APPLIED AND ENVIRONMENTAL MICROBIOLOGY, June 1996, p. 1889–1896 Vol. 62, No. 6
0099-2240/96/$04.00+0

Characterization of a *Neocallimastix patriciarum* Cellulase cDNA (*celA*) Homologous to *Trichoderma reesei* Cellobiohydrolase II

STUART DENMAN,[1,2] GANG-PING XUE,[1]* AND BHARAT PATEL[2]

CSIRO Division of Tropical Crops and Pastures, St. Lucia, Queensland 4067,[1] and Faculty of Science and Technology, Griffith University, Nathan, Queensland 4111,[2] Australia

Received 27 February 1996/Accepted 7 March 1996

The nucleotide sequence of a cellulase cDNA (*celA*) from the rumen fungus *Neocallimastix patriciarum* and the primary structure of the protein which it encodes were characterized. The *celA* cDNA was 1.95 kb long and had an open reading frame of 1,284 bp, which encoded a polypeptide having 428 amino acid residues. A sequence alignment showed that cellulase A (CELA) exhibited substantial homology with family B cellulases (family 6 glycosyl hydrolases), particularly cellobiohydrolase II from the aerobic fungus *Trichoderma reesei*. In contrast to previously characterized *N. patriciarum* glycosyl hydrolases, CELA did not exhibit homology with any other rumen microbial cellulases described previously. Primary structure and function studies in which deletion analysis and a sequence comparison with other well-characterized cellulases were used revealed that CELA consisted of a cellulose-binding domain at the N terminus and a catalytic domain at the C terminus. These two domains were separated by an extremely Asn-rich linker. Deletion of the cellulose-binding domain resulted in a marked decrease in the cellulose-binding ability and activity toward crystalline cellulose. When CELA was expressed in *Escherichia coli*, it was located predominantly in the periplasmic space, indicating that the signal sequence of CELA was functional in *E. coli*. Enzymatic studies showed that CELA had an optimal pH of 5.0 and an optimal temperature of 40°C. The specific activity of immunoaffinity-purified CELA against Avicel was 9.7 U/mg of protein, and CELA appeared to be a relatively active cellobiohydrolase compared with the specific activities reported for other cellobiohydrolases, such as *T. reesei* cellobiohydrolases I and II.

Cellulose is a polymer of glucose linked by β-1,4-glucosidic bonds. Cellulose chains form numerous intra- and intermolecular hydrogen bonds, which result in the formation of insoluble cellulose microfibrils. Microbial hydrolysis of cellulose to glucose involves the following three major classes of cellulases: (i) endoglucanases (EC 3.2.1.4), which cleave β-1,4-glucosidic links randomly throughout cellulose molecules; (ii) cellobiohydrolases (EC 3.2.1.91), which digest cellulose from the nonreducing end, releasing cellobiose; and (iii) β-glucosidases (EC 3.2.1.21), which hydrolyze cellobiose and low-molecular-mass cellodextrins to release glucose. These three types of enzymes work synergistically to degrade cellulose (2).

Cellulases are produced by many microorganisms and are often present in multiple forms. Recognition of the economic significance of the enzymatic degradation of cellulose has promoted an extensive search for microbial cellulases which may be used industrially. As a result, the enzymatic properties and the primary structures of a large number of cellulases have been investigated. On the basis of the results of a hydrophobic cluster analysis of the amino acid sequence of the catalytic domain, these cellulases have been placed into 11 families (18, 19). Most cellulases consist of a cellulose-binding domain (CBD) and a catalytic domain separated by a linker rich in proline and hydroxy amino acid residues. In the past decade, studies on the molecular biological aspects of cellulases have focused mainly on bacteria and aerobic fungi; however, relatively little is known about the molecular structure of the cellulases of anaerobic fungi.

The rumen anaerobic fungus *Neocallimastix patriciarum* has been shown to possess diverse plant polysaccharide hydrolase activities, a high capacity for cellulose degradation, and the ability to grow on cellulose as a sole carbohydrate source (44). The *Neocallimastix* cellulases have been shown to form a large multienzyme complex, which exhibits very high cellulase activity against crystalline cellulose (45). Studies on the enzymatic properties of individual native cellulases belonging to the rumen fungal cellulase complex have been hampered by difficulties encountered in obtaining purified native enzymes. Recently, two groups of workers have reported the cloning of cellulase and xylanase genes from the rumen fungus *N. patriciarum* (36, 47, 48). Analysis of the primary structure of one endoglucanase and two xylanases from this rumen fungus revealed that there is substantial sequence homology between rumen eukaryotic and prokaryotic genes (4, 16, 51); thus, there is evidence of horizontal transfer of genes between rumen fungi and bacteria. In this paper, we describe the characterization of the primary structure of an *N. patriciarum* cellulase cDNA (*celA*) which appears to share substantial sequence homology with cellobiohydrolase II (CBHII) from the aerobic fungus *Trichoderma reesei*. We also describe enzymatic properties of the cellulase encoded by *celA*.

MATERIALS AND METHODS

Microbial strains, vectors, and culture media. The anaerobic fungus *N. patriciarum* (the type species of the genus *Neocallimastix*) obtained from a sheep rumen was cultivated as described previously (48). *Escherichia coli* XL1-Blue (Stratagene) was grown in L broth (30). For expression studies, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to LBPG medium (5 g of Bacto Yeast Extract per liter, 10 g of Bacto Tryptone per liter, 5 g of NaCl per liter, 3.5 g of K_2HPO_4 per liter, 1.5 g of KH_2PO_4 per liter, 5 mg of thiamine per liter, 4 g of glucose per liter) to a final concentration of 1 mM. λZAPII and pBluescript SK vectors were supplied by Stratagene, and recombinant phage were grown in *E. coli* strains as specified by the supplier.

General DNA manipulation and DNA sequence analysis. DNA isolation, restriction endonuclease digestion, ligation, and transformation were performed by using the procedures described by Sambrook et al. (30). DNA amplification by PCR was performed by using *Taq* DNA polymerase (Boehringer). The methods used to construct the *N. patriciarum* cDNA library in λZAPII and isolate cellu-

* Corresponding author. Mailing address: CSIRO Division of Tropical Crops and Pastures, 306 Carmody Road, St. Lucia, Qld 4067, Australia. Phone: (617) 3377 0403. Fax: (617) 33713946. Electronic mail address: gang.ping.xue@tcp.csiro.au.

