**Microbiology** (2000), **146**, 1999–2008 Printed in Great Britain

# Primary sequence and enzymic properties of two modular endoglucanases, Cel5A and Cel45A, from the anaerobic fungus *Piromyces equi*

Ruth Y. Eberhardt,[1] Harry J. Gilbert[2] and Geoffrey P. Hazlewood[1]†

Author for correspondence: Ruth Y. Eberhardt. Tel: +44 1223 496478. Fax: +44 1223 496023.
e-mail: ruth.eberhardt@bbsrc.ac.uk

[1] Laboratory of Molecular Enzymology, The Babraham Institute, Babraham, Cambridge CB2 4AT, UK

[2] Department of Biological and Nutritional Sciences, The University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK

**Two endoglucanase cDNAs, designated *cel5A* and *cel45A*, were isolated from a cDNA library of the anaerobic fungus *Piromyces equi*. Sequence analysis revealed that *cel5A* has an open reading frame of 5142 bp and encodes a 1714 amino acid modular enzyme, Cel5A, with a molecular mass of 194847 Da. Cel5A consists of four catalytic domains homologous to family-5 glycosyl hydrolases, two C-terminal dockerins and one N-terminal dockerin. This is the first report of a complete gene containing tandem repeats of family-5 catalytic domains. The cDNA *cel45A* has an open reading frame of 1233 bp and encodes a 410 amino acid modular enzyme, Cel45A, with a molecular mass of 44380 Da. The catalytic domain, located at the C terminus, is homologous to the family-45 glycosyl hydrolases. Cel45A is the first family-45 enzyme to be described in an anaerobe. The presence of dockerins at the N and C termini of Cel5A and at the N terminus of Cel45A implies that both enzymes are part of the high-molecular-mass cellulose-degrading complex produced by *Piromyces equi*. The catalytic domain nearest the C terminus of Cel5A and the catalytic domain of Cel45A were hyperexpressed as thioredoxin fusion proteins, Trx-Cel5A′ and Trx-Cel45A′, and subjected to biochemical analysis. Trx-Cel5A′ has a broad substrate range, showing activity against carboxymethylcellulose, acid-swollen cellulose, barley β-glucan, lichenin, carob galactomannan, *p*-nitrophenyl β-D-cellobiopyranoside and xylan. Trx-Cel45A′ is active against carboxymethylcellulose, acid-swollen cellulose and the mixed linkage glucans, barley β-glucan and lichenin.**

**Keywords:** anaerobic fungi, cellulase, glycosyl hydrolases (families 5 and 45), endoglucanases, *Piromyces equi*

## INTRODUCTION

Cellulose, the most abundant molecule in the biosphere, is a polymer of β-1,4-linked glucose molecules. Chains of cellulose form insoluble fibrils, which are maintained by inter- and intramolecular hydrogen bonds and van der Waals interactions (Béguin & Lemaire, 1996). Three classes of enzymes are required for the complete hydrolysis of cellulose: endoglucanases (EC 3.2.1.4), which cleave β-1,4-glycosidic bonds randomly within the cellulose chain; cellobiohydrolases (EC 3.2.1.91), which cleave cellobiose from the ends of the cellulose chain; and β-glucosidases (EC 3.2.1.21), which convert cellobiose and other low molecular mass cellodextrins into glucose. Synergism between these three types of enzyme has been widely reported (Béguin, 1990; Leschine, 1995).

The sequences of over 1500 glycosyl hydrolases are known and have been assigned to at least 77 families on the basis of hydrophobic cluster analysis and amino acid sequence similarities. Twelve of these families contain endoglucanases (Henrissat, 1991, 1998; Henrissat &

---

†**Present address:** Finnfeeds International, PO Box 777, Marlborough SN8 1XN, UK.

**Abbreviations:** CMC, carboxymethylcellulose; CMCase, carboxymethylcellulase; DNSA, dinitrosalicylic acid reagent; pNPC, *p*-nitrophenyl β-D-cellobiopyranoside.

The EMBL accession numbers for the sequences reported in this paper are AJ277482 and AJ277483.

Bairoch, 1993, 1996). Many glycosyl hydrolases are modular enzymes consisting of one or more catalytic domains joined to non-catalytic domains via linker sequences rich in proline and hydroxy amino acids (Gilkes *et al*., 1991; Tomme *et al*., 1995; Warren, 1996).

Herbivores do not produce polysaccharide hydrolases, so rely on symbiotic relationships with micro-organisms (mostly bacteria and fungi), which inhabit their gastrointestinal tracts, to digest plant material. Anaerobic fungi were first isolated by Orpin (1975), from the rumen of a sheep. Since then they have been recovered from the digestive tracts of many different species of herbivores, including both ruminants and non-ruminants, where they are believed to be responsible for the digestion of 40–70% of the ingested plant material (Li & Heath, 1993; Trinci *et al*., 1994). At least five genera of anaerobic fungi have been described to date including *Neocallimastix*, *Orpinomyces*, *Piromyces*, *Caecomyces* and *Anaeromyces*. The polysaccharide hydrolases produced by anaerobic fungi are amongst the most active that have been described to date; they are capable of degrading a wide range of polysaccharides and can completely solubilize both amorphous and highly crystalline cellulose (Li & Heath, 1993; Selinger *et al*., 1996; Wubah *et al*., 1993). Unlike aerobic fungi, anaerobic fungi produce large, multienzyme cellulase–hemicellulase complexes, similar to the cellulosome of *Clostridium thermocellum* (Ali *et al*., 1995; Dijkerman *et al*., 1997; Fanutti *et al*., 1995; Hazlewood & Gilbert, 1998; Teunissen *et al*., 1993; Wilson & Wood, 1992).

The anaerobic fungus *Piromyces equi* was isolated from the caecum of a pony (Munn, 1994; Orpin, 1981). The cellulose–hemicellulose degrading system of *P. equi* consists of a large multienzyme complex, which accounts for up to 90% of the cellulase, mannanase and xylanase activities produced by the fungus. This complex consists of at least ten polypeptides ranging from 50 to 190 kDa, including a 97 kDa putative scaffolding protein (Ali *et al*., 1995; Fanutti *et al*., 1995; Hazlewood & Gilbert, 1998). Catalytic components of the complex are all modular enzymes and contain one, two or three copies of a highly conserved 40-residue non-catalytic sequence that is believed to function as a dockerin domain (Fanutti *et al*., 1995). These putative dockerin domains are located at either the N terminus, the C terminus or between catalytic domains of *P. equi* enzymes (Hazlewood & Gilbert, 1998).

Although genes encoding xylanases and mannanases from *P. equi* have been reported (Fanutti *et al*., 1995; Millward-Sadler *et al*., 1996), no endoglucanase genes have been described to date. In this study we aim to characterize the endoglucanase components of the *P. equi* multienzyme complex. Here we describe the elucidation of the primary structures and enzymic properties of two endoglucanases from *P. equi*, Cel5A and Cel45A. Cel5A has four similar family-5 catalytic domains and shows cellulase, xylanase and mannanase activities. Cel5A is the first anaerobic fungal enzyme to be reported which has both N-terminal and the C-terminal dockerin domains within the same polypeptide. Cel45A is the first family-45 endoglucanase to be isolated from an anaerobic organism.

## METHODS

**Microbial strains, vectors and culture conditions.** *Escherichia coli* strains SURE 2 (Stratagene), XL1-Blue and BL21(DE3) were cultured in Luria broth (LB) at 37 °C; the medium was supplemented with ampicillin (50 $\mu g\ ml^{-1}$) to select against non-recombinants. To induce expression of genes under the control of the *T7lac* promoter, *E. coli* strains were grown to mid-exponential phase, 0·5 mM IPTG was added, and the bacteria cultured for a further 3 h. Vectors used in this study were λZAPII (Stratagene), pBluescript SK− (Stratagene) and pET32a (Novagen). *P. equi* was cultured under anaerobic conditions at 37 °C in a medium containing rumen fluid (Kemp *et al*., 1984) and supplemented with Avicel (0·5%) and soluble xylan (0·1%).

