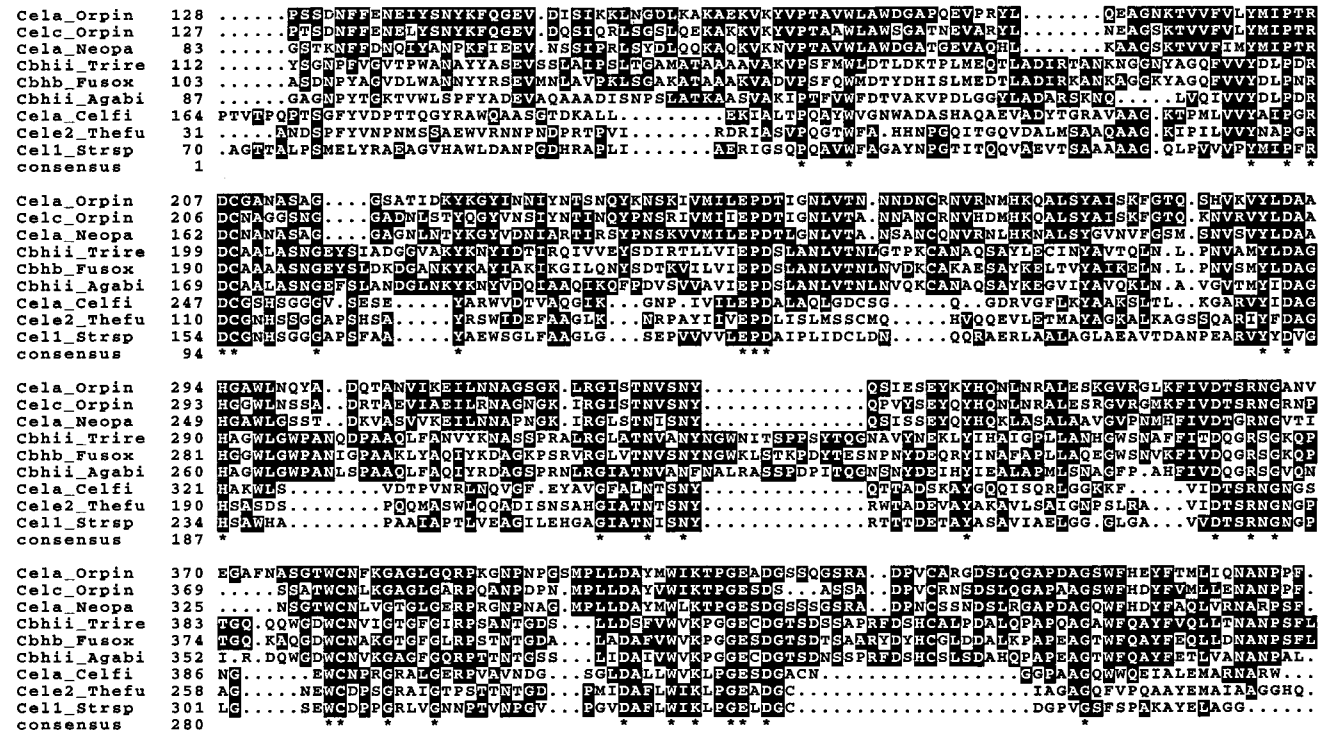

FIG. 5. Alignment of the amino acid sequences of the catalytic domains of *Orpinomyces* CelA (Celc-Orpin) and CelC (Celc-Orpin) with the amino acid sequences of other family 6 cellulases, including CELA of *N. patriciarum* (Cela-Neopa) (14), CBHIIs of *Trichoderma reesei* (Cbhii-Trire) (40), *Fusarium oxysporum* (Cbhb-Fusox) (38), and *Agaricus bisporus* (Cbhii-Agabi) (12), *Cellulomonas fimi* CenA (Cela-Celfi) (43), *Thermomonospora fusca* E2 (Cele2-Thefu) (27), and *Streptomyces* Ksm-9 (Cell-Strsp) (13).