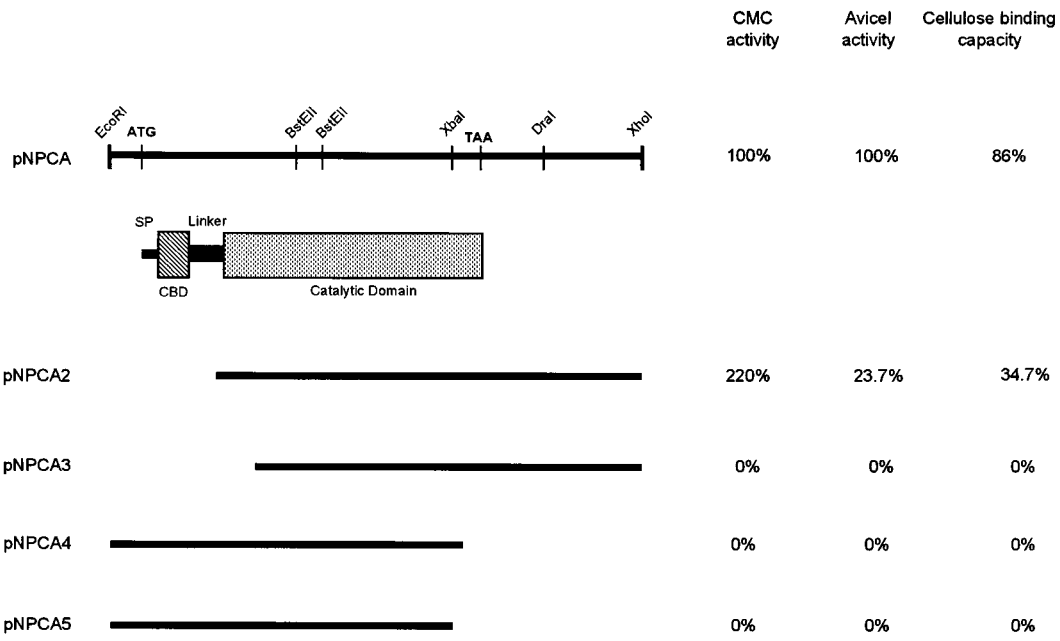

FIG. 1. Restriction map of *celA* cDNA and its deletion mutants. The precise positions of the *celA* deletion mutants are shown in Fig. 2. Values for CMC and Avicel activity are expressed as percentages compared with pNPCA activity. The Avicel activity experiment was carried out without adsorption to Avicel. Values for cellulose-binding capacity were calculated as the percentages of the retained activity after adsorption to Avicel and washing to remove unbound proteins compared with the amount of activity against Avicel without adsorption. SP, signal peptide.

lase cDNA clones have been described previously (48). Exonuclease III deletions of DNA were performed to generate overlapping sequences for DNA sequencing and for functional analysis as described in the *Promega Protocols and Applications Guide* (41).

DNA sequencing by the dideoxynucleotide chain termination method was performed by using single-stranded DNAs as templates and the T7 DNA polymerase sequencing system (Promega) as recommended by the supplier. The complete sequence of the cDNA contained in pNPCA was determined for both strands. The sequence data were analyzed with SEQAID (26). The deduced amino acid sequence was sent to the electronic mail server BLITZ (blitz@heidelberg-embl.de) (34), where the sequence was compared with the SWISS-PROT protein database. Multiple alignments were carried out by using Clustal W (39).

Construction of *celA* cDNA deletion mutants. pNPCA2 was constructed by exonuclease III deletion of the 5′ end of the *celA* cDNA in the pBluescript vector, and the 5′ end of the truncated *celA* cDNA was fused in frame to the sequence coding for the N terminus of the *lacZ* α-peptide present in the pBluescript vector. An N-terminal deletion mutant, pNPCA3, was constructed by PCR amplification of the *celA* cDNA in the pNPCA construction by using sense primer DCF1 (5′GGAATTCATGGCTCAACATCTTAAAGCTGC) and antisense primer CR1 (5′GAGCAAGCTTAAAATGATGGTCTAGC) (the underlined sequences in the primers are additional sequences that represent the restriction sites for *Eco*RI and *Hin*dIII, respectively). The PCR amplification product was digested with *Eco*RI and *Hin*dIII and ligated to the *Eco*RI-*Hin*dIII-digested pBTac2 vector (Boehringer). An ATG start codon was incorporated into the sense primer, DCF1, so that *celA* was under the control of the *tac* promoter. C-terminal deletion mutants pNPCA4 and pNPCA5 were constructed by PCR amplification of the 3′ region of the *celA* cDNA in plasmid pNPCA, using a sense primer (primer T3, corresponding to the vector sequence upstream of the *celA* cDNA) and two antisense primers (primer R2 [5′CGTCTCGAGTTAATCAGCTCTAGAACCAGAG] for pNPCA4 and primer R3 [5′CGTCTCGAGTTACATGTAAGCATCTAATAATGG] for pNPCA5) (the underlined sequences in the primers are additional sequences that represent an *Xho*I restriction site). The PCR amplification products were digested with *Bst*XI and *Xho*I and ligated to the *Bst*XI-*Xho*I-digested pBluescript vector. An in-frame stop codon (TAA) was incorporated into the antisense primers, so that the truncated cellulase A (CELA) molecules in pNPCA4 and pNPCA5 were synthesized as nonfusion proteins. In order to avoid possible point mutations arising during PCR amplification, multiple isolates (at least six isolates with the expected restriction pattern of the constructions) from two PCRs were analyzed for cellulase activity in the construction of pNPCA3, pNPCA4, and pNPCA5. The truncated *celA* cDNA was checked either by determining the restriction pattern or by partial nucleotide sequencing at the insertion termini. The deletion mutants are shown in Fig. 1.

Construction of Flag-tagged CELA. An expression-secretion vector (pSF-tac) was modified from vector pBTac2 (Boehringer) to include a synthetic signal peptide (MKRGKLWGRLVSAAGLSLSISSIGNVSTAYA) followed by the sequence for a Flag peptide (DYKDDDDK). The sequence encoding the mature CELA, as deduced from the nucleotide sequence, was amplified by PCR, using sense primer CS (5′CGGGATCCTGCTTGTGGTGGTGCCTGG), which corresponds to the N terminus of mature CELA, and anti-sense primer CR1 (5′GAGCAAGCTTAAAATGATGGTCTAGC) which corresponds to the carboxy terminus (the underlined sequences in primers CS and CR1 are additional sequences that represent restriction sites for *Bam*HI and *Hin*dIII, respectively). The PCR product was digested with *Bam*HI, which was followed by filling in with the Klenow enzyme, and was then digested with *Hin*dIII. The *Bam*HI-*Hin*dIII-digested *celA* PCR product was then ligated to the pSF-Tac vector which had previously been cut with *Bam*HI and filled in and then digested with *Hin*dIII. The resulting plasmid, designated pBSFCA, encoded the N terminus of mature CELA fused to the C terminus of the Flag peptide. Expression of the FLAG-tagged CELA was under the control of the *tac* promoter. Several cellulase-positive isolates were quantitatively analyzed for cellulase activity, and they all exhibited similar enzyme levels.

Enzyme analyses. *E. coli* cells harboring recombinant plasmids were grown in LBPG medium containing ampicillin (100 μg/ml) and 1 mM IPTG to the end of the exponential growth phase. Crude lysates were prepared by treatment with lysozyme (1 mg/ml) in a buffer containing 50 mM TrisCl (pH 8.0) and 1 mM EDTA. After incubation on ice for 30 min, the cells were frozen at −80°C and then thawed at room temperature to lyse the cells. The lysates were homogenized with a Polytron apparatus for 60 s to shear the DNA and centrifuged to separate the cell debris. Periplasmic and cytoplasmic fractions were prepared for cellular location studies by using the protocol described by Sambrook et al. (30). Measurement of cellulase activity and assays to determine the cellulose-binding capacities of the cloned cellulases were carried out as described previously (48) by using the following substrates: 0.4% (wt/vol) carboxymethyl cellulose (CMC) (low viscosity; Sigma), 1% (wt/vol) Avicel (Merck), 1% (wt/vol) amorphous cellulose (H_3PO_4-swollen Avicel), 0.4% (wt/vol) lichenan (practical grade; Sigma), 0.5% (wt/vol) oat spelt xylan (Sigma), and 0.05% *p*-nitrophenyl cellobioside (Sigma). All assays were carried out under reaction conditions within the range of a linear response with respect to incubation time and the amount of enzyme added. The rates of hydrolysis were essentially linear for at least 1 h, provided that the proper enzyme concentration was used. One unit of cellulase activity was defined as 1 μmol of reducing sugars released from substrates per min when glucose was used as a standard. The protein concentrations of the enzyme preparations were determined by a bicinchoninic acid (BCA) protein assay method in which a Micro BCA protein assay reagent kit (Pierce) was used according to the supplier's instructions.