**General recombinant DNA methodology.** Agarose-gel electrophoresis, Southern hybridization, Northern hybridization and the transformation of *E. coli* were performed as described by Sambrook *et al.* (1989). Plasmid DNA was prepared by an adaptation of the alkaline lysis method of Birnboim & Doly (1979) or by using Qiagen resin columns (Qiagen), following the protocol recommended by the manufacturer. Restriction endonucleases and other DNA-modifying enzymes were used in accordance with the manufacturer's instructions. Total RNA was extracted from a 2 l culture of *P. equi* by grinding fungal mycelia to a powder in liquid nitrogen, resuspending in GTC [4 M guanidine thiocyanate, 25 mM sodium citrate, 0·5 mM EDTA pH 7·5, 0·5% (w/v) sodium lauryl sarcosine] and homogenizing. The resulting solution was centrifuged at 2000 ***g*** for 10 min. CsCl was added to the supernatant to a final concentration of 0·1 $g\ ml^{-1}$, transferred to tubes containing 5·7 M CsCl, 0·1 M EDTA and centrifuged at 80000 ***g*** for 24 h. The supernatant was discarded and the RNA pellet resuspended in TE (10 mM Tris, 1 mM EDTA) pH 7·6, 0·1% SDS. The RNA was further purified by chloroform/n-butanol extraction and ethanol precipitation, before resuspension in 1 mM EDTA pH 7·6. A cDNA library was constructed in λZAPII from *P. equi* mRNA as described previously (Ali *et al*., 1995) and screened for carboxymethylcellulase (CMCase) activity by plating onto NZY medium in a soft agar overlay containing 0·5 mM IPTG. After 16 h incubation at 37 °C, the plaques which had formed were overlaid with molten agar containing 0·5% carboxymethylcellulose (CMC), 50 mM $K_2HPO_4$ and 12 mM citric acid pH 6·5. After a further 4 h incubation at 37 °C, plates were stained with Congo red and destained with 1 M NaCl (Teather & Wood, 1982) to identify positive clones. Clones expressing CMCase activity were excised and recircularized to form pBluescript phagemids.

**Nucleotide sequencing.** Nucleotide sequencing was performed on plasmid DNA using a PRISM ready reaction dyedeoxy terminator cycle sequencing kit (Applied Biosystems). To sequence *cel5A*, nested deletions were generated with exonuclease III using the double-stranded nested deletion kit (Amersham-Pharmacia). *cel45A* was sequenced by chromosome walking: synthetic oligonucleotide primers were designed according to known sequence. Sequences were compiled and ordered using software written by the Genetics Computer Group at the University of Wisconsin.

**Construction of pRE1 and pRE2.** The primers 5′-CGGGATCCGAACCAACTGGTAATATTCGTG-3′ (primer A1) and

5′-ACTACCGCTCGAGTTATGGTTCTTCTTCAATAG-3′ (primer A2) were used to amplify the region of *cel5A* encoding the catalytic domain nearest the C terminus (between nucleotides 3790 and 4884) to construct pRE1. To construct pRE2, the region of *cel45A* encoding the catalytic domain (between nucleotides 535 and 1260) was amplified by PCR using the primers 5′-CGGGATCCACTGGTTCGGGTGGTAACAGC-3′ (primer B1) and 5′-GCCGCCGCTCGAGAGATAGGGATATACTGG-3′ (primer B2). The reaction conditions for PCR were: denaturation 30 s at 94 °C, annealing 45 s at 50 °C, elongation 75 s at 72 °C, for 25 cycles. The amplified DNA sequences were digested with *Bam*HI and *Xho*I and ligated into the thioredoxin fusion expression vector, pET32a, digested with the same restriction enzymes to form pRE1 and pRE2, encoding Trx-Cel5A′ and Trx-Cel45A′, respectively.

**Purification of fusion proteins.** *E. coli* strains harbouring the plasmids pRE1 and pRE2 were grown to mid-exponential phase in 3 l (pRE1) or 6·7 l (pRE2) LB medium containing 50 µg ampicillin $ml^{-1}$ at 37 °C. IPTG was added to each culture to a concentration of 0·5 mM and cultures were grown for a further 3 h at 37 °C (pRE1) or 30 °C (pRE2). Cells were pelleted by centrifugation at 10000 ***g*** for 10 min and then resuspended in 75 ml 100 mM NaCl, 20 mM Tris pH 8·0. The cells were then lysed by sonication. Triton X-100 was added to the pRE2 extract to a final concentration of 0·1% (v/v), followed by an incubation at 4 °C for 20 min. Cell debris was removed by centrifugation at 10000 ***g*** for 20 min and the resulting cell-free extract was incubated with 5 ml TALON metal affinity resin (Clontech) for 2 h at 4 °C. Unbound proteins were removed by 5 min centrifugation at 150 ***g***, four washes with 100 mM NaCl, 20 mM Tris/HCl pH 8·0 and two washes with 10 mM imidazole, 100 mM NaCl, 20 mM Tris/HCl pH 8·0. To elute bound proteins, the TALON resin was applied to a column and 5 ml 50 mM imidazole, 100 mM NaCl, 20 mM Tris/HCl pH 8·0 was added. Five 1 ml eluates were collected; elution was repeated with 5 ml 100 mM imidazole, 100 mM NaCl, 20 mM Tris/HCl pH 8·0 to obtain a further five eluates. Eluates of Trx-Cel5A′ and Trx-Cel45A′ were checked by SDS-PAGE as described by Laemmli (1970). The eluates containing fusion protein were pooled and dialysed twice against 10 mM Tris pH 8·0 to remove imidazole, Trx-Cel45A′ was further purified by anion-exchange chromatography using a 2 ml BioQ column (Bio-Rad) on an automated Biologic workstation (Bio-Rad) and eluted in 10 mM Tris/HCl pH 8·0 with a gradient of 0–500 mM NaCl.

**Assays.** Protein concentrations of Trx-Cel5A′ and Trx-Cel45A′ were determined by dye binding (Bradford, 1976). Enzyme activity was assayed in 50 mM $K_2HPO_4$, 12 mM citric acid (PC buffer) pH 6·5 in the presence of 1 mg BSA at 37 °C (Trx-Cel5A′) or 60 °C (Trx-Cel45A′), unless otherwise stated. Reducing sugar release was measured with the dinitrosalicylic acid reagent (DNSA) (Miller, 1959). All assays were performed for 5, 10, 15, 30 and 60 min to determine the initial rate of activity. Substrates tested were CMC (BDH), acid-swollen cellulose (Sigmacell, treated according to Wood, 1971), bacterial microcrystalline cellulose [harvested from *Acetobacter hansenii*, washed with 4% (v/v) NaOH, neutralized with HCl and washed in distilled $H_2O$], Avicel (Merck), soluble birchwood xylan, barley *β*-glucan (Megazyme), lichenin (*Cetraria islandica*; Sigma), laminarin (*Laminaria digitata*; Sigma), carob galactomannan (Megazyme), debranched arabinan (Megazyme) and potato galactan (Megazyme). Concentrations of substrate used ranged from 0·1% to 1·125% w/v. Cellobiohydrolase activity was measured using 100 µl 1 mM *p*-nitrophenyl *β*-D-cellobiopyranoside (pNPC) (Sigma) in a 1 ml reaction volume, stopping the reaction with 500 µl 1 M $Na_2CO_3$ and measuring *p*-nitrophenyl release spectrophotometrically at 400 nm. Assays to determine pH optima were performed in citric acid, $Na_2HPO_4$ buffer (pH 4·0–7·6) or sodium barbitone, HCl buffer (pH 7·9–8·1). Temperature inactivation studies were carried out by preincubating enzyme and buffer at a given temperature for 0, 2, 5, 10, 15, 20, 30, 45 and 60 min, before assaying for activity against CMC as described above. One unit of enzyme activity was defined as the quantity of enzyme required to liberate 1 µmol product in 1 min.

