Mycoscience (2005) 46:154–161
DOI 10.1007/s10267-005-0225-0

© The Mycological Society of Japan and Springer-Verlag Tokyo 2005

FULL PAPER

**Hiroki Ishihara · Tadanori Aimi · Keisuke Takahashi
Yutaka Kitamoto**

Heterologous expression and characterization of the endocellulase encoding gene *cel3A* from the basidiomycete *Polyporus arcularius*

Received: October 12, 2004 / Accepted: January 15, 2005

Abstract The genomic and cDNA clones encoding the carboxymethyl cellulase (CMCase) gene (*cel3A*) of *Polyporus arcularius* were sequenced and characterized. The coding region of *cel3A*, composed of 1329 bp, was found to encode a polypeptide of 243 amino acids that has similarity with FI-CMCase of *Aspergillus aculeatus.* Expression of the *cel3A* cDNA in *Escherichia coli* led to production of a nonglycosylated protein as an active form; moreover, CMCase activity measured by the viscometric method was enhanced 4.47 times by addition of cellobiose. These results indicate that glycosylation and any modification of protein such as processing are not required for Cel3A activity.

Introduction

Cellulose, which is a polymer of glucose units linked by β-1,4-glycosidic bonds, is the most abundant carbohydrate in the biosphere. Cellulose is produced at an estimated rate of 4×10^7 tons per year. Cellulose is the most promising large-scale and renewable carbon source available for long-range solutions to energy, chemical, and food resource problems (Murai et al. 1998). The most useful technology to utilize cellulose is biological conversion to alcohol via cellooligosaccharides; however, depolymerization of cellulose into cellooligosaccharides with commercial enzymes remains expensive (Wood and Ingram 1992). Thus, identification of a new microbial cellulase has been needed for some time.

Cellulose depolymerization is mediated by endoglucanases, which randomly cleave the internal β-1,4-glucosidic links, and cellobiohydrolases, which act on the free ends of cellulose polymer chains. Cellulases belong to the large group of glycosyl hydrolases, of which there are several families based on amino acid sequence similarities. Over the years, the number of glycosyl hydrolase families has grown steadily, and currently there are 82 known families (Henrissat and Davies 1997).

Cellulases, including the members of glycosyl hydrolase family 12, are widely used in a number of industrial applications such as detergent, textile, and food processing (Goedegebuur et al. 2002). Gene sequences of family 12 glycosyl hydrolases have been reported in numerous filamentous fungi, such as *Aspergillus aculeatus* Iizuka (Ooi et al. 1990), *A. kawachii* Kitahara & Yoshida (Sakamoto et al. 1995), *Trichoderma viride* Persoon: Fries (Goedegebuur et al. 2002), *T. reesei* Simmons (Ward et al. 1993), *Humicola grisea* Traaen (Goedegebuur et al. 2002), *H. insolens* Cooney et Emerson (Dalboge et al. 1993), and *Fusarium equiseti* (Corda) Saccardo (Goedegebuur et al. 2002). However, the nucleotide sequence of the gene for the family 12 glycosyl hydrolase from basidiomycete has not yet been reported.

The polypore mushroom *Polyporus arcularius* Fr. is a wood-decomposing basidiomycete that produces four types (I, II, and IIIa and IIIb) of carboxymethyl cellulase (CMCase). Although CMCases I and II display similar enzymatic properties, CMCase IIIa has unique enzymatic characteristics. The molecular mass of CMCases I, II, and IIIa are 39.1, 36.3, and 24.3 kDa, respectively. Furthermore, the enzymatic activities of CMCases I and II are inhibited by addition of reaction products, such as cellooligosaccharides and glucose, whereas CMCase IIIa activity is enhanced by the accumulation of reaction products (Ishihara et al. 2005). Based on these qualities, the mass production and protein engineering of this new enzyme are expected to allow its application to the industrial degradation of cellulose.