A zymogram analysis was carried out by using a sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) gel containing 0.3% (wt/vol) lichenan to separate the proteins. Lichenan was used as the substrate in the

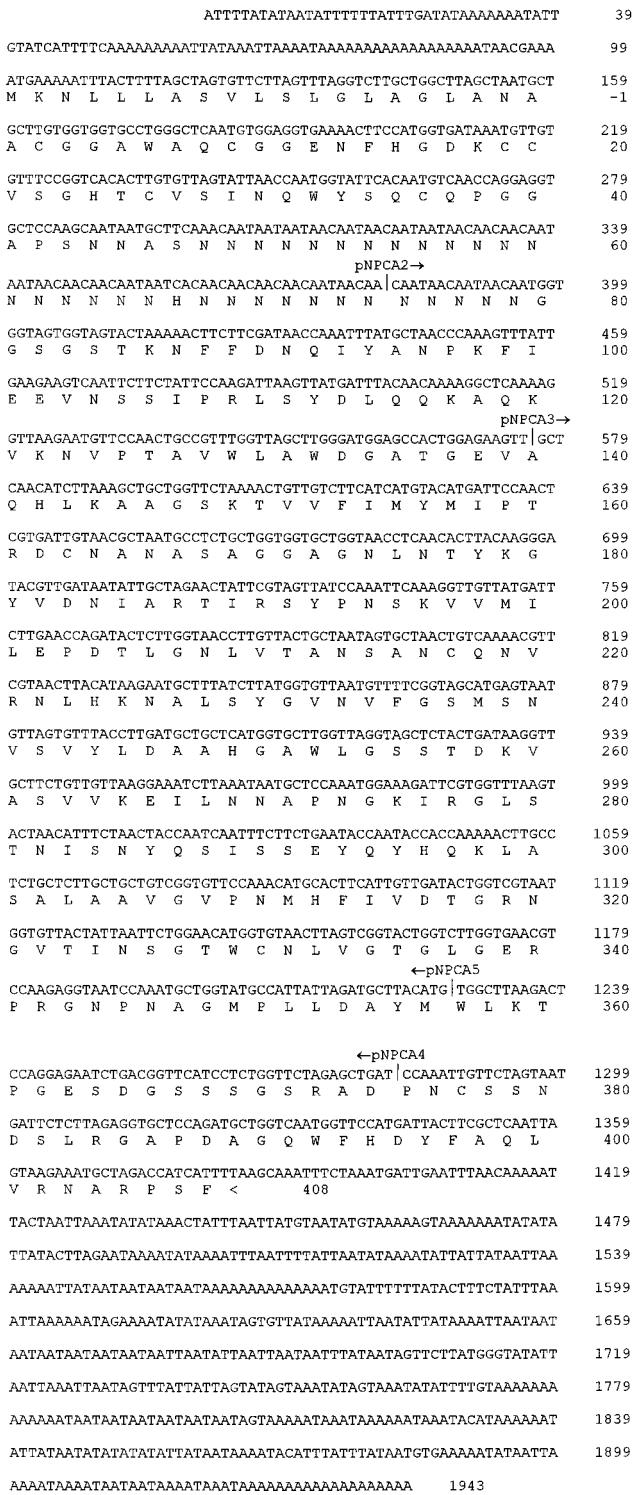

FIG. 2. Nucleotide sequence of *celA* cDNA aligned with the deduced amino acid sequence. The numbers for the nucleic acid sequence start with the first base of the cDNA, and the amino acid numbering begins with the predicted signal peptide cleavage site for the mature protein. The CBD spans residues 2 to 37, the linker spans residues 38 to 81, and the catalytic domain spans residues 82 to 408.

zymogram analysis because CELA is more active against this compound than against CMC and thus the sensitivity of the analysis was increased. The proteins were denatured at 50°C for 10 min in the presence of SDS and β-mercaptoethanol. The denatured proteins were loaded onto an SDS denaturing gel for separation. After electrophoresis the proteins were renatured and stained to reveal cellulase activity as described by Schwarz et al. (31).

The Flag-tagged CELA was purified from the periplasmic fraction of the enzyme preparation by using an anti-Flag M1 affinity gel (IBI) according to the supplier's instructions. Purified protein samples were electrophoresed on a gradient SDS–10 to 15% polyacrylamide PAGE gel by using the Phast System (Pharmacia). To remove the Flag peptide from the recombinant protein, 2 μg of Flag-tagged CELA was incubated for 8 h at 37°C in a 50-μl reaction mixture containing 10 mM Tris (pH 8), 10 mM $CaCl_2$ buffer, and 600 ng of enterokinase (Boehringer).

The extent of Avicel hydrolysis was assayed at 40°C for 24 h in a 1-ml reaction mixture containing 0.2% Avicel, 50 mM sodium citrate (pH 6), and 3 mg of crude CELA extract (total cell lysate) from the Flag-tagged *celA* construct. An extract of *E. coli* cells containing only the vector was used as a control. A zero blank was also used so that the amount of reducing sugars present in the enzyme extract could be subtracted. The amounts of the reducing sugars released from Avicel digestion were determined to calculate the extent of hydrolysis.

pH and temperature profiles were analyzed by varying the pH values of the buffers (50 mM sodium citrate [pH 4 to 7] and 25 mM Tris-Cl–50 mM NaCl [pH 7.5 to 9.5]) and the hydrolysis temperatures and using the periplasmic fraction of the Flag-tagged CELA. The assays were performed under conditions which gave a linear response for hydrolysis of Avicel. An incubation temperature of 40°C was used for the pH profile analysis, while the temperature profile was analyzed at pH 6.

Nucleotide sequence accession number. The GenBank accession number for the *celA* sequence of *N. patriciarum* is U29872.

RESULTS

Nucleotide sequence. The *celA* cDNA was isolated previously from an *N. patriciarum* cDNA library (48), and the restriction map is shown in Fig. 1. The cDNA was 1.95 kb long and contained an open reading frame (ORF) of 1,284 nucleotides. The nucleotide sequence and deduced amino acid sequence are shown in Fig. 2. The enzyme encoded by the *celA* cDNA consisted of a polypeptide of 428 amino acids with a calculated molecular mass of 45,681 Da. The cloned CELA enzyme appeared to be synthesized from its own translational start codon under the control of the *lacZ* promoter in the pBluescript vector. Assignment of the proposed translation initiation codon was based on the following observations: (i) there are no ATG sequences upstream of the proposed ORF; (ii) an ORF shift at the vector polyclonal site (the vector was cut with *Bam*HI upstream of the putative start and then filled in and religated, which led to a frameshift of +1) did not affect the expression of cellulase activity, suggesting that there is an internal translation start codon; (iii) translation stop codons are present in all three ORFs upstream of the proposed translation initiation codon; (iv) a typical signal peptide occurs at the N terminus of the ORF; and (v) correct assignment of the ORF is supported by deletion of the 5′ untranslated region of *celA* cDNA and an in-frame fusion with the N-terminal sequence of *lacZ* (the pNPCA2 construction). Fungal mRNAs do not contain a prokaryotic Shine-Dalgarno sequence for translation initiation. Examination of the *celA* sequence upstream of the proposed translational initiation codon did not reveal a typical *E. coli* Shine-Dalgarno-like sequence. However, the sequence TAACGA, which is 2 bp upstream of the ATG start codon, partially matches the 3′ end of the *E. coli* 16S rRNA and may act as a very weak ribosomal binding site to allow the expression of the CELA enzyme in *E. coli.*