## RESULTS AND DISCUSSION

### Isolation of *cel5A* and *cel45A*

A cDNA library in λZAPII, consisting of $1·8 \times 10^7$ clones, was constructed as described previously using mRNA isolated from *Pir. equi* mycelia cultured in medium containing soluble xylan and Avicel (Ali *et al.*, 1995). Twenty-nine clones producing CMCase activity were isolated from this library. Restriction mapping and Southern hybridization indicated that the fungal cDNAs contained in these clones were derived from three different genes. Nucleic acid sequencing of the 5′ and 3′ ends of representatives of each of the three groups revealed sequences characteristic of enzymes in glycosyl hydrolase families 5, 6 and 45. Initial biochemical studies revealed that the family-5 enzyme had a broad substrate range (data not shown), so this enzyme was chosen for further study. The only family-45 enzymes described to date are from aerobic fungi and the aerobic bacterium *Pseudomonas fluorescens* (Davies *et al.*, 1993; Gilbert *et al.*, 1990; Saloheimo *et al.*, 1994; Schauwecker *et al.*, 1995; Schülein, 1997; Sheppard *et al.*, 1994), which makes the *Pir. equi* family-45 enzyme particularly interesting for further study. The longest cDNAs encoding enzymes from families 5 and 45 (pKPC1 and pKPC28, respectively) were estimated by restriction

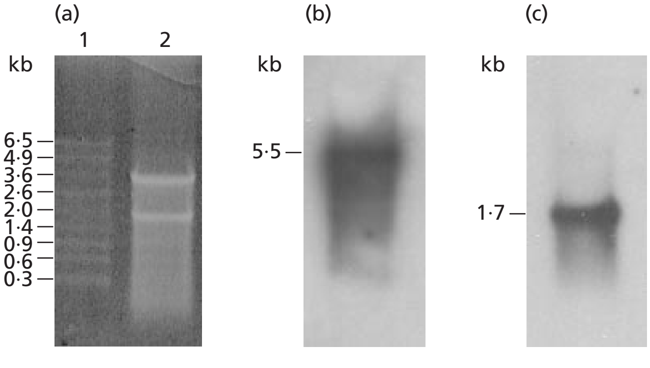

***Fig. 1.*** Northern blot analysis of *cel5A* and *cel45A* expression in *P. equi*. (a) Ethidium bromide stained 1% agarose/2·2 M formaldehyde gel. Lanes: 1, RNA molecular mass markers (Sigma); 2, *P. equi* RNA. (b) Northern blot probed with the amplified region of *cel5A* cDNA encoding a catalytic domain, labelled with [α-$^{32}$P]dCTP. (c) Northern blot probed with the amplified region of *cel45A* cDNA encoding the catalytic domain, labelled with [α-$^{32}$P]dCTP.

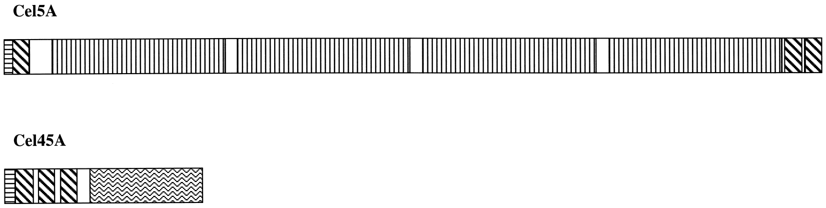

***Fig. 2.*** Schematic representation of the molecular architecture of Cel5A and Cel45A. The locations of the signal peptides (horizontal lines), dockerin domains (diagonal lines), linker sequences (white), family-5 catalytic domains (vertical lines) and family-45 catalytic domains (wavy lines) are shown.

mapping to be 5·7 kb and 1·5 kb, respectively. Northern blot analysis was performed subsequently using nucleic acid probes derived from the two cDNA sequences, in order to deduce the size of the full-length transcripts of these genes. *Pir. equi* produces RNA transcripts of 5·5 kb and 1·7 kb which hybridize to these regions, confirming that the cDNAs contained in pKPC1 and pKPC28 are complete (Fig. 1).

## Nucleotide sequence

A set of nested deletions was created with exonuclease III to facilitate the sequencing of the cDNA derived from pKPC1, designated *cel5A*. The complete nucleotide sequence of the cDNA derived from pKPC28, designated *cel45A*, was determined in both strands by chromosome walking. Translation of the nucleotide sequences of *cel5A* and *cel45A* revealed ORFs of 5142 bp and 1230 bp, respectively, encoding polypeptides of $M_r$ 194848 and $M_r$ 44380 (EMBL accession numbers AJ277482 and AJ277483). The assignment of the proposed translation-initiation codons was based on the following observations: (i) the presence of translational stop codons in all three ORFs upstream of the proposed translation-initiation codon; (ii) the absence of ATG sequences upstream of the ORF; (iii) a very high A+T content preceding the proposed translation-initiation codon, as found previously in cDNAs encoding cellulases, xylanases and mannanases from anaerobic fungi (Fanutti *et al*., 1995; Li *et al*., 1997a). The non-coding regions of *cel5A* and *cel45A* had A+T contents of 90% and 88% respectively, compared with 62% and 57% for the ORFs. The codon utilization of *cel5A* and *cel45A* is very similar to that of other polysaccharide hydrolase genes in *Piromyces*, *Neocallimastix* and *Orpinomyces* (Black *et al*., 1994; Chen *et al*., 1997; Denman *et al*., 1996; Gilbert *et al*., 1992; Li *et al*., 1997a; Millward-Sadler *et al*., 1996; Zhou *et al*., 1994); 16 of the 61 amino acid codons are absent from *cel5A* and 20 are absent from *cel45A*. There is a marked preference for T in the wobble position; 47% of *cel5A* and 51% of *cel45A* codons end in T, whereas only 12% (*cel5A*) and 9% (*cel45A*) end in G.

## Molecular architecture

The deduced primary structures of Cel5A and Cel45A, encoded by *cel5A* and *cel45A*, respectively, revealed that the N-terminal regions conformed to typical signal peptides, which is consistent with the observation that endoglucanases are components of the extracellular cellulase system of *P. equi*. Cel5A has three copies of a highly conserved approximately 40-residue sequence, one between residues 23 and 57, and a further two between residues 1631 and 1708. Cel45A has three repeats of this 40-residue sequence between residues 22 and 152. This highly conserved 40-residue sequence is present in many of the polysaccharide hydrolases isolated from anaerobic fungi and is believed to function as a dockerin domain, which mediates binding of the catalytic subunits of the fungal multienzyme cellulase complex to a scaffolding protein (Dalrymple *et al*., 1997; Fanutti *et al*., 1995; Hazlewood & Gilbert, 1998; Li *et al*., 1997b). Fungal dockerins have been previously found in one, two or three copies at either the N terminus, the C terminus or between two catalytic domains (Fig. 2). Cel5A is the first reported example of an enzyme with dockerins at two separate locations within the molecule. This may have arisen via the fusion of two endoglucanase genes encoding proteins with dockerin domains. Alternatively, the presence of dockerins at both ends of the enzyme could provide a means for linking complexes together to form clusters analogous to the polycellulosome of *C. thermocellum* (Mayer *et al*., 1987; Shoham *et al*., 1999).