In this article, we describe characterization of the genomic DNA sequence of the CMCase-encoding gene from *P. arcularius* (*cel3A*), the cloning of its cDNA, and its expression in *Escherichia coli* (Migula) Castellani and Chalmers as the active and native enzyme. The present study is part of our attempt to utilize CMCase from a basidiomycete for industrial applications.

H. Ishihara · T. Aimi · K. Takahashi · Y. Kitamoto (✉)
Laboratory of Microbial Technology, Faculty of Agriculture, Tottori University, 4-101 Minami, Koyama-cho, Tottori 680-8553, Japan
Tel. +81-857-31-5371; Fax +81-857-31-5371
e-mail: kitamoto@muses.tottori-u.ac.jp

Materials and methods

Strains and plasmids

The monokaryotic strain *P.* (= *Favolus*) *arcularius* 69B-8 was a single basidiospore isolate from the fruiting body of the dikaryotic strain *P. arcularius* 69B (ATCC 24461). The *E. coli* strain DH5α (Takara Bio, Shiga, Japan) was used as a host for plasmid amplification. The plasmid pT7blue T-vector (Novagen, Madison, WI, USA) was used for cloning of polymerase chain reaction (PCR) products. *E. coli* strain BL21(DE3) (Novagen) was used as a host for recombinant protein expression, and pET-26b(+) (Novagen) vector was used for cloning and expression of recombinant protein.

Amino acid sequencing

CMCase IIIa was purified from culture filtrate of *P. arcularius* 69B, and 10ml was treated with 8M urea in 10mM Tris-HCl (pH 9.0) and incubated at 37°C for 2h. This mixture was diluted twofold with the same buffer, mixed with lysyl endpeptidase (Wako Pure Chemical, Osaka, Japan) at a protein/protease ratio of 300:1 (w/w) and incubated at 37°C for 24h. The resulting proteolytic digests were separated by reverse-phase high-pressure liquid chromatography (HPLC) on a YMC-PackC4-AP column (100mm × 1.0-mm inner diameter; YMC, Kyoto, Japan) using a gradient of 0%–80% acetonitrile in 0.1% trifluoroacetic acid at a flowrate of 0.5ml/min. Peptide fragments separated by reverse-phase HPLC were pooled and concentrated. The peptide fragments were subjected to sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (Laemmli 1970), blotted onto a polyvinylidene difluoride membrane, and sequenced. The sequence analysis was performed using an Applied Biosystems model 477A protein sequencer with a model 120A online phenylthiohydantoin analyzer (Applied Biosystems, Tokyo, Japan).

DNA and RNA preparation

Fungal cultures were grown in potato dextrose broth (potato extract with 2% glucose). Genomic DNA was isolated from lyophilized mycelium of *P. arcularius* 69B-8 using a Genomic Prep Cells and Tissue DNA Isolation Kit (Amersham Biosciences, Tokyo, Japan) according to the manufacturer's instructions.

To prepare total RNA from *P. arcularius* 69B-8, the mycelium was grown on potato dextrose agar (PDA) at 25°C for 1 week, after which the mycelium along with ten square agar blocks (5 × 5mm) were transferred to 50ml glucose-peptone medium (20g/l glucose, 2g/l peptone, 0.1g/l yeast extract, 1g/l KH_2PO_4, 0.3g/l $MgSO_4 \cdot 7H_2O$, and 0.1g/l $CaCl_2 \cdot 2H_2O$, pH 5.6) in a Sakaguchi flask and grown at 25°C for 1 week. The mycelium was harvested by centrifugation at 3000*g* for 10min and washed three times with sterilized distilled water. The washed mycelium was suspended in 20ml cellulase-producing medium (5g/l avicel, 2g/l peptone, 0.1g/l yeast extract, 1g/l KH_2PO_4, 0.5g/l $MgSO_4 \cdot 7H_2O$, and 0.1g/l $CaCl_2 \cdot 2H_2O$, pH 5.8; Enokibara et al. 1992) and homogenized with a Marusan Homoblender HG2 (Sakura Seisakusho, Tokyo, Japan). The medium containing the homogenized mycelium was added to 20ml fresh liquid cellulase-producing medium in a Sakaguchi flask and incubated at 25°C for 4 days. The mycelium was harvested by filtration, frozen in liquid nitrogen, and ground in a mortar and pestle to a fine powder. RNA extraction buffer [0.1mM citrate, 0.5% SDS, 1mM ethylene diamine tetraacetic acid (EDTA), pH 4.7] was added, the RNA-containing solution was extracted several times with RNA extraction buffer-saturated phenol, and total RNA was recovered by ethanol precipitation.