The G+C content of the *celA* ORF was 37%, compared with 27% for the entire *celA* cDNA. The overall G+C content of *Neocallimastix* DNA is approximately 15% (3, 5), indicating that noncoding regions of the genome are generally very A+T rich. Like other *N. patriciarum* cellulase and xylanase genes (4, 16, 51), *celA* showed a clear bias in codon usage, with a marked

```
CELA        86  ----KNFFDNQIYANPKFI--EEVNSSIPRLSYDLQQKAQKVKNVPTAVWLA----WDGA
CBHII       87  TYSGNPFVGVTPWANAYYA-SEVSSLAIPSLTGAMATAAAAVAKVPSFMWLDTLDKTPLM
MICBI       35  -------YDSPFYVDPQSN-AAKWVAANPNDPR-TPVIRDRIAAVPTGRWFA-NYNPSTV
THEFU       32  -------NDSPFYVNPNMS-SAEWVRNNPNDPR-TPVIRDRIASVPQGTWFA-HHNPGQI
CELFI      134  --TVTPQPTSGFYVDPTTQGYRAWQAASGTD----KALLEKIALTPQAYWVGNWADASHA
STRSQ       71  ---GTTALPSMELYRAEAG-VHAWLDANPGDHR-APLIAERIGSQPQAVWFAGAYNPGTI
STRHA       28  --ADPTTMTNGFYADPDSS-ASRWAAANPGDGR-AAAINASIANTPMARWFG--SWSGAI
                                                 .   *   *

CELA       137  TGEVAQHLKAAG---SKTVVFIMYMIPTRDCNANASAGG-----AGNLNTYKGYVDNIAR
CBHII      146  EQTLADIRTANKNGGNYAGQFVVYDLPDRD-CAALASNGEYSIADGGVAKYKNYIDTIRQ
MICBI       85  RAEVDAYVGAAAAA-GKIPIMVVYAMPNRD-CGGPSAGG-----APNHTAYRAWIDEIAA
THEFU       82  TGQVDALMSAAQAA-GKIPILVVYNAPGRD-CGNHSSGG-----APSHSAYRSWIDEFAA
CELFI      188  QAEVADYTGRAVAA-GKTPMLVVYAIPGRD-CGSHSGGG-----VS-ESEYARWVDTVAQ
STRSQ       126 TQQVAEVTSAAAAA-GQLPVVVPYMIPFRD-CGNHSGGG-----APSFAAYAEWSGLFAA
STRHA       82  GTAAGAYAGAADGR-DKLPILVAYNIYNRDYCGGHSAGG-----AASPSAYADWIARFAG
                              . *    **     . .*          *

CELA       188  TIRSYPNSKVVMILEPDTLGNLVTANSA-NCQNVRNLHKNALSYGVNVFGSMSNVSVYLD
CBHII      205  IVVEYSDIRTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPN-VAMYLD
MICBI      138  GLRNRPA---VIILEPDALPIMTNCMSP-SEQAEVQASAVGAGKKFKAASSQ--AKVYFD
THEFU      135  GLKNRPA---YIIVEPDLISLMSSCMQ--HVQQEVLETMAYAGKALKAGSSQ--ARIYFD
CELFI      240  GIKGNP----IVILEPDALAQLGDCSG----QGDRVGFLKYAAKSLTLKG----ARVYID
STRSQ       179 GLGSEPV---VVVLEPDAIPLIDCLDN--QQRAERLAALAGLAEAVTDANPE--ARVYYD
STRHA       136 GIAARPA---VVILEPDSLGDYGCMNP--AQIDEREAMLTNALVQFNRQAPN--TWVYMD
                 .         ...*** .                                  .* *

CELA       247  AAHGAWLGS------STDKVASVVKEILNNAPNGKIRGLSTNISNYQ-------------
CBHII      264  AGHAGWLGWPANQDPAAQLFANVYKNASSPR---ALRGLATNVANYNGWNITSPPSYTQG
MICBI      192  AGHDAWV--------PADEMASRLRGADIAN---SADGIALNVSNYR-------------
THEFU      188  AGHSASD--------SPQQMASWLQQADISN---SAHGIATNTSNYR-------------
CELFI      288  AGHAKWL--------SVDTPVNRLNQVGFE----YAVGFALNTSNYQ-------------
STRSQ      232  VGHSAWH---------APAAIAPTLVEAGILE---HGAGIATNISNYR-------------
STRHA      189  AGNPRWA--------DAATMARRLHEAGLR----QAHGFSLNVSNYI-------------
                  .                              * . * .**

CELA       288  -SISSEYQYHQKLASALAAVGVPNMHFIVDTGRNGVTINSG----TWCNLVGTGLGERPR
CBHII      321  NAVYNEKLYIHAIGPLLANHGWSNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRP-
MICBI      228  -YTSGLISYAKSVLSAIGASH---LRAVIDTSRNGNGPLGS----EWCDPPGRATGT-W-
THEFU      224  -WTADEVAYAKAVLSAIGNPS---LRAVIDTSRNGNGPAGN----EWCDPSGRAIGT-P-
CELFI      323  -TTADSKAYGQQISQRLGGK-----KFVIDTSRNGNGSNG-----EWCNPRGRALGERP-
STRSQ       268 -TTTDETAYASAVIAELGGG----LGAVVDTSRNGNGPLGS----EWCDPPGRLVGN-N-
STRHA       224 -TTAENTAYGNAVNNELAARYGYTKPFVVDTSRNGNGSNG-----EWCNPSGRRIGT-P-
                        *   .   .          . *  * *           **   *   *

CELA       343  GNPNAGMPLLDAYMWLKTPGESDGSS--SGSRADPNCSSNDSLRGAPDAGQWFHDYFAQL
CBHII      380  -SANTGDSLLDSFVWVKPGGECDGTSDSSAPRFDSHCALPDALQPAPQAGAWFQAYFVQL
MICBI      278  STTDTGDPAIDAFLWIKPPGEADG------------CIAT--------PGVFVPDRAYEL
THEFU      274  STTNTGDPMIDAFLWIKLPGEADG------------CIAG--------AGQFVPQAAYEM
CELFI      371  VAVNDGS-GLDALLWVKLPGESDGA-----------CNGG------PAAGQWWQEIALEM
STRSQ      317  PTVNPGVPGVDAFLWIKLPGELDG------------CDGP--------VGSFSPAKAYEL
STRHA      276  TRTGGG---AEMLLWIKTPGESDGN-----------CGVGS----GSTAGQFLPEVAYKM
                     *    .  .*.*  ** **            *           *         .

CELA       401  VRNARPSF-
CBHII      439  LTNANPSFL
MICBI      318  AMNAAPPTY
THEFU      314  AIAAGGH--
CELFI      413  ARNARW---
STRSQ      357  AGG------
STRHA      318  IYGY-----
```

FIG. 3. Clustal W (1.5) multiple-sequence alignment of the catalytic domains of *N. patriciarum* CELA (CELA), *Trichoderma reesei* CBHII (27) (CBHII), *M. bispora* CELA (49) (MICBI), *Thermomonospora fusca* CELB (22) (THEFU), *C. fimi* CENA (46) (CELFI), *Streptomyces* sp. strain Ksm-9 CASA (9) (STRSQ), and *S. halstedii* CELAI (14) (STRHA). A dot indicates similar residues, an asterisk indicates identical residues, and slightly unaligned Cys residues that are involved in disulfide formation are indicated by boldface type.

preference for a T in the third position (56% of all codons contained T in the third position).