Based on sequence homology with known glycosyl hydrolases, Cel5A has four putative catalytic domains, spanning residues 106–462, 495–851, 880–1236 and 1269–1625, which are more than 99% identical to each other on the nucleic acid level. Comparison of the catalytic domains of Cel5A with sequences in the SWISS-PROT database revealed homology with the family-5 endoglucanases (Fig. 3). Aylward *et al*. (1999) recently reported the incomplete sequence of *celD* from *Neocallimastix patriciarum*. The protein encoded by this sequence consists of three family-5 catalytic domains followed by two dockerin domains and shows a high degree of similarity with the C-terminal part of Cel5A. In an earlier study, Xue *et al*. (1992) demonstrated that the full-length *N. patriciarum celD* transcript was 5·5 kb, which is the same length as the *cel5A* transcript produced by *P. equi*. This suggests that the full-length *cel5A* from *P. equi* is of a similar size to the full-length *N. patriciarum celD*. In addition to *N. patriciarum celD*, there are a number of other genes encoding family-5 endoglucanases from anaerobic fungi in the EMBL database including *celB* from *Orpinomyces* sp. PC-2 (accession no. U57818), *Orpinomyces joyonii celA* (accession no. U59432), *celB2* (accession no. AF015249) and *celB29* (accession no. AF015248), *celB* from *N.*

```
P. equi Cel5A 1       IRDISSKELIKEMNFGWNLGNTLDAECTSWMNYEKDPIGSETCWGNPKTTEDMYKILMDN   165
P. equi Cel5A 2       IRDISSKELIKEMNFGWNLGNTLDAECTSWMNYEKDPIGSETCWGNPKTTEDMYKILMDN   554
P. equi Cel5A 3       IRDISSKKLIKEMNFGWNLGNTLDAECTSWMNYEKDPIGSETCWGNPKTTEDMYKILMDN   939
P. equi Cel5A 4       IRDISSKELIKEMNFGWNLGNTLDAECTSWMNYEKDPIGSETCWGNPKTTEDMYKILMDN  1328
C. longisporum CelA   MR..SASEIVQEMGVGWNLGNTLDAKITN.LSYNTSPISFETGWGNPVTTKAMIDKIKNA    92
R. albus EGA          VRDISAMELVGEMKTGWNLGNSLDATGAPGN.......ASEVNWGNPKTTKEMIDAVYNK    74
N. patriciarum CelB   MKNISSKELVKDLTIGWSLGNTLDATCFETLDYNKNQIASETCWGNVKTTQELYYKLSDL    78
C. thermocellum EGE   MRDISAIDLVKEIKIGWNLGNTLDA.............PTETAWGNPRTTKAMIEKVREM    98


P. equi Cel5A 1       QFNVFRIPTTWTGHIGEAPDYKINEKWMKRVHEIVDYPYKNGAFVILNIHHE.SWNHAFE   224
P. equi Cel5A 2       QFNVFRIPTTWTGHIGEAPDYKINEKWMKRVHEIVDYPYKNGAFVILNIHHE.SWNHAFE   613
P. equi Cel5A 3       QFNVFRIPTTWTGHIGEAPDYKINEKWMKRVHEIVDYPYKNGAFVILNIHHE.SWNHAFE   998
P. equi Cel5A 4       QFNVFRIPTTWTGHIGEAPDYKINEKWMKRVHEIVDYPYKNGAFVILNIHHE.SWNHAFE  1387
C. longisporum CelA   GFKTIRIPTTWGEHLDG..NNKLNEEWVKRVKEVVDYCIADDLYVILNTHHEGNWVIPTY   150
R. albus EGA          GFDVIRIPVTWGGHVGDAPDYKIDDEWIARVQEVVNYAYDDGAYVIINSHHEEDWRIPDN   134
N. patriciarum CelB   GFNTFRIPTTWSGHFGNAPDYKINDQWMKRVHEIVDYAINTGGYAILNIHHE.TWNHAFQ   137
C. thermocellum EGE   GFNAVRVPVTWDTHIGPAPDYKIDEAWLNRVEEVVNYVLDCGMYAIINLHHDNTWIIPTY   158


P. equi Cel5A 1       ETVEEAKVELAKVWAQIAEEFKDYDEHLIFEGQNEPRKNDTPVEWNGGDQEGWDVVNAMN   284
P. equi Cel5A 2       ETVEEAKVELAKVWAQIAEEFKDYDEHLIFEGQNEPRKNDTPVEWNGGDQEGWDVVNAMN   673
P. equi Cel5A 3       ETVEEAKVELAKVWAQIAEEFKDYDEHLIFEGQNEPRKNDTPVEWNGGDQEGWDVVNAMN  1058
P. equi Cel5A 4       ETVEEAKVELAKVWAQIAEEFKDYDEHLIFEGQNEPRKNDTPVEWNGGDQEGWDVVNAMN  1447
C. longisporum CelA   AKESSVTPKLKTLWTQISEAFKDYDDHLIFETLNEPRLEGTPYEWTGGTSESRDVVNKYN   210
R. albus EGA          EHIDAVDEKTAAIWKQVAERFKDYGDHLIFEGLNEPRVKGSPQEWNGGTEEGRRCVDRLN   194
N. patriciarum CelB   KNLESAKKILVAIWKQIAAEFADYDEHLIFEGMNEPRKVGDPAEWNGGDYEGWNFVNEMN   197
C. thermocellum EGE   ANEQRSKEKLVKVWEQIATRFKDYDDHLLFETMNEPREVGSPMEWMGGTYENRDVINRFN   218


P. equi Cel5A 1       AVFMKTVRSSGGNNAKRHLMIPPYAAACNKNSFDNFDFPEDDDKVIASVHAYSPYNFALN   344
P. equi Cel5A 2       AVFMKTVRSSGGNNAKRHLMIPPYAAACNQNSFDHFDFPEDDDKVIASVHAYSPYNFALN   733
P. equi Cel5A 3       AVFMKTVRSSGGNNAKRHLMIPPYAAACNKNSFDNFDFPEDDDKVIASVHAYSPYNFALN  1118
P. equi Cel5A 4       AVFMKTVRSSGGNNAKRHLMIPPYAAACNKNSFDNFDFPEDDDKVIASVHAYSPYNFALN  1507
C. longisporum CelA   AAALESIRKTGGNNLSRAVMMPTYAASGSSTTMNDFKVP.DDKNVIASVHAYSPYFFAMD   269
R. albus EGA          KTFLDTVRATGGNNEKRLLLMTTYASSSMSNVIKDTAIPEDD.HIGFSIHAYTPYAFTYN   253
N. patriciarum CelB   DLFVKTIRATGGNNALRHLMIPTYAACINDGAINNFKFPSGDDKVIVSLHSYSPYNFALN   257
C. thermocellum EGE   LAVVNTIRASGGNNDKRFILVPTNAATGLDVALNDLVIPNNDSRVIVSIHAYSPYFFAMD   278


P. equi Cel5A 1       .NGEGAVDKF.DATGKNELDYNLGLIKKRFVSKGIPVIMGEYGAMNRDNE.EVRATWAEY   401
P. equi Cel5A 2       .NGEGAVDKF.DATGKNELDYNLGLIKKRFVSKGIPVIMGEYGAMNRDNE.EERATWAEY   790
P. equi Cel5A 3       .NGEGAVDKF.DATGKNELDYNLGLIKKRFVSKGIPVIMGEYGAMNRDNE.EERATWAEY  1175
P. equi Cel5A 4       .NGEGAVDKF.DATGKNELDYNLGLIKKRFVSKGIPVIMGEYGAMNRDNE.EERATWAEY  1564
C. longisporum CelA   TSSN.SVNTWGSSYDKYSLDVELDSYLNTFKSKGVPVVIGEFGSINK.NNTSSRAELAEY   327
R. albus EGA          ANADWELFHWDDSHD.GELVSLMTNLKENYLDKDIPVIITEYGAVNKDNNDEDRAKWVSS   312
N. patriciarum CelB   .NGAGAISNF...YDGSEIDWAMNTINSKFISRGIPVIIGEFGAMNRNNE.DDRERWAEY   312
C. thermocellum EGE   VNG...TSYWGSDYDKASLTSELDAIYNRFVKNGRAVIIGEFGTIDK.NNLSSRVAHAEH   334


P. equi Cel5A 1       YMKEITAL.GVPQVWWDNGIFE.GEGERFGLIDRKNLKVVYPSIVAALQKGRGLEVNVLH   459
P. equi Cel5A 2       YMKEITAL.GIPQVWWDNGIFE.GEGERFGLIDRKNLKVVYPSIVAALQKGRGLEVNVLH   848
P. equi Cel5A 3       YMKEITAL.GIPQVWWDNGIFE.GEGERFGLIDRKNLKVVYPSIVAALQKGRGLEVNVLH  1233
P. equi Cel5A 4       YMKEITAL.GIPQVWWDNGVFE.GEGERFGLIDRKNLKVVYPSIVAALQKGRGLEVNVLH  1622
C. longisporum CelA   YVTAAQK.RGIPCVWWDNNYAETNKGETFGLLNRSTLNWYFSDIKDALIRGYK...NVHP   383
R. albus EGA          YIEYAELLGGIPCVWWDNGYYSSGN.ELFGIFDRNTCTWFTDTVTDAIIENAK          365
N. patriciarum CelB   YIKKATSI.GVPCVIWDNGYFE.GEGERFGLINRSTLQVVYPKLVNGLIKGLGNSIKTRT   370
C. thermocellum EGE   YAREA.VSRGIAVFWWDNGYYNPGDAETYALLNRKTLSWYYPEIVQALMRGAGVEPLVSP   393


P. equi Cel5A 1       AIE    462
P. equi Cel5A 2       AIE    851
P. equi Cel5A 3       AIE   1236
P. equi Cel5A 4       AIE   1625
C. longisporum CelA   EAT    386
R. albus EGA                 365
N. patriciarum CelB   TIR    373
C. thermocellum EGE   TPT    396
```