TABLE 1. Substrate specificities of *Orpinomyces* cellulases produced in *E. coli*

Substrate[a]	% of activity with CMC		
	CelA	ΔCelA	CelC
CMC[b]	100	100	100
Avicel	5.6	6.6	10.3
ASC	54.4	63.2	63.7
Laminarin	ND[c]	ND	ND
Lichenan	139	142	171
Barley β-glucan	696	710	812
Arabinogalactan	ND	ND	10.7
Araban	ND	ND	28.4
Galactan	ND	ND	16.7
Pullulan	11.0	8.2	20.3
Gum arabic	ND	ND	17.6
Pachyman	ND	ND	21.1
Pustulan	ND	ND	17.2

[a] The rate of hydrolysis of mannan, starch, oat spelt xylan (0.7%, wt/vol), *p*-nitrophenyl-β-D-xylopyranoside, or *p*-nitrophenyl-β-D-cellobiose (1 mM) was less than 1.0% of the rate of hydrolysis of CMC.

[b] The activities of extracts containing CelA, ΔCelA, and CelC with CMC were 1.2, 1.3, and 1.1 U/ml, respectively.

[c] ND, not detected (hydrolysis rate was less than 1.0% of the hydrolysis rate for CMC).

cellobiose and cellotriose. Oligosaccharides larger than cellotriose were also formed during the hydrolysis of CMC by CelA and CelC but not when ASC was the substrate. Glucose was not formed with these two polymeric substrates by CelA or CelC.

Neither of the enzymes hydrolyzed cellobiose. Different product profiles for the two enzymes were obtained when cellotriose and cellopentaose were the substrates. CelA hydrolyzed cellotriose to cellobiose and glucose, but CelC was not able to cleave this substrate. Cellotetraose was cleaved predominantly to cellobiose by CelA or CelC; in addition, trace amounts of glucose and cellotriose were formed in the case of CelA. A trace amount of cellotriose and possibly some higher oligosaccharides were observed during hydrolysis of cellotetraose by CelC, which suggests that CelC may have transglycosylation activity. CelA hydrolyzed cellopentaose to cellotriose, cellobiose, and glucose, while CelC produced only cellotriose and cellobiose and no glucose, which confirmed the inability of CelC to hydrolyze cellotriose.

The change in viscosity and the accumulation of reducing

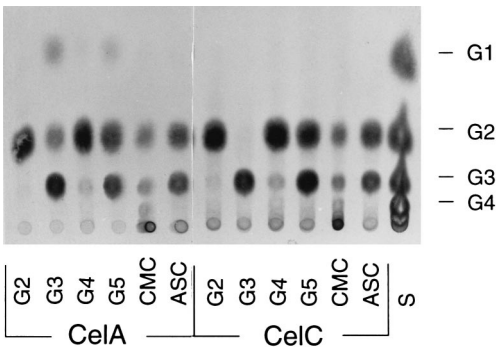

FIG. 6. TLC analysis of products of CMC, ASC, and cellodextrins hydrolyzed by CelA and CelC. The procedures used for enzyme and substrate preparation, hydrolysis, TLC, and visualization are described in Materials and Methods. Glucose (G1) and cellodextrins (Sigma Chemical Co.), including cellobiose (G2), cellotriose (G3), cellotetraose (G4), and cellopentaose (G5), were used as standards (lane S) at equal molarities and as substrates.