Homology, primary structure, and function studies. In contrast to the previously characterized *N. patriciarum* enzymes, the comparison of the deduced amino acid sequence of CELA with the sequences in the SWISS-PROT protein database showed that CELA did not share sequence homology with any of the other rumen microbial enzymes that have been characterized to date. However, the CELA sequence exhibited homology with sequences in the catalytic domain of family B cellulases or family 6 glycosyl hydrolases (19). The sequence with the highest level of homology to CELA was the CBHII sequence of *Trichoderma reesei* (level of similarity, 53%; level of identity in the catalytic domain region, 37%), followed by the sequences of *Microbispora bispora* CELA, *Thermomonospora fusca* CELB, *Cellulomonas fimi* CENA, *Streptomyces* sp. strain Ksm9 CASA, and *Streptomyces halstedii* CELA1. An alignment of the sequences of these enzymes and *N. patriciarum* CELA is shown in Fig. 3.

Analysis of the deduced amino acid sequence showed that CELA contained a typical eukaryotic signal peptide sequence consisting of 20 amino acid residues (Fig. 2). The predicted

```
                                                                #
CELA          2   CGGAWAQCGGENFHGDKCCVSGHTCVSINQWYSQCQ
TrEgI       424   TQTHWGQCGGIGYSGCKTCTSGTTCQYSNDYYSQCL
TrEgIII       1   QQTVWGQCGGIGWSGPTNCAPGSACSTLNPYYAQCI
HgCbhI      491   KAGRWQQCGGIGFTGPTQCEEPYICTKLNDWYSQCL
TrCbhII       3   CSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCL
TrCbhI      462   TQSHYGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL
TvCbhI      461   TQTHYGQCGGIGYIGPTVCASGSTCQVLNPYYSQCL
NcCbhI      478   GAAHWAQCGGIGFSGPTTCPEPYTCAKDHDIYSQCV
PcCbhI      481   TVPQWGQCGGIGYTGSTTCASPYTCHVLNPYYSQCY
PjCbhI      502   GARDWAQCGGNGWTGPTTCVSPYTCTKQNDWYSQCL
                        **** .  *   *     *   .  *.**
```

FIG. 4. Clustal W (1.5) multiple-sequence alignment of the fungal CBD type 2 family. CELA, *N. patriciarum* CELA; TrEgI, *Trichoderma reesei* EGLI (24); TrEgIII, *Trichoderma reesei* EGLIII (29); HgCbhI, *Humicola grisea* CBHI (11); TrCbhII, *Trichoderma reesei* CBHII (38); TrCbhI, *Trichoderma reesei* CBHI (32); TvCbhI, *Trichoderma viride* CBHI (6); NcCbhI, *Neurospora crassa* CBHI (35); PcCbhI, *Phanerochaete chrysosporium* CBHI (33); PjCbhI, CBHI, *Penicillium janthinellum* CBHI (20). Dots indicate similar residues, asterisks indicate identical residues, and the number sign indicates the position of Tyr-492 in *Trichoderma reesei.*

cleavage site of the signal peptide is located at alanine −1. This is consistent with the expectation that CELA is an extracellular enzyme of the rumen fungus *N. patriciarum.*

Similar to the primary structure of *Trichoderma reesei* CBHII, the CBD of CELA is located at the N terminus of the mature enzyme, whereas the catalytic domain is at the C terminus (Fig. 1). These two domains are separated in CELA by an Asn-rich linker.

The sequence alignment shown in Fig. 4 revealed that the CBDs of fungal enzymes appear to be highly conserved, as the CELA CBD exhibited a high degree of homology with the CBD of aerobic fungal cellulases, particularly *Trichoderma reesei* CBHII (level of identity, 57%; level of similarity, 77%). This type of CBD belongs to the type 2 family, which so far is exclusive to fungal cellulases (17).

A deletion analysis of the function of the CELA CBD revealed that the Avicel-hydrolyzing activity of CELA without the CBD (the pNPCA2 construct) was markedly reduced, whereas the CMC-hydrolyzing activity of the pNPCA2 enzyme was higher than that of full-length CELA (Fig. 1). Thus, the ratio of Avicel-hydrolyzing activity to CMC-hydrolyzing activity in CELA lacking the CBD was reduced ninefold compared with the ratio obtained with full-length CELA. This indicates that there was a marked reduction in crystalline cellulose-hydrolyzing activity after the CBD was removed. To assess the effect of removal of the CBD on the cellulose-binding capacity of CELA, the relative activities (Avicel-hydrolyzing activities before and after adsorption) of full-length CELA and CELA lacking the CBD were compared. As shown in Fig. 1, markedly reduced recovery of pNPCA2 cellulase activity after preadsorption and washing was observed compared with the recovery of full-length CELA activity, clearly indicating that the cellulose-binding capacity was reduced after the CBD was removed.

Deletion of the amino acid residues of the catalytic domain at either end in the constructs pNPCA3, pNPCA4, and pNPCA5 resulted in a complete loss of catalytic activity (Fig. 1). These three deletion mutants were constructed by using the PCR method; an in-frame TAA stop codon was introduced into the 3′ end of the truncated *celA* cDNA (pNPCA4 and pNPCA5), or a translation start codon (ATG) was introduced into the 5′ end of the truncated cDNA (pNPCA3). The C termini of pNPCA4 and pNPCA5 end at Asp-374 and Met-356, respectively, while the N terminus of pNPCA3 starts at Ala-140 (Fig. 2). Several isolates that produced the expected restriction pattern were screened on agar plates containing CMC, and none of them exhibited cellulase activity. The pNPCA3, pNPCA4, and pNPCA5 constructs were designed as nonfusion proteins to avoid any possible effect of a protein fusion on the truncated CELA activity.

Enzymic properties. Studies on the substrate specificity and catalytic mode of CELA have shown that this enzyme possesses the properties of a cellobiohydrolase, similar to the properties of *Trichoderma reesei* CBHII (48). The cellulase activity of the cloned CELA was found predominantly in the periplasm (87% of the total cellular activity), indicating that the CELA signal peptide mediated efficient secretion in *E. coli* to the periplasm. The zymogram analysis showed that the apparent molecular mass of CELA produced in *E. coli* was approximately 43 kDa (Fig. 5), which appears to be consistent with the deduced molecular mass of the mature CELA after removal of the proposed signal peptide. This also provides evidence which supports our prediction that CELA in pNPCA (*celA* cDNA in the pBluescript vector) was translated from its native ATG start codon.

In order to simplify purification of the cloned CELA for assessment of the specific activity of the pure enzyme, we constructed a *celA* expression plasmid in which CELA was tagged with a Flag peptide at the N terminus of the mature CELA under the control of the *tac* promoter. The Flag peptide is an antigenic epitope consisting of eight highly hydrophilic amino acid residues which allows affinity purification of tagged recombinant proteins. The Flag-tagged CELA was purified to near homogeneity (>95% pure) by using anti-Flag antibody affinity gel (Fig. 6). The substrate specificity of the Flag-tagged CELA shown in Table 1 appears to be similar to that of the nonfusion CELA (48). Both CELA and the Flag-tagged CELA were found to be more active on lichenan than on amorphous cellulose. The purified Flag-tagged CELA exhibited relatively high specific activity toward crystalline cellulose (Avicel) compared with the specific activity of purified *N. patriciarum* CELD (9.7 U/mg of protein for Flag-tagged CELA, compared with 2.3 U/mg of protein for Flag-tagged CELD domain II [unpublished CELD data]). In order to assess whether the Flag peptide tagging affected the specific activity of CELA, Flag-tagged CELA was incubated with enterokinase, which cleaves Flag from CELA. No significant changes in the specific activity of the enzyme during hydrolysis of CMC were observed after digestion with enterokinase (data not shown).