***Fig. 3.*** Comparison of the amino acid sequences of the putative catalytic domains of *P. equi* Cel5A (between residues 106 and 462, 495 and 851, 880 and 1236, and 1269 and 1625) with the sequences of the catalytic domains of *Clostridium longisporum* CelA (SWISS-PROT P54937) (Mittendorf & Thomson, 1993), *Ruminococcus albus* EGA (SWISS-PROT P23660) (Poole *et al.*, 1990), *N. patriciarum* CelB (SWISS-PROT Q12647) (Zhou *et al.*, 1994) and *C. thermocellum* EGE (SWISS-PROT P10477) (Hall *et al.*, 1988); conserved residues are shown in bold.

```
P. equi Cel45A          QTGKTTRYWDCCLASCSWQENCKNDGAQGVVRSCNVDGITPFTDLSNLWRVKSGCN.GGS 255
U. maydis EG1           RAGMATRYWDCCLASASWE...GKAPVYAPVDACKADGVT.LIDSKKDPSGQSGCN.GGN 79
H. insolens EGV         ADGRSTRYWDCCKPSCGW...AKKAPVNQPVFSCNAN.FQRITDFD....AKSGCEPGGV 52
P. fluorescens EGB      CNGYATRYWDCCKPHCGWSANV..PSLVSPLQSCSANN.TRLSDVS....VGSSCDGGG. 318
F. oxysporum Kfam1      GSGHSTRYWDCCKPSCSW...SGKAAVNAPALTCDKN.DNPISNTN....AVNGCEGGGS 71

P. equi Cel45A          VYMCNDQQPW..AINDNVAYGFVA.....SHEK..CCTCQRLKFT..........SGPI 295
U. maydis EG1           KFMCSCMQPFDDETDPTLAFGFGA..FTTGQESDTDCACFYAEFEHDAQGK.......AM 130
H. insolens EGV         AYSCADQTPW..AVNDDFALGFAATSIAGSNEAGWCCACYELTFT..........SGPV 99
P. fluorescens EGB      GYMCWDKIPF..AVSPTLAYGYAAT.....SSGDVCGRCYQLQFTGSSYNAPGDPGSAAL 371
F. oxysporum Kfam1      AYACTNYSPW..AVNDELAYGFAATKISGGSEASWCCACYALTFT..........TGPV 118

P. equi Cel45A          AGKQMIVQTTNTGGDLSSNHFDIQMPGGGFGIF.DGCTSQFGGS.YQWGERYGGISSA.. 351
U. maydis EG1           KRNKLIFQVTNVGGDVQSQNFDFQIPGGGLGAFPKGCPAQWGVEASLWGDQYGGVKSA.. 188
H. insolens EGV         AGKKMVVQSTSTGGDLGSNHFDLNIPGGGVGIF.DGCTPQFG.G.LP.GQRYGGISSR.. 152
P. fluorescens EGB      AGKTMIVQATNIGYDVSGGQFDILVPGGGVGAF.NACSAQWGVSNAELGAQYGGFLAACK 430
F. oxysporum Kfam1      KGKKMIVQSTNTGGDLGDNHFDLMMPGGGVGIF.DGCTSEFGKA.LG.GAQYGGISSR.. 173

P. equi Cel45A          .................SQCANL.....PPQLKAGCEWRF.NWFKNADNPAVVF..ERVQ 386
U. maydis EG1           .................TECSKL.....PKPLQEGCKWRFSEW...GDNPVLKGSPKRVK 223
H. insolens EGV         .................NECDRF.....PDALKPGCYWRF.DWFKNADNPSFSF..RQVQ 188
P. fluorescens EGB      QQLGYNASLSQYKSCVLNRCDSVFGSRGLTQLQQGCTWFA.EWFEAADNPSLKY..KEVP 487
F. oxysporum Kfam1      .................SECDSY.....PELLKDGCHWRF.DWFENADNPDFTF..EQVQ 208

P. equi Cel45A          CPKELTEITGCVP...GDDAS..A 405
U. maydis EG1           CPKSLIDRSGCQRK...DDNTISP 244
H. insolens EGV         CPAELVARTGCRR...NDDGNFPA 209
P. fluorescens EGB      CPAELTTRSGMNRSILNDIRNTCP 511
F. oxysporum Kfam1      CPKALLDISGCKR...DDDSSFPA 329
```

***Fig. 4.*** Comparison of the amino acid sequence of the putative catalytic domain of *P. equi* Cel45A (between residues 197 and 405) with the sequences of the catalytic domains of *Ustilago maydis* EG1 (SWISS-PROT P54424) (Schauwecker *et al.*, 1995), *Humicola insolens* EGV (SWISS-PROT P43316) (Davies *et al.*, 1993, 1995; Schülein, 1997), *Pseudomonas fluorescens* EGB (SWISS-PROT P18126) (Gilbert *et al.*, 1990) and *Fusarium oxysporum* Kfam1 (SWISS-PROT P45699) (Sheppard *et al.*, 1994); conserved residues are shown in bold.