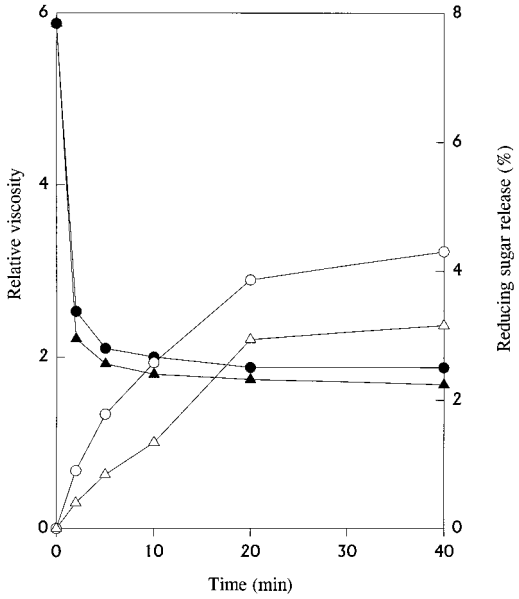

FIG. 7. Viscosity reduction and reducing sugar production during the hydrolysis of high-viscosity CMC (Sigma) by CelA (● and ○) and CelC (▲ and △). Relative viscosity = time for reaction sample/time for buffer. The reducing sugar production is expressed as the percentage of reducing ends generated compared with the total theoretical number of ends.

sugars during hydrolysis of CMC by the two enzymes were determined (Fig. 7). The viscosity of CMC was reduced rapidly by the enzymes during the first 5 min of hydrolysis. After this, the reduction in viscosity was small. This behavior is typical of endoglucanase activity. The levels of reducing sugars increased most rapidly during the first 20 min of the reaction. However, an increase, although at a low rate, was observed for 40 min. The formation of reducing sugars may be due to low endoglucanase activity, but the production of cellobiose and cellotriose as observed in Fig. 6 suggests that exoglucanase activity is present.

The adsorption of the expressed cellulase activities in a lysate of *E. coli* cells was investigated by using Avicel. More than 90% of the CelA and ΔCelA activities were recovered after the Avicel treatment, indicating that these preparations did not possess enzymes with strong cellulose binding affinity. Less than 50% of the CelC activity was recovered after the Avicel adsorption treatment, and this adsorption was not hindered by bovine serum albumin at a concentration four times that of the *E. coli* proteins. This result indicates that CelC has affinity for Avicel.

The activities of the two enzymes with CMC at various pH values and temperatures are shown in Fig. 8 and 9, respectively. The highest activities were at pH 4.8 for CelA and at pH 5.6 to 6.2 for CelC. Enzyme activities greater than 50% of the maximum activity were attained at pH 4.3 to 6.8 for CelA and at pH 4.6 to 7.0 for CelC. After preincubation at 40°C at pH 3.5 to 9.6 for 1 h, both enzymes retained 80% or more of the maximal activity.

CelA and CelC displayed their highest activities at 50 and 40°C, respectively. The activities rapidly diminished at 60°C, a temperature at which CelA and CelC were irreversibly inactivated during preincubation. The enzymes retained more than 90% of their activities after preincubation at 45°C for 24 h in the absence of substrate. CelA and CelC retained 92 and 83%, respectively, of their activities after 5 h of preincubation at

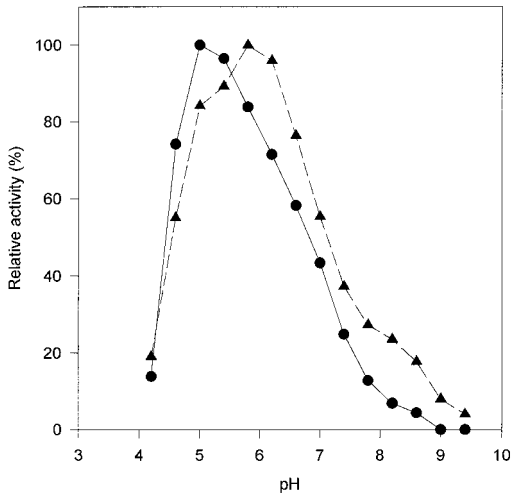

FIG. 8. Effect of pH on the activities of CelA (●) and CelC (▲) during the hydrolysis of CMC. Assays were performed at 39°C for 15 min in 0.4 ml of 50 mM SPB (pH 6.0) containing 0.7% (wt/vol) CMC (medium viscosity) and enzyme preparations (20 to 100 μg of *E. coli* cell extract). Reactions were terminated by adding DNS reagent and boiling for 5 min.