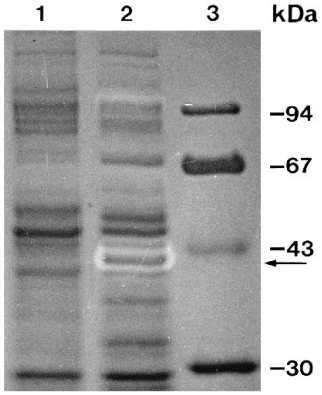

FIG. 5. Zymogram of cloned CELA analyzed on an SDS–10% polyacrylamide PAGE gel containing lichenan. Lane 1, total cellular protein from *E. coli* containing pBluescript SKII; lane 2, total cellular protein from *E. coli* harboring pNPCA; lane 3, molecular mass standards. The position of the lichenan hydrolysis band is indicated by an arrow. The weak band above the 94-kDa band was due to some residual nondenatured CELA, as denaturation was carried out at a low temperature (50°C).

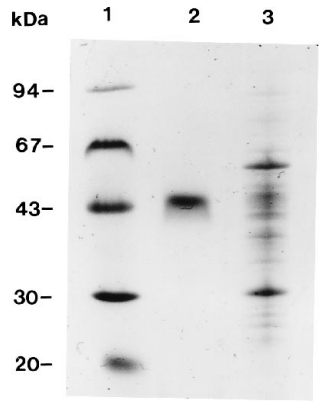

FIG. 6. SDS-PAGE analysis of the Flag-tagged CELA produced by *E. coli* harboring plasmid pBSFCA. Lane 1, molecular mass standards; lane 2, antibody affinity-purified CELA; lane 3, crude CELA prepared as total cell lysate from *E. coli* harboring pBSFCA.

The pH profile of CELA was examined at pH values between 4 and 9.5. CELA had an acid optimal pH (pH around 5.0), and there was substantial activity (>40% of the optimum activity) at pH values between 4.5 and 6.5. The temperature profile revealed that CELA was most active at 40°C. CELA was almost inactive at a hydrolysis temperature of 60°C.

To assess the ability of CELA to hydrolyze the crystalline regions of cellulose, a prolonged hydrolysis assay was performed. Hydrolysis of Avicel by CELA for 24 h resulted in 9.9% ± 0.1% hydrolysis (mean ± standard deviation; $n = 3$). The main hydrolysis product was cellobiose (data not shown), which is in agreement with our previous results (48).

DISCUSSION

It has been shown that the amino acid sequences of cellulolytic and xylanolytic enzymes (CELB, XYNA, and XYNB) from *N. patriciarum* exhibit similarity to the sequences of enzymes isolated from rumen bacteria (4, 16, 51), which suggests that horizontal transfer of genes may have occurred during evolution (15). Conditions within the rumen are potentially favorable for gene transfer because of the large populations of various microorganisms that live in close proximity to each other. This is the first report of a rumen anaerobic fungal cellulase that exhibits substantial homology with an aerobic fungal enzyme. As anaerobic rumen fungi and aerobic microorganisms live in separate ecological niches, this finding strongly suggests that a common ancestral precursor of cellulolytic aerobic fungi and rumen anaerobic fungi may have existed.

Primary structure and function analysis revealed that CELA has two distinct domains, a CBD and a catalytic domain. This is in contrast to the other *N. patriciarum* cellulases that have been characterized, CELB (51) and CELD (17a), in which there are no CBDs. The two domains of CELA are connected by a linker sequence. Unlike the usual Ser/Thr/Pro-rich linker found in most other cellulases, the CELA linker sequence is extremely Asn rich. Of the 44 amino acid residues in the linker, 33 (75%) are Asn. To date, the only other glycosyl hydrolase that contains an Asn-rich linker sequence is a xylanase from *Ruminococcus flavefaciens*, which has a linker consisting of 45% Asn (50). *Trichoderma reesei* CBHII has an abundance of serine and threonine residues in its linker region; both of these compounds are polar, uncharged, hydrophilic amino acids with a hydroxyl group. Asn is also a polar, uncharged, hydrophilic amino acid, but it contains an amide group.

It has been shown that the Ser/Thr-rich linkers of *Trichoderma reesei* cellulases are heavily O glycosylated (13, 42). Glycosylation may increase the resistance of a linker to proteolytic cleavage. The CELA linker is unlikely to be heavily glycosylated, as there are only a few Ser residues present in the linker region. One potential N-glycosylation site (Asn-45) is present in the linker (Fig. 2), as predicted from the tripeptide sequence (Asn-Xaa-Ser/Thr, where Xaa is any residue). There are six more potential N-glycosylation sites present in CELA, but all are located in the catalytic domain. Many glycoproteins require glycosylation for their biological function to occur. It has been reported that screening for functional expression of *Trichoderma reesei* cellulases in *E. coli*, which lacks the necessary glycosylation mechanism, was unsuccessful (37). In contrast, an active form of the cloned CELA is produced in *E. coli*, indicating that glycosylation is not absolutely required for CELA activity. It remains to be seen whether glycosylation affects the specific activity or stability of CELA.

It has been shown by three-dimensional structure analysis that three tyrosine residues present on the flat face of the wedge-shaped CBD of *Trichoderma reesei* CBHI are important for binding to cellulose (7). Site-directed mutagenesis of the Tyr-492 residue of CBHI leads to a reduction in the binding capacity equivalent to the binding capacity of the core enzyme (25). Unlike most members of the type 2 family, CELA contains only one Tyr residue, but still exhibits strong cellulose-binding capacity (48). The other Tyr residues thought to be directly involved in binding of *Trichoderma reesei* CBHI are replaced with Trp residues in CELA. These Trp residues in CELA may play a role in cellulose binding, as Saloheimo et al. (28) have suggested that both Tyr and Trp residues interact readily with carbohydrates. This type of amino acid substitution is also seen in other fungal CBDs (Fig. 4). Another Tyr residue (Tyr-474) in the CBD of *Trichoderma reesei* CBHI, which is not located on the flat face of the CBD, is replaced with a Phe residue in CELA. It has been proposed that this Tyr residue has an important structural role and that it is involved in forming a hydrogen bond with the conserved Gly (Gly-476) (21). The conservation of an aromatic residue at this position in all fungal CBDs (Fig. 4) appears to support this hypothesis. Two other conserved residues (Asn-490 and Gln-495) found on the flat face of the CBHI CBD are also conserved in CELA. It has been suggested that these two residues are essential for tight binding of the CBD to cellulose (23).

TABLE 1. Specific activity of Flag-tagged CELA expressed in *E. coli*

Substrate	Sp act (U/mg of protein)[a]	
	Crude CELA[b]	Purified CELA[c]
Avicel	1.7	9.7
CMC	3.1	15.5
Amorphous cellulose	7.72	42.82
Lichenan	66.7	343.5
Xylan	0	0
p-Nitrophenyl cellobioside	0	0

[a] Cellulase activity was measured at 40°C in 50 mM sodium citrate (pH 6) for 30 to 60 min; the values are the means of the values obtained in at least two separate assays.

[b] Crude CELA was prepared as the total cell lysate from *E. coli* XL1-Blue harboring the cellulase expression plasmid (pBSFCA).

[c] Flag-tagged CELA was purified from the clone pBSFCA by using an immunoaffinity column (M1 anti-Flag antibody affinity gel). The purity of the enzyme was >95%, as judged on an SDS-PAGE gel (Fig. 6).

The removal of the CBD of CELA (the pNPCA2 construct) markedly reduced the enzyme activity toward crystalline cellulose and the cellulose-binding capacity, indicating that the role of the CBD in hydrolysis of crystalline cellulose is important. These results are consistent with observations made with *Trichoderma reesei* cellobiohydrolases. Tomme et al. (42) have shown that the intact cellobiohydrolases and the isolated core catalytic domains have similar activities toward a soluble substrate, but that the activity against crystalline cellulose and the cellulose-binding capacity of the core enzymes are severely impaired. It has been proposed that the CBD of *Trichoderma reesei* cellobiohydrolases either acts as an anchor which raises the effective substrate concentration or plays an active role in breaking the crystalline region of cellulose and liberating cellulose chains (25).