*patriciarum* (accession no. Z31364) and *celA* from *Neocallimastix frontalis* (accession no. U38843). *Orpinomyces* sp. PC-2 *celB*, *O. joyonii celB*29 and *N. patriciarum celB* are all complete; their proposed start codons are preceded by A+T-rich DNA typical of the non-coding regions of anaerobic fungi (Li *et al.*, 1997b; Zhou *et al.*, 1994). Furthermore, all three genes encode modular enzymes consisting of a single family-5 catalytic domain at the C terminus followed by two dockerin domains at the N terminus. The gene encoded by *O. joyonii celA* consists of a C-terminal catalytic domain followed by a repeated domain of unknown function (Liu *et al.*, 1997). *N. frontalis celA* and *O. joyonii celB*2 are both incomplete. *N. frontalis celA* encodes a protein which consists of a 37-residue linker, followed by a family-5 catalytic domain and two dockerins (Fujino *et al.*, 1998). *O. joyonii celB*2 encodes a protein which consists of a family-5 catalytic domain followed by a single C-terminal dockerin. Based on comparison with the sequence of *P. equi cel5A*, we believe that the proposed start codon of *celB2* actually encodes residue 13 of the catalytic domain. Further support for this view comes from the fact that there is no stop codon upstream of the proposed start codon, and the DNA upstream of this codon is not A+T rich, suggesting that it may be part of a longer ORF. It is possible that both *N. frontalis celA* and *O. joyonii celB*2 are part of much larger genes. The presence of a single band at 5·5 kb on a Northern blot of *P. equi* RNA (Fig. 1), when probed with a truncated variant of *cel5A* encoding only a single catalytic domain (nucleotides 3790–4884), implies that *P. equi* does not produce any other enzymes homologous to the family-5 glycosyl hydrolases. Genes encoding xylanases with duplicated catalytic domains have been isolated from *P. equi* and *N. patriciarum* (Fanutti *et al.*, 1995; Gilbert *et al.*, 1992), but *cel5A* is the first complete cDNA to be cloned from an anaerobic fungus encoding an endoglucanase with multiple catalytic domains. The presence of reiterated catalytic domains within an enzyme could enable the domains to act synergistically; alternatively, the repetition could provide a means for increasing expression of the particular catalytic domain. The high degree of similarity between Cel5A and the other family-5 cellulases from anaerobic fungi reinforces the view that the anaerobic fungi are extremely closely related to each other and have evolved from a common ancestral plant-cell-wall-degrading fungus. An alternative hypothesis is that lateral transfer of genes has occurred between anaerobic fungi; the evidence for horizontal transfer of genes from rumen-inhabiting

***Table 1.*** Substrate specificity of Trx-Cel5A′ and Trx-Cel45A′

| Substrate | Trx-Cel5A′ | | Trx-Cel45A′ | |
|---|---|---|---|---|
| | Specific activity* [U (mg protein)$^{-1}$] | Relative activity† (%) | Specific activity* [U (mg protein)$^{-1}$] | Relative activity† (%) |
| CMC | 3·2 | 100 | 42·5 | 100 |
| Acid-swollen cellulose | 0·7 | 21·9 | 30·0 | 70·6 |
| Avicel | 0 | 0 | 0 | 0 |
| Bacterial microcrystalline cellulose | 0 | 0 | 0 | 0 |
| Barley *β*-glucan | 59·4 | 1856·3 | 97·4 | 229·2 |
| Lichenin | 2·9 | 90·6 | 69·4 | 163·3 |
| Laminarin | 0 | 0 | 0 | 0 |
| Carob galactomannan | 1·1 | 34·4 | 0 | 0 |
| Galactan | 0 | 0 | 0 | 0 |
| Xylan | 0·2 | 6·3 | 0 | 0 |
| Arabinan | 0 | 0 | 0 | 0 |
| pNPC | 0·1 | 3·1 | 0 | 0 |

* Activity was measured at intervals over a period of an hour at 37 °C (Trx-Cel5A′) or 60 °C (Trx-Cel45A′) in PC buffer pH 6·5; the initial rate of activity was taken.

† Values expressed as percentages of the activity against CMC.

bacteria to fungi suggests that anaerobic fungi are capable of the uptake of genetic material (Chen *et al.*, 1997; Gilbert *et al.*, 1992; Hazlewood & Gilbert, 1998; Li *et al.*, 1997b; Zhou *et al.*, 1994).

Cel45A has a putative C-terminal catalytic domain located between residues 195 and 410. Comparison of the sequence of this catalytic domain with sequences in the SWISS-PROT database revealed homology with family-45 endoglucanases (Fig. 4). Cel45A is the first family-45 endoglucanase to be reported from an anaerobic organism. This suggests that Cel45A has not evolved by horizontal transfer of genes from rumen bacteria but may have originated from an aerobic fungal gene.

## Characterization of Cel5A′ and Cel45A′

To determine the biochemical properties of a single putative catalytic domain of Cel5A (located between residues 1269 and 1625) and the putative catalytic domain of Cel45A (located between residues 195 and 410), fusion genes consisting of the DNA sequences encoding these domains, fused to the thioredoxin gene, were constructed using the pET32a vector. Purified Trx-Cel5A′ and Trx-Cel45A′ appeared as single bands on SDS-PAGE; their apparent molecular masses of 56 kDa (Trx-Cel5A′) and 45 kDa (Trx-Cel45A′) were consistent with their predicted molecular masses. The initial rates of activity of Trx-Cel5A′ and Trx-Cel45A′ on various substrates were determined (Table 1). Trx-Cel5A′ showed cellulase, xylanase and mannanase activities; CMC, acid-swollen cellulose, barley *β*-glucan, lichenin, xylan, galactomannan and pNPC were all hydrolysed, with specific activities of between 0·1 and 59·4 U (mg protein)$^{-1}$. No hydrolysis was observed with laminarin, arabinan, galactan, bacterial microcrystalline cellulose or Avicel as substrates. Most of the family-5 enzymes reported to date are endoglucanases, but the family also includes exoglucanases (Mackenzie *et al.*, 1997; Vazquez de Aldana *et al.*, 1991), mannanases (Hilge *et al.*, 1998) and some enzymes which are primarily endoglucanases but with a broad substrate specificity. *P. equi* Cel5A shows more sequence similarity to the family-5 endoglucanases than to the exoglucanases or the mannanases. Trx-Cel45A′ hydrolysed CMC, acid-swollen cellulose and the mixed-linkage glucans, lichenin and barley *β*-glucan efficiently, with specific activities of between 30·0 and 97·4 U (mg protein)$^{-1}$. No hydrolysis was observed with laminarin, xylan, galactomannan, galactan, arabinan, bacterial microcrystalline cellulose, Avicel, or pNPC as substrates. The family-45 enzyme, endoglucanase B, from *Pseudomonas fluorescens* subspecies *cellulosa*, has activity against CMC, acid-swollen cellulose, barley *β*-glucan, lichenin and Avicel (Gilbert *et al.*, 1990). The lack of activity of Cel45A′ against Avicel may be explained by the fact that, unlike EGB, Cel45A′ has no cellulose-binding domain. Trx-Cel5A′ and Trx-Cel45A′ are both more active against barley *β*-glucan than CMC; a likely reason for this is that CMC is highly substituted with methoxy side chains which may interfere with enzyme activity. The specific activities of these enzymes against acid-swollen cellulose are less than the specific activities against CMC; acid-swollen cellulose has a more highly ordered structure than CMC and is therefore less accessible to these enzymes.