50°C. The pH and temperature profiles indicate that the *Orpinomyces* enzymes are active under the physiological conditions found in the rumen (pH 6 to 7 and 38 to 42°C) (46).

DISCUSSION

The cellulases encoded by *celA* and *celC* of the polycentric anaerobic fungus *Orpinomyces* sp. strain PC-2 exhibit structural similarities with each other, XynA and CelB from the same fungus, and hydrolytic enzymes from other anaerobic fungi. Like the sequences of these other enzymes, their sequences can be dissected into several regions. Starting with the N-terminal end, they contain a basic residue at the second position, which is followed by a peptide rich in hydrophobic residues. This region is present in extracellular proteins and functions as a transmembrane signal. In both enzymes, it is

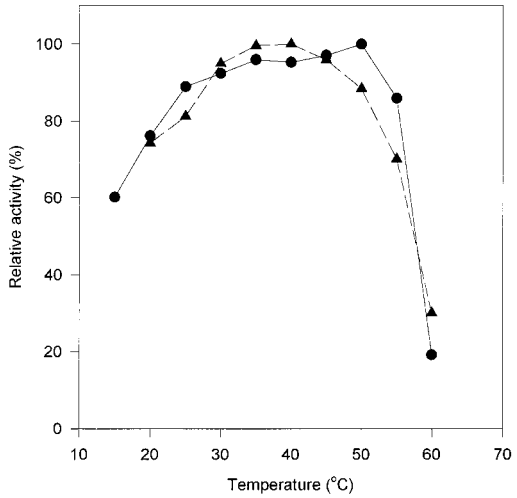

FIG. 9. Effect of temperature on the activities of CelA (●) and CelC (▲) during the hydrolysis of CMC. Assays were performed as described in the legend to Fig. 8 except that the incubation temperature was varied.

followed by an NCRPD which has been postulated to have a function similar to that of the dockerin domain of catalytic subunits of the cellulosome of *Clostridium thermocellum* (4, 11, 19, 25). The NCRPD is followed by a linker region which connects the NCRPD to the catalytic domain.

The most striking similarity between the two cellulases discussed here is the presence of NCRPDs. These domains have been found in several hydrolases of anaerobic fungi (17, 20, 31, 47). Western blot analysis in which a polyclonal antibody against the first peptide sequence of the NCRPD of *Orpinomyces* XynA was used demonstrated that this sequence is present in numerous extracellular proteins of *Orpinomyces* sp. strain PC-2 and *N. frontalis* (31). Our present work and observations by other workers (17, 20) have shown that the NCRPD is not involved in catalysis or cellulose binding. Fanutti et al. (17) showed that the NCRPD of a *Piromyces* xylanase binds to other polypeptides of high-molecular-weight complexes of *N. patriciarum* and a *Piromyces* strain. Cellulase-hemicellulase complexes with masses of more than 2×10^6 Da have been found in *Orpinomyces* sp. strain PC-2 (30). These complexes contain more than 20 polypeptides having masses ranging from 35 to 130 kDa. It can be postulated that CelA, CelB, CelC, and XynA are components of the complexes of *Orpinomyces* sp. strain PC-2. The NCRPDs of CelB and XynA of *Orpinomyces* sp. (31), XYLA of *N. patriciarum* (20), and MANA of *Piromyces* sp. (17) are located at the C-terminal ends. In *Piromyces* XYLA the NCRPD is located between two catalytic domains (17). The presence of the NCRPD at the N termini of mature CelA and CelC indicates that the location of this domain in the fungal enzymes is not critical for activity. Assuming that the NCRPDs of various hydrolases bind to a scaffolding protein with the same orientation, varying the NCRPD locations may provide conformational variations for the catalytic subunits in the complexes.