The results of the deletion analysis of the catalytic region sequence of the *celA* cDNA appear to support the assignment of the CELA catalytic domain based on sequence alignment with *Trichoderma reesei* CBHII. The loss of activity in the pNPCA5 construct may be explained by the deletion of a putative catalytic residue (Asp-365) which corresponds to Asp-392 in *C. fimi* CenA (10). It has been demonstrated with site-directed mutagenesis that Asp-392 in CenA acts as a base catalyst and that replacement of Asp-392 with an Ala residue results in a 34,000-fold decrease in CMC-hydrolyzing activity (10). The amino acid residues that were removed from the pNPCA4 enzyme do not appear to be directly involved in the catalytic activity of the family B cellulases analyzed to date (10, 27). However, the Cys-377 residue in CELA is a highly conserved residue in family B cellulases (Fig. 3). The secondary-structure analysis of *Trichoderma reesei* CBHII showed that the corresponding Cys residue (Cys-415) forms a disulfide bridge with the conserved Cys-368 residue. By analogy, it is likely that Cys-377 in CELA forms a disulfide bridge with Cys-330. The disulfide bridge in CBHII has been shown to stabilize loop structures involved in formation of a tunnel which contains the catalytic site (27). Rouvinen et al. (27) also showed that there is a second disulfide bridge in CBHII between Cys-176 and Cys-235. These Cys residues are also found in CELA (Cys-163 and Cys-217) and appear to be present in all family B cellulases, although they are not perfectly aligned (Fig. 3). The loss of activity in the pNPCA3 construction suggests that the deleted amino residues in the N-terminal region of the CELA catalytic domain may also be essential for enzyme activity. It has been shown that the Trp-135 residue in *Trichoderma reesei* CBHII is located in the active site of the enzyme and is involved in substrate binding (27). This Trp residue is strictly conserved in the catalytic domain of all family B cellulases (Fig. 3) and was deleted in the pNPCA3 construct. It should be noted that some amino residues at the terminal regions of the catalytic domain may be essential for maintaining the correct three-dimensional structure of the enzyme active site, although they are not directly involved in binding or catalytic activity. Therefore, deletion analysis defines only the essential sequence region required for enzyme activity.

The Flag peptide tagging used in this study provided a simple means for purifying the recombinant CELA. It appears that CELA activity was not affected by fusion to this small peptide (length, 8 amino acids). An analysis of the specific activity of the purified Flag-tagged enzyme showed that CELA appeared to be a relatively active enzyme toward Avicel when the data were compared with the limited previously published data on the specific activities of native cellobiohydrolases purified from the highly cellulolytic fungus *Trichoderma reesei*. For example, CBHI and CBHII purified from *Trichoderma reesei* exhibited specific activities on Avicel of 0.014 and 0.027 U/mg of pure enzyme, respectively, as reported by Tomme et al. (42), and 0.26 and 0.48 U/mg of pure enzyme, respectively, as reported by El-Gogary et al. (12). The pH and temperature profiles of CELA basically match the physiological conditions of the rumen, where the enzyme may be expected to work at a relatively efficient rate.

The results of this study demonstrated that the strong cellulolytic capacity of rumen fungi can be at least partially attributed to the high specific activities of some of the cellulases that they produce. Although rumen fungi are highly cellulolytic, their contribution to the total cellulolytic capacity of the rumen is generally considered to be relatively low. This is because the fungal populations in the rumen are very small compared with the populations of the rumen bacteria. Recent advances in the development of transformation systems for rumen bacteria (1, 8, 40, 43) have made it possible to modify rumen microorganisms for improvement of plant fiber degradation in the rumen. In view of its high specific activity against crystalline cellulose and enzyme activity with no absolute requirement for glycosylation, the *celA* cDNA is currently being explored as a possible candidate for introduction into rumen bacteria to enhance their cellulolytic capacity in the rumen.

ACKNOWLEDGMENTS

This work was supported by a grant from the Australian Meat Research Corporation, and S.D. is a recipient of a Meat Research Corporation Australasian Postgraduate Research Fellowship.

We thank Kari Gobius, Brian Dalrymple, and Jim Aylward for their helpful discussions.

REFERENCES

1. **Beard, C. E., M. A. Hefford, R. J. Forster, S. Sontakke, R. M. Teather, and K. Gregg.** 1995. A stable and efficient transformation system for *Butyrivibrio fibrisolvens* OB156. Curr. Microbiol. **30:**105–109.
2. **Béguin, P.** 1990. Molecular biology of cellulose degradation. Annu. Rev. Microbiol. **44:**219–248.
3. **Billon-Grand, G., J. B. Fiol, A. Breton, A. Bruyere, and Z. Oulhaj.** 1991. DNA of some anaerobic rumen fungi G+C content determination. FEMS Microbiol. Lett. **82:**267–270.
4. **Black, G. W., G. P. Hazelwood, 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.
5. **Brownlee, A. G.** 1989. Remarkably AT-rich genomic DNA from the anerobic fungus *Neocallimastix*. Nucleic Acids Res. **35:**1327–1335.
6. **Cheng, C., N. Tsukagoshi, and S. Udaka.** 1990. Nucleotide sequence of the cellobiohydrolase gene from *Trichoderma viride*. Nucleic Acids Res. **18:**5559.
7. **Claeyssens, M., and P. Tomme.** 1989. Structure-function relationships of cellulolytic proteins from *Trichoderma reesei*, p. 1–11. *In* C. P. Kubicek et al. (ed.), *Trichoderma reesei* cellulases: biochemistry, genetics, physiology and applications. Royal Society of Chemistry, Cambridge.
8. **Cocconcelli, P. S., E. Ferrari, F. Rossi, and V. Bottazzi.** 1992. Plasmid transformation of *Ruminococcus albus* by means of high-voltage electroporation. FEMS Microbiol. Lett. **94:**203–208.
9. **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.
10. **Damude, H. G., S. G. Withers, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren.** 1995. Site directed mutation of the putative catalytic residues of endoglucanase CenA from *Cellulomonas fimi*. Biochemistry **34:**2220–2224.
11. **de Oliviera, M., and A. Radford.** 1990. Sequence of *cbh-1* gene of *Humicola grisea* var. *thermoidea*. Nucleic Acids Res. **18:**668.
12. **El-Gogary, S., A. Leite, O. Crivellaro, D. E. Everleigh, and H. El-Dorry.** 1989. Mechanism by which cellulose triggers cellobiohydrolase I gene expression in *Trichoderma reesei*. Proc. Natl. Acad. Sci. USA **86:**6138–6141.
13. **Fägerstam, L. G., L. G. Pettersson, and J.-A. Engström.** 1984. The primary structure of a 1,4-β-glucan cellobiohydrolase from the fungus *Trichoderma reesei* QM9414. FEBS Lett. **167:**309–315.
14. **Fernández-Abalos, J. M., P. Sánchez, P. M. Coll, J. R. Villanueva, P. Pérez, and R. I. Santamaría.** 1992. Cloning and nucleotide sequence of $celA_1$, an endo-β-1,4-glucanase-encoding gene from *Streptomyces halstedii* JM8. J. Bacteriol. **174:**6368–6376.
15. **Flint, H. J.** 1994. Molecular genetics of obligate anaerobes from the rumen.