The pH profiles of Trx-Cel5A′ and Trx-Cel45A′ were

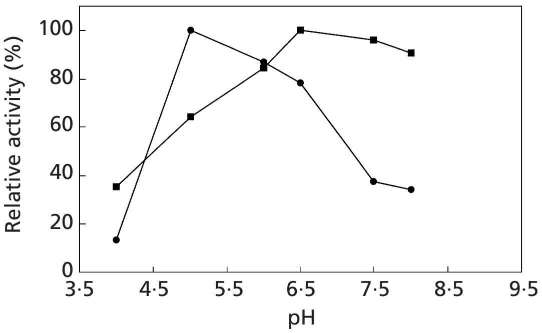

***Fig. 5.*** pH profiles of Trx-Cel5A′ and Trx-Cel45A′. Trx-Cel45A′ (■) and Trx-Cel5A′ (●) activities against CMC were determined at 37 °C and 60 °C, respectively, in citric acid/$Na_2HPO_4$ buffer (pH 4·0–7·6) and sodium barbitone/HCl buffer (pH 7·9–8·1). Assays were performed for 5, 10, 15, 30 and 60 min to determine the initial rate of activity at each pH.

examined between pH 4·0 and 8·0 using CMC as a substrate. Trx-Cel5A′ is active over a narrow pH range (Fig. 5), with an optimum of around pH 5·1; approximately 78 % of maximum activity is seen at pH 6·4, but activity falls off sharply at lower pH values. Trx-Cel45A′ is active over a broader pH range, with an optimum of around pH 6·5 and approximately 65 % of maximum activity between pH 5·2 and 7·9. The effect of temperature on Trx-Cel5A′ and Trx-Cel45A′ activity on CMC was determined. The initial rate of hydrolysis of CMC by Trx-Cel5A′ was highest at 45 °C. Preincubation at 50 °C and at 55 °C caused a decrease in Trx-Cel5A′ activity; an almost complete loss of Trx-Cel5A′ activity was seen after 20 min preincubation at 55 °C. All further assays with Trx-Cel5A′ were performed at 37 °C since the enzyme was unstable at higher temperatures. The initial rate of Trx-Cel45A′ activity increased with increasing temperature between 40 °C and 70 °C. Trx-Cel45A′ was more thermostable than Trx-Cel5A′; the enzyme was stable for at least an hour at 65 °C, but after 20 min incubation at 70 °C only 10 % of the original enzyme activity remained. For this reason all assays with Trx-Cel45A′ were performed at 60 °C.

The data presented in this report provide further insights into the complex nature of the cellulase systems expressed by anaerobic fungi. It is apparent that the cellulase–hemicellulase complex produced by these microbial eukaryotes includes family-45 enzymes, dockerin sequences can be present at both the N and C termini of the catalytic components of the complex, and evidence is presented which indicates that certain cellulases produced by these fungi have evolved by extensive duplication of an ancestral core sequence.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support of BBSRC and Finnfeeds International. We thank Dr Peter White, Jo Durgan, Keith Davidson, Judith Laurie and Babraham Technix for their assistance.

## REFERENCES

**Ali, B. R. S., Zhou, L., Graves, F. M., Freedman, R. B., Black, G. W., Gilbert, H. J. & Hazlewood, G. P. (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**, 15–21.

**Aylward, J. H., Gobius, K. S., Xue, G.-P., Simpson, G. D. & Dalrymple, B. P. (1999).** The *Neocallimastix patriciarum* cellulase, CelD, contains three almost identical catalytic domains with high specific activities on avicel. *Enzyme Microb Technol* **24**, 609–614.

**Béguin, P. (1990).** Molecular biology of cellulose degradation. *Annu Rev Microbiol* **44**, 219–248.

**Béguin, P. & Lemaire, M. (1996).** The cellulosome: an exocellular, multiprotein complex specialized in cellulose degradation. *Crit Rev Biochem Mol Biol* **31**, 201–236.

**Birnboim, H. C. & Doly, J. (1979).** A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**, 1513–1522.

**Black, G. W., Hazlewood, G. P., Xue, G.-P., Orpin, C. G. & Gilbert, H. J. (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.

**Bradford, M. M. (1976).** A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal Biochem* **72**, 248–254.

**Chen, H., Li, X.-L. & Ljungdahl, L. G. (1997).** Sequencing of a 1,3-1,4-β-D-glucanase (lichenase) from the anaerobic fungus *Orpinomyces* strain PC-2: properties of the enzyme expressed in *Escherichia coli* and evidence that the gene has a bacterial origin. *J Bacteriol* **179**, 6028–6034.

**Dalrymple, B. P., Cybinski, D. H., Layton, I., McSweeney, C. S., Xue, G.-P., Swadling, Y. J. & Lowry, J. B. (1997).** Three *Neocallimastix patriciarum* esterases associated with the degradation of complex polysaccharides are members of a new family of hydrolases. *Microbiology* **143**, 2605–2614.

**Davies, G. J., Dodson, G. G., Hubbard, R. E. & 7 other authors (1993).** Structure and function of endoglucanase V. *Nature* **365**, 362–364.

**Davies, G. J., Tolley, S. P., Henrissat, B., Hjort, C. & Schülein, M. (1995).** Structures of oligosaccharide-bound forms of the endoglucanase V from *Humicola insolens* at 1·9Å resolution. *Biochemistry* **34**, 16210–16220.

**Denman, S., Xue, G.-P. & Patel, B. (1996).** Characterization of a *Neocallimastix patriciarum* cellulase cDNA (*CelA*) homologous to *Trichoderma reesei* cellobiohydrolase II. *Appl Environ Microbiol* **62**, 1889–1896.

**Dijkerman, R., Op den Camp, H. J. M., Van der Drift, C. & Vogels, G. D. (1997).** The role of the cellulolytic high molecular mass (HMM) complex of the anaerobic fungus *Piromyces* sp. strain E2 in the hydrolysis of microcrystalline cellulose. *Arch Microbiol* **167**, 137–142.

**Fanutti, C., Ponyi, T., Black, G. W., Hazlewood, G. P. & Gilbert, H. J. (1995).** The conserved noncatalytic 40-residue sequence in cellulases and hemicellulases from anaerobic fungi functions as a protein docking domain. *J Biol Chem* **270**, 29314–29322.

**Fujino, Y., Ogata, K., Nagamine, T. & Ushida, K. (1998).** Cloning, sequencing, and expression of an endoglucananse gene from the rumen anaerobic fungus *Neocallimastix frontalis* MCH3. *Biosci Biotechnol Biochem* **62**, 1795–1798.

**Gilbert, H. J., Hall, J., Hazlewood, G. P. & Ferreira, L. M. A. (1990).** The N-terminal region of an endoglucanase from *Pseudomonas* *fluorescens* subspecies *cellulosa* constitutes a cellulose-binding domain that is distinct from the catalytic centre. *Mol Microbiol* **4**, 759–767.

**Gilbert, H. J., Hazlewood, G. P., Laurie, J. I., Orpin, C. G. & Xue, G.-P. (1992).** Homologous catalytic domains in a rumen fungal xylanase: evidence for gene duplication and prokaryotic origin. *Mol Microbiol* **6**, 2065–2072.

**Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C. & Warren, R. A. J. (1991).** Domains in microbial *β*-1,4-glycanases: sequence conservation, function, and enzyme families. *Microbiol Rev* **55**, 303–315.

**Hall, J., Hazlewood, G. P., Barker, P. J. & Gilbert, H. J. (1988).** Conserved reiterated domains in *Clostridium thermocellum* endoglucanases are not essential for catalytic activity. *Gene* **69**, 29–38.

**Hazlewood, G. P. & Gilbert, H. J. (1998).** Structure–function relationships in the cellulase–hemicellulase system of anaerobic fungi. In *Carbohydrases from Trichoderma reesei and Other Microorganisms*, pp. 147–155. Edited by M. Claeyssens, W. Nerinckx & K. Piens. Cambridge: The Royal Society of Chemistry.

**Henrissat, B. (1991).** A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* **280**, 309–316.

**Henrissat, B. (1998).** Glycosidase families. *Biochem Soc Trans* **26**, 153–156.

**Henrissat, B. & Bairoch, A. (1993).** New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* **293**, 781–788.

**Henrissat, B. & Bairoch, A. (1996).** Updating the sequence-based classification of glycosyl hydrolases. *Biochem J* **316**, 695–696.