Figure 5 shows that the catalytic domains of *Orpinomyces* CelA and CelC are very similar to each other and to the catalytic domain of CELA of the monocentric fungus *N. patriciarum* (14). However, the *N. patriciarum* enzyme has a CBD, whereas the *Orpinomyces* enzymes contain NCRPDs. This suggests that the NCRPD and the catalytic domains have different origins. The parts of the genes encoding the catalytic domains of *Orpinomyces* CelA and CelC and CELA of *N. patriciarum* may have been acquired by horizontal gene transfer between the fungi, and there may have been subsequent duplication in *Orpinomyces* sp. In *Orpinomyces* sp. the CBD part of the gene might have been replaced with the part encoding the NCRPD. Evidence has now been provided by studies of hydrolases of anaerobic fungi that gene transfer, recombination, and duplications among these organisms have occurred fairly frequently (20, 31, 34). Consequently, events have allowed the fungi to diversify the strategies used for plant cell wall breakdown. Cellulases and xylanases with homologous tandem catalytic domains in single polypeptides have been found in *N. patriciarum* (20) and *Piromyces* sp. (17). The finding that CelA and CelC, as well as the three mannanases of *Piromyces* sp. (34), are encoded by separated genes but have highly similar catalytic domains indicates that there has been another type of gene duplication. The presence of a large number of sequences in the genomes of different anaerobic fungi encoding similar NCRPDs and catalytic domains reflects the fact that these sequences are highly selected and that microorganisms in the rumen have shared genetic sources during evolution.

CelA and CelC have both endoglucanase and cellobiohydrolase activities. They rapidly decrease the viscosity of CMC solutions and exhibit activities with ASC and Avicel. The products of CMC and ASC hydrolysis were mainly cellobiose and

cellotriose. No or very little glucose was detected. Cellotriose was slowly hydrolyzed by CelA but was not hydrolyzed at all by CelC. Cellotetraose was hydrolyzed to cellobiose by both enzymes, and cellopentaose was hydrolyzed to cellobiose and cellotriose. These enzymes have high homology with the family 6 glycosyl hydrolases, which consist of endoglucanases and cellobiohydrolases (23, 24). It is apparent that *Orpinomyces* CelA and CelC have both activities.

The three-dimensional structures of the catalytic domains of two family 6 hydrolases, CBHII, a cellobiohydrolase from *Trichoderma reesei* (36), and E2, an endoglucanase from *Thermomonospora fusca* (39), have been determined. The overall topologies of these two enzymes overlap to a high degree despite the fact that they exhibit only 26% sequence identity and the fact that one is a cellobiohydrolase and the other is an endoglucanase. Four aspartic acid residues, Asp-199, Asp-245, Asp-287, and Asp-425, of *Trichoderma reesei* CBHII (the numbering includes the 24 residues of the signal peptide) are present in *Orpinomyces* CelA and CelC. They are conserved in the two types of enzymes in family 6 (Fig. 5) and are found near the proposed site of cleavage of the cellulose chain (36, 39). The catalytic modes of the two types of enzymes are distinguished by the fact that the active-site tunnel of CBHII is enclosed by two surface loops that block the access by long cellulose chains (36). One of the loops is absent in E2, while the other is present but is pulled away due to a deletion adjacent to this loop (39). As a consequence of these changes, the tunnel in E2 is easily accessed by cellulose chains. The loop absent in E2 corresponds to residues from Ser-418 to Gly-436 of *Orpinomyces* CelA (Fig. 5). Deletions of two amino acids for *Orpinomyces* CelA and *N. patriciarum* CELA and of five amino acids for *Orpinomyces* CelC suggest that in the cellulases of anaerobic fungi this loop might only partially enclose the tunnel of the active site. The other loop, which covers the other end of the tunnel of CBHII but is pulled away in E2, is related to the region corresponding to Pro-204 to Ser-217 of *Orpinomyces* CelA (Fig. 5). The three cellulases from anaerobic fungi all have deletions of four amino acids, which may form a loop distinct from the loops of either the aerobic fungal cellobiohydrolases or the bacterial endoglucanases. Nevertheless, the regions of the three cellulases involved in loop formation are distinct from those of cellobiohydrolases and endoglucanases and may allow access of both long cellulose chains and their ends. As a result, these changes may allow the three enzymes of anaerobic fungi to display both endo and exotypes of activities. It should also be pointed out that deletions and insertions of regions other than the loop regions of the three cellulases may contribute to structural changes resulting in enzymes which display both activities.