FEMS Lett. **121:**259–268.
16. **Gilbert, H. J., J. 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.
17. **Gilkes, N. R., B. Henrissat, D. G. Kilburn, R. C. Miller, and R. A. J. Warren.** 1991. Domains in microbial β-1,4-glycanases: sequence conservation, function, and enzyme families. Microbiol. Rev. **55:**303–315.
17a. **Grobius, K. S., and G. P. Xue.** Unpublished data.
18. **Henrissat, B.** 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. **280:**309–316.
19. **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.
20. **Koch, A., C. T. O. Weigel, and G. Schulz.** 1993. Cloning, sequencing, and heterologous expression of a cellulase-encoding cDNA (*cbh1*) from *Penicillum janthinellum*. Gene **124:**57–65.
21. **Kraulis, P. J., G. M. Clore, M. Nilges, T. A. Jones, G. Pettersson, J. Knowles, and A. M. Gronenborn.** 1989. Determination of the three-dimensional structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. Biochemistry **28:**7241–7257.
22. **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.
23. **Linder, M., M. Mattinen, M. Kontteli, G. Lindeberg, J. Stahlberg, T. Drakenberg, T. Reinikainen, G. Pettersson, and A. Annila.** 1995. Identification of functionally important amino acids in the cellulose-binding domain of *Trichoderma reesei* cellobiohydrolase I. Protein Sci. **4:**1056–1064.
24. **Penttilä, M., P. Lehotovaara, H. Nevalainen, R. Bhikhabhai, and J. Knowles.** 1986. Homology between cellulase genes of *Trichoderma reesei*: complete nucleotide sequence of the endoglucanase I gene. Gene **45:**253–263.
25. **Reinikainen, T., L. Ruhonen, T. Nevanen, L. Laaksonen, P. Kraulis, T. A. Jones, J. K. C. Knowles, and T. T. Teeri.** 1992. Investigations of the function of mutated cellulose-binding domains of *Trichoderma reesei* cellobiohydrolase I. Protein Struct. Funct. Genet. **14:**475–482.
26. **Roufa, D. D., and D. D. Rhoads.** 1991. SEQAID version 3.81. Division of Biology, Kansas State University, Manhattan.
27. **Rouvinen, J., T. Bergforgs, J. Teeri, K. C. Knowles, and T. A. Jones.** 1990. Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. Science **249:**380–386.
28. **Saloheimo, A., B. Henrissat, A. M. Hoffrén, O. Teleman, and M. Penttilä.** 1994. A novel small endoglucanase gene, *egl5*, from *Trichoderma reesei* isolated by expression in yeast. Mol. Microbiol. **13:**219–228.
29. **Saloheimo, M., P. Lehotovaara, M. Penttilä, T. T. Teeri, J. Ståhlberg, G. Johansson, G. Pettersson, M. Claeyssens, P. Tomme, and J. K. C. Knowles.** 1988. EGIII, a new endoglucanase from *Trichoderma reesei*: the characterisation of both gene and enzyme. Gene **63:**11–21.
30. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **Schwarz, W. H., K. Bronnenmeier, F. Grabnitz, and W. L. Staudenbauer.** 1987. Activity staining of cellulases in polyacrylamide gels containing mixed linkage β-glucans. Anal. Biochem. **164:**50–56.
32. **Shoemaker, S., V. Schweickart, M. Lander, D. Gelfend, S. Kwok, K. Myambo, and M. Innis.** 1983. Molecular cloning of Exo-cellobiohydrolase I derived from *Trichoderma reesei* strain L27. Bio/Technology **1:**691–696.
33. **Sims, P., C. James, and P. Broda.** 1988. The identification, molecular cloning and characterisation of a gene from *Phanerochaete chrysosporium* that shows strong homology to the exo-cellobiohydrolase I gene from *Trichoderma reesei*. Gene **74:**411–422.
34. **Sturrock, S. S., and J. F. Collins.** 1993. MPsrch version 1.3. Biocomputer Research Unit, University of Edinburgh, Edinburgh.
35. **Taleb, F., and A. Radford.** 1995. The cellulase complex of *Neurospora crassa*: *cbh-1* cloning, sequencing and homologies. Gene **161:**137–138.
36. **Tamblyn Lee, J. M., Y. Hu, H. Zhu, K. J. Cheng, P. J. Krell, and C. W. Forsberg.** 1993. Cloning of a xylanase gene from the ruminal fungus *Neocallimastix patriciarum* 27 and its expression in *E. coli*. Can. J. Microbiol. **39:**134–139.
37. **Teeri, T. T.** 1987. Ph.D. thesis. University of Helsinki, Helsinki, Finland.
38. **Teeri, T. T., P. Lehtovaara, S. Kauppinen, I. Salovuori, and J. Knowles.** 1987. Homologous domains in *Trichoderma reesei* cellulolytic enzymes: gene sequence and expression of cellobiohydrolase II. Gene **51:**43–52.
39. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
40. **Thomson, A. M., and H. J. Flint.** 1989. Electroporation induced transformation of *Bacteroides ruminicola* and *Bacteroides uniformis* by plasmid DNA. FEMS. Microbiol. Lett. **61:**101–104.
41. **Titus, D. E. (ed.).** 1991. Promega protocols and applications guide, 2nd ed. Promega, Madison, Wis.
42. **Tomme, P., H. Van Tilberugh, G. Petterson, J. Van Damme, J. Vandekerckhove, J. Knowles, T. Teeri, and M. Claeyssens.** 1988. Studies of the celluloytic system of *Trichoderma reesei* QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. Eur. J. Biochem. **170:**575–581.
43. **Whitehead, T. R.** 1992. Genetic transformation of the ruminal bacteria *Butyrivibrio fibrisolvens* and *Streptococcus bovis* by electroporation. Lett. Appl. Microbiol. **15:**186–189.
44. **Williams, A. G., and C. G. Orpin.** 1987. Polysaccharide-degrading enzymes formed by three species of anaerobic rumen fungi grown on a range of carbohydrate substrates. Can. J. Microbiol. **33:**418–426.
45. **Wilson, C. A., and T. M. Wood.** 1992. Studies on the cellulase of the rumen anaerobic fungus *Neocallimastix frontalis*, with special reference to the capacity of the enzyme to degrade crystalline cellulose. Enzyme Microb. Technol. **14:**258–264.
46. **Wong, W. K. R., B. Gerhard, Z. M. Guo, D. G. Kilburn, R. A. J. Warren, and R. C. Miller, Jr.** 1986. Characterisation and structure of an endoglucanase gene cenA of *Cellulomonas fimi*. Gene **44:**315–324.
47. **Xue, G. P., K. S. Gobius, and C. G. Orpin.** 1992. A novel polysaccharide hydrolase cDNA (*celD*) from *Neocallimastix patriciarum* encoding three multi-functional catalytic domains with high endoglucanase, cellobiohydrolase and xylanase activities. J. Gen. Microbiol. **138:**2397–2403.
48. **Xue, G. P., C. G. Orpin, K. S. Gobius, J. H. Aylward, and G. D. Simpson.** 1992. Cloning and expression of multiple cellulase cDNAs from the anaerobic rumen fungus *Neocallimastix patriciarum* in *Escherichia coli*. J. Gen. Microbiol. **138:**1413–1420.
49. **Yablonsky, M. C., T. Bartley, K. O. Elliston, S. K. Kahrs, Z. P. Shalita, and D. E. Eveleigh.** 1988. Characterisation and cloning of the cellulase complex of *Microbispora bispora*. FEMS Symp. **43:**249–266.
50. **Zhang, J. X., and H. J. Flint.** 1992. A bifunctional xylanase encoded by the *xynA* gene of the rumen cellulolytic bacterium *Ruminococcus flavefaciens* 17 comprises two dissimilar domains linked by an asparagine/glutamine-rich sequence. Mol. Microbiol. **6:**1013–1023.
51. **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.