**Hilge, M., Gloor, S. M., Rypniewski, W., Sauer, O., Heightman, T. D., Zimmermann, W., Winterhalter, K. & Piontek, K. (1998).** High-resolution native and complex structures of thermostable *β*-mannanase from *Thermomonospora fusca* – substrate specificity in glycosyl hydrolase family 5. *Structure* **6**, 1433–1444.

**Kemp, P., Lander, D. J. & Orpin, C. G. (1984).** The lipids of the anaerobic rumen fungus *Piromonas communis*. *J Gen Microbiol* **130**, 27–37.

**Laemmli, U. K. (1970).** Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

**Leschine, S. B. (1995).** Cellulose degradation in anaerobic environments. *Annu Rev Microbiol* **49**, 399–426.

**Li, J. & Heath, I. B. (1993).** Chytridiomycetous gut fungi, oft overlooked contributors to herbivore digestion. *Can J Microbiol* **39**, 1003–1013.

**Li, X.-L., Chen, H. & Ljungdahl, L. G. (1997a).** Monocentric and polycentric anaerobic fungi produce structurally related cellulases and xylanases. *Appl Environ Microbiol* **63**, 628–635.

**Li, X.-L., Chen, H. & Ljungdahl, L. G. (1997b).** Two cellulases, CelA and CelC, from the polycentric anaerobic fungus *Orpinomyces* strain PC-2 contain N-terminal docking domains for a cellulase-hemicellulase complex. *Appl Environ Microbiol* **63**, 4721–4728.

**Liu, J.-H., Selinger, L. B., Hui, Y.-H., Molony, M. M., Cheng, K.-J. & Beauchemin, K. A. (1997).** An endoglucanase from the anaerobic fungus *Orpinomyces joyonii*: characterization of the gene and its product. *Can J Microbiol* **43**, 477–485.

**Mackenzie, L. F., Brooke, G. S., Cutfield, J. F., Sullivan, P. A. & Withers, S. G. (1997).** Identification of Glu-330 as the catalytic nucleophile of *Candida albicans* exo-*β*-(1,3)-glucanase. *J Biol Chem* **272**, 3161–3167.

**Mayer, F., Coughlan, M. P., Mori, Y. & Ljungdahl, L. G. (1987).** Macromolecular organization of the cellulolytic enzyme complex of *Clostridium thermocellum* as revealed by electron microscopy. *Appl Environ Microbiol* **53**, 2785–2792.

**Miller, G. L. (1959).** Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem* **31**, 426–428.

**Millward-Sadler, S. J., Hall, J., Black, G. W., Hazlewood, G. P. & Gilbert, H. J. (1996).** Evidence that the *Piromyces* gene family encoding endo-1,4-mannanases arose through gene duplication. *FEMS Microbiol Lett* **141**, 183–188.

**Mittendorf, V. & Thomson, J. A. (1993).** Cloning of an endo-(1→4)-*β*-glucanase gene, *celA*, from the rumen bacterium *Clostridium* sp. ('*C. longisporum*') and characterization of its product, CelA, in *Escherichia coli*. *J Gen Microbiol* **139**, 3233–3242.

**Munn, E. A. (1994).** The ultrastructure of anaerobic fungi. In *Anaerobic Fungi: Biology, Ecology, and Function*, pp. 47–105. Edited by D. O. Mountfort & C. G. Orpin. New York: Marcel Dekker.

**Orpin, C. G. (1975).** Studies on the rumen flagellate *Neocallimastix frontalis*. *J Gen Microbiol* **91**, 249–262.

**Orpin, C. G. (1981).** Isolation of cellulolytic phycomycete fungi from the caecum of the horse. *J Gen Microbiol* **123**, 287–296.

**Poole, D. M., Hazlewood, G. P., Laurie, J. I., Barker, P. J. & Gilbert, H. J. (1990).** Nucleotide sequence of the *Ruminococcus albus* SY3 endoglucanase genes *celA* and *celB*. *Mol Gen Genet* **223**, 217–223.

**Saloheimo, A., Henrissat, B., Hoffrén, A.-M., Teleman, O. & Penttilä, M. (1994).** A novel, small endoglucanase gene, *egl5*, from *Trichoderma reesei*, isolated by expression in yeast. *Mol Microbiol* **13**, 219–228.

**Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989).** *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

**Schauwecker, F., Wannner, G. & Kahmann, R. (1995).** Filament-specific expression of a cellulase gene in the dimorphic fungus *Ustilago maydis*. *Biol Chem Hoppe-Seyler* **376**, 617–625.

**Schülein, M. S. (1997).** Enzymatic properties of cellulases from *Humicola insolens*. *J Biotechnol* **57**, 71–81.

**Selinger, L. B., Forsberg, C. W. & Cheng, K.-J. (1996).** The rumen: a unique source of enzymes for enhancing livestock production. *Anaerobe* **2**, 263–284.

**Sheppard, P. O., Grant, F. J., Oort, P. J., Sprecher, C. A., Foster, D. C., Hagen, F. S., Upshall, A., McKnight, G. L. & O'Hara, P. J. (1994).** The use of conserved cellulase family-specific sequences to clone cellulase homologue cDNAs from *Fusarium oxysporum*. *Gene* **150**, 163–167.

**Shoham, Y., Lamed, R. & Bayer, E. A. (1999).** The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. *Trends Microbiol* **7**, 275–281.

**Teather, R. M. & Wood, P. J. (1982).** Use of Congo red–polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Appl Environ Microbiol* **43**, 777–780.

**Teunissen, M. J., Hermans, J. M. H., Huis in 't Veld, J. H. J. & Vogels, G. D. (1993).** Purification and characterization of a complex bound and a free *β*-1,4-endoxylanase from the culture fluid of the anaerobic fungus *Piromyces* sp. strain E2. *Arch Microbiol* **159**, 265–271.

**Tomme, P., Warren, R. A. J. & Gilkes, N. R. (1995).** Cellulose hydrolysis by bacteria and fungi. *Adv Microb Physiol* **37**, 1–81.

**Trinci, A. P. J., Davies, D. R., Gull, K., Lawrence, M. I., Nielsen, B. B., Rickers, A. & Theodorou, M. K. (1994).** Anaerobic fungi in herbivorous animals. *Mycol Res* **98**, 129–152.

**Vazquez de Aldana, C. R., Correa, J., San Segundo, P., Bueno, A., Nebreda, A. R., Mendez, E. & del Rey, F. (1991).** Nucleotide sequence of the exo-1,3-β-glucanase-encoding gene, *EXG1*, of the yeast *Saccharomyces cerevisiae*. *Gene* **97**, 173–182.

**Warren, R. A. J. (1996).** Microbial hydrolysis of polysaccharides. *Annu Rev Microbiol* **50**, 183–212.

**Wilson, C. A. & Wood, T. M. (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.

**Wood, T. M. (1971).** The cellulase of *Fusarium solani*. Purification and specificity of the β-(1-4)-glucanase and the β-D-glucosidase components. *Biochem J* **121**, 353–362.

**Wubah, D. A., Akin, D. E. & Borneman, W. S. (1993).** Biology, fiber-degradation, and enzymology of anaerobic zoosporic fungi. *Crit Rev Microbiol* **19**, 99–115.

**Xue, G.-P., Gobius, K. S. & Orpin, C. G. (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.

**Zhou, L., Xue, G.-P., Orpin, C. G., Black, G. W., Gilbert, H. J. & Hazlewood, G. P. (1994).** Intronless *celB* from the anaerobic fungus *Neocallimastix patriciarum* encodes a modular family A endoglucanase. *Biochem J* **297**, 359–364.

Received 17 February 2000; revised 11 May 2000; accepted 17 May 2000.