ACKNOWLEDGMENTS

This work was supported by grant DE-FG05-93ER20127 from the U.S. Department of Energy to L.G.L. and by a grant from the Advanced Technology Development Center, Georgia Institute of Technology, to X.L.L. and L.G.L. Support by a Georgia Power Distinguished Professorship in Biotechnology (to L.G.L.) is also gratefully acknowledged.

REFERENCES

1. **Ali, B. R. S., L. Zhou, F. M. Graves, R. B. Freedman, G. W. Black, H. J. Gilbert, and G. P. Hazlewood.** 1995. Cellulases and hemicellulases of the anaerobic fungus *Piromyces* constitute a multiprotein cellulose-binding complex and are encoded by multigene families. FEMS Microbiol. Lett. **125:**15–22.
2. **Aubert, J.-P., P. Béguin, and J. Millet.** 1992. Genes and proteins involved in cellulose and xylan degradation by *Clostridium thermocellum*, p. 412–422. *In* M. Sebald (ed.), Genetics and molecular biology of anaerobic bacteria. Springer Verlag, New York, N.Y.
3. **Barichievich, E. M., and R. E. Calza.** 1990. Supernatant protein and cellulase activities of the anaerobic ruminal fungus *Neocallimastix frontalis* EB188. Appl. Environ. Microbiol. **56:**43–48.
4. **Bayer, E. A., E. Morag, and R. Lamed.** 1994. The cellulosome—a treasure-trove for biotechnology. Trends Biotechnol. **12:**379–386.
5. **Béguin, P.** 1983. Detection of cellulase activity in polyacrylamide gels using Congo red-stained agar replicas. Anal. Biochem. **131:**333–336.
6. **Béguin, P.** 1990. Molecular biology of cellulose degradation. Annu. Rev. Microbiol. **44:**219–248.
7. **Black, G. W., G. P. Hazlewood, G.-P. Xue, C. G. Orpin, and H. J. Gilbert.** 1994. Xylanase B from *Neocallimastix patriciarum* contains a non-catalytic 455-residue linker sequence comprised of 57 repeats of an octapeptide. Biochem. J. **299:**381–387.
8. **Borneman, S., and D. E. Akin.** 1994. The nature of anaerobic fungi and their polysaccharide degrading enzymes. Mycoscience **35:**199–211.
9. **Borneman, W. S., D. E. Akin, and L. G. Ljungdahl.** 1989. Fermentation products and plant cell wall-degrading enzymes produced by monocentric and polycentric anaerobic ruminal fungi. Appl. Environ. Microbiol. **55:**1066–1073.
10. **Chen, H., X.-L. Li, and L. G. Ljungdahl.** 1995. A cyclophilin from the polycentric anaerobic rumen fungus *Orpinomyces* sp. strain PC-2 is highly homologous to vertebrate cyclophilin B. Proc. Natl. Acad. Sci. USA **92:**2587–2591.
11. **Choi, S. K., and L. G. Ljungdahl.** 1996. Structural role of calcium for the organization of the cellulosome of *Clostridium thermocellum*. Biochemistry **35:**4906–4910.
12. **Chow, C.-M., E. Yague, S. Raguz, D. A. Wood, and C. F. Thurston.** 1994. The *cel3* gene of *Agaricus bisporus* codes for a modular cellulase and is transcriptionally regulated by the carbon source. Appl. Environ. Microbiol. **60:**2779–2785.
13. **Damude, H. G., N. R. Gilkes, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren.** 1993. Endoglucanase CasA from alkalophilic *Streptomyces* strain KSM-9 is a typical member of family B of β-1,4-glucanases. Gene **123:**105–107.