PCR

All amplified DNA fragments were subcloned into the pT7Blue(R) T-vector (Novagen), and the plasmids were sequenced. All PCRs were performed with a Takara PCR Thermal Cycler (Takara Bio). The PCR primers used in this study are shown in Table 1. Initially, we amplified fragments of genomic DNA encoding *cel3A* using PCR with degenerate oligonucleotide primers BOX1 and P3 (Table 1). The BOX1 primer was designed based on the conserved family 12 glycosyl hydrolase sequence SVNQNLW (Goedegebuur et al. 2002), and the P3 primer was designed based on the internal amino acid sequence DFFNYLY from purified *P. arcularius* CMCase IIIa (Fig. 1). PCR was carried out in 100µl containing 1× PCR buffer (Takara Bio), 100ng extracted genomic DNA, 100pmol each primer, 0.2mM each deoxynucleoside triphosphate (dNTP), and 2.5U Ex Taq polymerase (Takara Bio). The PCR was carried out by an initial denaturation for 3min at 95°C, followed by 30 cycles of denaturation at 95°C for 30s, annealing at 45°C for 2min, and extension at 72°C for 1min. Finally, the reaction mixture was incubated at 72°C for 10min. The resulting 674-bp fragment was subcloned, and the clone (pF12-2) was sequenced.

To clone the entire genomic sequence of *cel3A*, gene walking by inverse PCR was performed essentially according to Triglia et al. (1988) and Parker et al. (1991) and with oligonucleotide primers F12Up1 and F12Dn1, which were constructed based on the partial nucleotide sequences of *cel3A* (see Table 1). Genomic DNA (1.5µg) from *P. arcularius* 69B-8 strain was digested with *Pst*I, extracted with phenol/chloroform, precipitated with ethanol, and resuspended in TE buffer (10mM Tris-HCl, pH 8.0, and 1mM EDTA). The DNA was added to 50µl 10× ligation buffer (Takara Bio), H_2O was added to a total volume of 500µl, and 1000U T4 DNA ligase (Takara Bio) was added. The reaction mixtures were incubated overnight at 16°C, concentrated by ultrafiltration with a Microcon-100 (Amicon, Beverly, MA, USA), and used as templates for inverse PCR amplification. Genomic walking by PCR was carried out in 100µl containing 1× Ex Taq buffer (Takara Bio), 100ng template DNA, 100pmol each primer, 0.2mM each dNTP,

Table 1. Polymerase chain reaction (PCR) primers used in this study

Primer	Sequence	Remark
Primer BOX1 Primer P3	5′-WSNGTNAAYCARAAYYTNTGG-3′ 5′-TANARRTARTTRAARAARTC-3′	Used for initial amplification
F12Up1 F12Dn1	5′-CGGTAGACAGCATATTGACGA-3′ 5′-GAGGAGGTTAGCGTAGCTCTTG-3′	Used for inverse PCR and direct sequencing
F12F1 F12R1	5′-CAGAAGTGGCTTCTGCGATGCAGT-3′ 5′-CAGCCATATCGCTCGTGCAAAGGG-3′	Used for amplification of the full-length genomic clone of *cel3A* and direct sequencing
F12 3′-RACE	5′-GTTTTCATATCACCGAGGACAATC-3′	Used for 3′-RACE
F12RTPP	5′-TGCCCAACCAGATGTCGTAG-3′	Used for 5′-RACE; the 5′-end of this oligonucleotide was phosphorylated
F12 5′-RACEUp1	5′-ACACGGTGTTGCCGCTTGTGCTGA-3′	Used for 5′-RACE
F12 5′-RACEDn1	5′-AACAACGTGAAGAGCTACGCTAAC-3′	
F12 5′-RACEUp2	5′-AAGTTCTGGCAGAGCGTGTAGG-3′	
F12 5′-RACEDn2	5′-GCAGTGGGAGTACGAGAGCAAG-3′	
F12 RTF F12 RTR	5′-TATACGACCAACTCAAACATGCAC-3′ 5′-GTTTTCATATCACCGAGGACAATC-3′	Used for RT-PCR
F12 *Nde*N F12 *Xho*C	5′-*CATATG*CACTTCTCCACTCTCGCC-3′ 5′-G*CTCGAG*ACAACTGCTTTATCACTGGT-3′	Used for amplification of the full-length cDNA clone to generate *Nde*I and *Xho*I sites in cDNA

RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcription-polymerase chain reaction

and 2.5 U Ex Taq polymerase (Takara Bio). To amplify large fragments, the reaction was initiated with 2 min of denaturation at 94°C, followed by 30 cycles of denaturation for 30 s at 94°C, annealing at 58°C for 2 min, and extension at 72°C for 4 min. Finally, the reaction mixture was incubated at 72°C for 10 min. The resulting fragment was subcloned and sequenced.

To amplify the whole genomic clone of *cel3A*, oligonucleotide primers F12F1 and F12R1 (see Table 1) were designed based on the nucleotide sequence of DNA fragments amplified by the inverse PCR method. PCR was carried out in 100 µl containing 1× Ex Taq buffer, 100 ng extracted genomic DNA, 100 pmol each primer, 0.2 mM each dNTP, and 2.5 U Ex Taq polymerase. The reaction was initiated by an initial denaturation for 2 min at 94°C for 2 min, followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 58°C for 2 min, and extension at 72°C for 2 min. The amplified *cel3A* DNA fragments were purified with a Microcon-100 filter (Millipore, Bedford, MA, USA) and used as a DNA template for direct sequencing with the oligonucleotide primers F12Up1, F12Dn1, F12F1, and F12R1 (see Table 1).

DNA sequencing and computer analysis of nucleotide and protein sequences

DNA sequencing was carried out in an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Tokyo, Japan) by chain termination with a BigDye Terminator Cycle Sequencing version 3.1 kit (Applied Biosystems) according to manufacturer's instructions. Nucleotide and protein sequence data were analyzed by Genetyx-SV/RC version 6 (Software Development, Tokyo, Japan). The subcellular localization of Cel3A was predicted by PSORTII (Nakai and Kanehisa 1992; http://psort.nibb.ac.jp). Potential N- and O-glycosylation sites were predicted by NetNGlyc 1.0 Server (Gupta et al.; http://www.cbs.dtu.dk/services/NetNGlyc/) and DictyOGlyc 1.1 Server (Gupta et al. 1999; http://www.cbs.dtu.dk/services/DictyOGlyc/).