14. **Denman, S., G.-P. Xue, and B. Patel.** 1996. Characterization of a *Neocallimastix patriciarum* cellulase cDNA (*celA*) homologous to *Trichoderma reesei* cellobiohydrolase II. Appl. Environ. Microbiol. **62:**1889–1896.
15. **Doi, R. H., M. Goldstein, S. Hashida, J.-S. Park, and M. Takagi.** 1994. The *Clostridium cellulovorans* cellulosome. Crit. Rev. Microbiol. **20:**87–93.
16. **Durand, R., M. Fischer, C. Rascle, and M. Fevre.** 1995. *Neocallimastix frontalis* enolase gene, *enol*: first report of an intron in an anaerobic fungus. Microbiology **141:**1301–1308.
17. **Fanutti, G., T. Ponyi, G. W. Black, G. P. Hazlewood, and H. J. Gilbert.** 1995. The conserved noncatalytic 40-residue sequence in cellulases and hemicellulases from anaerobic rumen fungi functions as a docking domain. J. Biol. Chem. **270:**29314–29322.
18. **Felix, C. R., and L. G. Ljungdahl.** 1993. The cellulosome: the exocellular organelle of *Clostridium*. Annu. Rev. Microbiol. **47:**791–819.
19. **Fujino, T., P. Béguin, and J.-P. Aubert.** 1993. Organization of a *Clostridium thermocellum* gene cluster encoding the cellulosomal scaffolding protein CipA and a protein possibly involved in attachment of the cellulosome to the cell surface. J. Bacteriol. **175:**1891–1899.
20. **Gilbert, H. J., G. P. Hazlewood, J. I. Laurie, C. G. Orpin, and G. P. Xue.** 1992. Homologous catalytic domains in a rumen fungal xylanase: evidence for gene duplication and prokaryotic origin. Mol. Microbiol. **6:**2065–2072.
21. **Gilkes, N. R., B. Henrissat, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren.** 1991. Domains in microbial β-1,4-glycanases: sequence conservation, function, and enzyme families. Microbiol. Rev. **55:**303–315.
22. **Hansen, S. A.** 1975. Thin-layer chromatographic method for identification of oligosaccharides in starch hydrolyzates. J. Chromatogr. **105:**388–390.
23. **Henrissat, B., M. Claeyssens, P. Tomme, L. Lemesle, and J.-P. Mornon.** 1989. Cellulase families revealed by hydrophobic cluster analysis. Gene **81:**83–85.
24. **Henrissat, B., and A. Bairoch.** 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. **293:**781–788.
25. **Kruus, K., A. C. Lua, A. L. Demain, and J. H. D. Wu.** 1995. The anchorage function of CipA (CelL), a scaffolding protein of the *Clostridium thermocellum* cellulosome. Proc. Natl. Acad. Sci. USA **92:**9254–9258.
26. **Lake, B. D., and H. J. Goodwin.** 1976. Lipids, p. 345–366. *In* I. Smith and J. M. T. Seakin (ed.), Chromatographic and electrophoretic techniques, 4th ed., vol. 1. Pitman Press, Bath, England.
27. **Lao, G., G. S. Ghangas, E. D. Jung, and D. B. Wilson.** 1991. DNA sequences of three β-1,4-endoglucanase genes from *Thermomonospora fusca*. J. Bacteriol. **173:**3397–3407.
28. **Li, X.-L., and R. E. Calza.** 1991. Fractionation of cellulases from the ruminal fungus *Neocallimastix frontalis* EB188. Appl. Environ. Microbiol. **57:**3331–3336.
29. **Li, X.-L., and L. G. Ljungdahl.** 1996. Expression of *Aureobasidium pullulans xynA* in, and secretion of the xylanase from, *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. **62:**209–213.
30. **Li, X.-L., H. Z. Chen, Y. He, D. L. Blum, and L. G. Ljungdahl.** 1997. High

molecular weight cellulase/hemicellulase complexes of anaerobic fungi, abstr. O-31, p. 424. *In* Abstracts of the 97th General Meeting of the American Society for Microbiology 1997. American Society for Microbiology, Washington, D.C.
31. **Li, X.-L., H. Chen, and L. G. Ljungdahl.** 1997. Monocentric and polycentric anaerobic fungi produce structurally related cellulases and xylanases. Appl. Environ. Microbiol. **63:**628–635.
32. **Lowe, S. E., M. K. Theodorou, and A. P. J. Trinci.** 1987. Cellulases and xylanase of an anaerobic rumen fungus grown on wheat straw, wheat straw holocellulose, cellulose, and xylan. Appl. Environ. Microbiol. **53:**1216–1223.
33. **Miller, G. L.** 1959. Use of dinitrosalylic acid reagent for determination of reducing sugars. Anal. Chem. **31:**127–132.
34. **Millward-Sadler, S. J., J. Hall, G. W. Black, G. P. Hazlewood, and H. J. Gilbert.** 1996. Evidence that the *Piromyces* gene family encoding endo-1,4-mannase arose through gene duplication. FEMS Microbiol. Lett. **141:**183–188.
35. **Orpin, C. G.** 1975. Studies on the rumen flagellate *Neocallimastix frontalis*. J. Gen. Microbiol. **91:**249–262.
36. **Rouvinen, J., T. Bergfors, T. Teeri, J. K. C. Knowles, and T. A. Jones.** 1990. Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. Science **249:**380–386.
37. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
38. **Sheppard, P. O., F. J. Grant, P. J. Oort, C. A. Sprecher, D. C. Foster, F. S. Hagen, A. Upshall, G. L. Mcknight, and P. J. O'Hara.** 1994. The use of conserved cellulase family-specific sequences to clone cellulase homologue cDNAs from *Fusarium oxysporum*. Gene **150:**163–167.
39. **Spezio, M., D. B. Wilson, and P. A. Karplus.** 1993. Crystal structure of the catalytic domain of a thermophilic endocellulase. Biochemistry **32:**9906–9916.
40. **Teeri, T. T., P. Lehtovaara, S. Kauppinnen, I. Salouori, and J. Knowles.** 1987. Homologous domains in *Trichoderma reesei* cellulolytic enzymes: gene sequence and expression of cellobiohydrolase II. Gene **51:**43–52.
41. **Warren, R. A. J.** 1996. Microbial hydrolysis of polysaccharides. Annu. Rev. Microbiol. **50:**183–212.
42. **Wilson, C. A., and T. M. Wood.** 1992. The anaerobic fungus *Neocallimastix frontalis*: isolation and properties of a cellulosome-type enzyme fraction with the capacity to solubilize hydrogen-bond-ordered cellulose. Appl. Microbiol. Biotechnol. **37:**125–129.
43. **Wong, W. K. R., B. Gerhard, Z. M. Guo, D. G. Kilburn, R. A. J. Warren, and R. C. Miller, Jr.** 1986. Characterization and structure of an endoglucanase gene *cenA* of *Cellulomonas fimi*. Gene **44:**315–324.
44. **Wood, T. M.** 1988. Preparation of crystalline, amorphous, and dyed cellose substrates. Methods Enzymol. **160:**19–25.
45. **Wood, T. M., C. A. Wilson, S. I. McCrae, and K. N. Joblin.** 1986. A highly active extracellular cellulase from the anaerobic rumen fungus *Neocallimastix frontalis*. FEMS Microbiol. Lett. **34:**37–40.
46. **Yokoyama, M. T., and K. A. Johnson.** 1988. Microbiology of the rumen and intestine, p. 125–144. *In* D. C. Church (ed.), The ruminant animal: digestive physiology and nutrition. Reston Book, Prentice Hall, Englewood Cliffs, N.J.
47. **Zhou, L., G.-P. Xue, C. G. Orpin, G. W. Black, H. J. Gilbert, and G. P. Hazlewood.** 1994. Intronless *celB* from the anaerobic fungus *Neocallimastix patriciarum* encodes a modular family A endoglucanase. Biochem. J. **297:**359–364.