Southern hybridization

Genomic DNA from *P. arcularius* 69B-8 was isolated and digested with *Sal*I, *Pst*I, and *Xho*I, and the fragments were

```
   1  cagaagtggcttctgcgatgcagtcgacgacttgggttggtccgcctgagtgcacccacgggcttatgctgctgcgtggt    80
               F12F1
  81  acaactcgagagcacgactttttccgacgtgcatccctcgcgtgctcgagtctcagcgcaagaagatgctcgcgggtata   160
 161  tatacgccgctgttcggagaggctgacttaccaggttccagcgtcgaacccggcttcgctatacgaccaactcaaacATG   240
                                                                                  M
 241  CACTTCTCCACTCTCGCCACCCTTGTCGCCGTCGCGGCCACCGCGAGCGCACAAACGCTCTCGGGCCAGTACGACTGCGC   320
      H  F  S  T  L  A  T  L  V  A  V  A  A  T  A  S  A  Q  T  L  S  G  Q  Y  D  C  A

 321  GCCCGCTGGTGCCTACACGCTCTGCCAGAACTTGTGGGGCGAATgtgcgtgacggtttcttcctatgtgtcgtgagcgcg   400
       P  A  G  A  Y  T  L  C  Q  N  L  W  G  E  S
 401  actcacagtgtctcctacgcagCCTCCGGCGTCGGCAACCAGAACTCGACGCTGATCAGCACAAGCGGCAACACCGTGTC   480
                               S  G  V  G  N  Q  N  S  T  L  I  S  T  S  G  N  T  V  S
 481  GTGGCGCACGCAGTGGCAGTGGCAGAACAACCCGAACAACGTGAAGAGCTACGCTAACCTCCTCTCCAACTCGGCCAAGG   560
       W  R  T  Q  W  Q  W  Q  N  N  P  N  N  V  K  S  Y  A  N  L  L  S  N  S  A  K [G
                                                                                  P1

 561  GAGTTCAGgtgcgacactcagcctcccctagcatctatggaggagtctgaccgtttgccacagCTCTCCGCCGTCAGGGC   640
       V  Q]                                                          [L  S  A  V  R  A]
                                                                             P1

 641  CGCGCCGACAGCGTGGCAGTGGGAGTACGAGAGCAAGTCTGACGGTATTCGCGCGGACGTGTCCTACGACATCTGGTTGG   720
       A  P  T  A  W  Q  W  E  Y  E  S  K  S  D  G  I  R  A  D  V  S  Y  D  I  W  L  G

 721  GCACTGCGCCGTCTGGTGATCCCGCGTCCAGAGCATCCTCGTACGAGATCATGATCTGGCTCTCCGGCCTCGGCGGgtac   800
        T  A  P  S  G  D  P  A  S  R  A  S  S  Y  E  I  M  I  W  L  S  G  L  G  G
 801  gttctccccactgtacaggctcggtagacaggcatattgacgagccttgtgcccagCATCCAACCCGTCGGTTCGAAGAT   880
                                                              I  Q  P  V  G  S  K [I]
 881  CCTCTCCGGAGTGAACGTCGCAGGCCACACCTGGGACCTCTGGAAGGGCCCGAACTCGAACTGGCAGGTGCTCTCGTTCG   960
     [L  S  G  V  N  V  A  G  H  T  W  D  L  W  K] G  P  N  S  N  W  Q  V  L  S  F  V
                P2

 961  TGAGCTCCACGGGCGACATCACGGACTTCAACGTCGACCTCAAGGACTTCTTCAACTACCTCACGCAAAGCCAGGGTGTC  1040
        S  S  T  G  D  I  T  D  F  N  V  D  L  K [D  F  F  N  Y  L  T  Q] S  Q  G  V
                                                              P3

1041  GCGGCTTCGCAGgtgcgtgttctgtttactgtggcgcgcgagatctgaggatgctgacgatctatgggtcgcagTATGTG  1120
      A  A  S  Q                                                                 Y  V
1121  CAAGCCATCCAGACGGGCACGGAGCCGTTCACGGGAAGCGCGAGCTTGTTCACGAAGGCGTACAGCGTTGCTATCAACCA  1200
      Q  A  I  Q  T  G  T  E  P  F  T  G  S  A  S  L  F  T  K  A  Y  S  V  A  I  N  Q
1201  GTGAagcagttgtctctagcaattgttgcgagattgtcctcggtgatatgaaaacagcctctctctgttgtccccatgcg  1280
       *
1281  Tggcgctcgtcgtggcaaaggcatgccctttgcacgagcgatatggctg                                 1329
```

Fig. 1. Nucleotide and amino acid sequences of *cel3A* gene from *Polyporus arcularius*. *Capital* and *lowercase letters* indicate exons and introns, respectively. The stop codon is indicated by an *asterisk*. The two primers (F12F1 and F12R1) used for amplification of the CMCase IIIa gene are *underlined*. The three partial amino acid sequences (P1, GVQLSAVRTA; P2, ILSGVNVAGHTTDLLK; P3, DFFNYLYQ) of native CMCase IIIa are *boxed*. The amino acid sequence of the putative signal sequence is *double underlined*

separated by agarose gel electrophoresis and blotted onto a nylon membrane (Hybond-N+; Amersham Biosciences). The blot was probed with pF12-2 labeled with a DIG-High Prime DNA Labeling and Detection Kit (Roche Diagnostics, Tokyo Japan). Probe labeling, hybridization, and detection were carried out according to the manufacturer's recommendations.

Reverse transcription-PCR (RT-PCR)

Amplification of full-length cDNA by RT-PCR and 3′-rapid amplification of cDNA ends (RACE) were performed with a Takara RNA LA PCR (AMV) version 1.1 kit (Takara Bio). The 5′-RACE was carried out with a 5′-Full RACE Core Set (Takara Bio). The reverse transcription and PCR were carried out according to the manufacturer's instructions using the primers listed in Table 1. The amplified fragments were subcloned and sequenced. We selected the pT7blue-*cel3A* plasmid that contained the full-length cDNA for the genomic sequences of *cel3A* but lacked the intron sequences.

Expression of *cel3A* cDNA in *E. coli*

For the construction of the *cel3A* cDNA expression plasmid in *E. coli*, a DNA fragment containing the *cel3A* coding region, framed with restriction enzyme sites at both ends, was obtained by PCR using pT7Blue-*cel3A* as a template and with the sequence-specific primers F12 NdeN and F12 XhoC. The PCR fragment was digested with *Nde*I and *Xho*I and ligated with *Nde*I/*Xho*I-digested pET-26b (+); this generated plasmid pET-*cel3A*, which was used to transform *Escherichia coli* BL21(DE3). The transformants carrying pET-*cel3A* and pET-26b (+) were grown at 25°C in LB medium (1% tryptone, 0.5% yeast extract, and 1% NaCl) supplemented with 50 μg/ml kanamycin. When the cell density reached an OD_{660} of 0.6, the culture was adjusted to 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and further cultivated for 4 h at 25°C. The cells were harvested by centrifugation at 3000 *g* for 10 min, suspended in 20 mM potassium phosphate buffer (pH 7.0) supplemented with 2 mM glutathione and 100 μg/ml lysozyme, and disrupted with a sonicator. Following centrifugation at 3000 *g* for 10 min, the resulting supernatant was analyzed for CMCase activity and protein concentration by the method of Lowry et al. (1951).

Determination of enzyme activities

A colorimetric assay of CMCase activity was used as the standard assay of cellulase activity. The reaction mixture contained 0.6% CMC, 50 mM sodium acetate buffer (pH 4.5), and enzyme solution in a total volume of 1.0 ml. After incubation at 37°C for 10 min, the reducing sugars released from the substrate were measured as glucose by the Somogyi–Nelson method (Somogyi 1959). One unit of enzyme activity was defined as the quantity of enzyme that liberates 1 μmol of reducing sugar equivalents per minute. The specific activity was expressed as units of enzyme activity per milligram protein.

We used a viscometric assay for determination of endoglucanase activity. The reaction mixture contained 1.5% CMC, 50 mM sodium acetate buffer (pH 4.5), and 20 μl enzyme solution in a total volume of 2.0 ml. The mixture was incubated at 37°C for 10 min, and the reaction was terminated by heating in a boiling water bath for 60 min. After cooling, the viscosity of the mixture was determined at 20°C by using a rotary viscometer (Type E; Tokyo Instrument, Tokyo, Japan). The enzyme activity of each cellulase component was calculated as the reciprocal of η (1/cP), which, for CMCase IIIa, is proportional to the corresponding value from the CMCase colorimetric assay (Ishihara et al. 2005).

Polyacrylamide gel electrophoresis (PAGE) and active staining

Native PAGE was carried out according to Davis (1964) on a 10% polyacrylamide slab gel containing 20% glycerol. Active staining for CMCase was performed as described by Béguin (1983) with slight modifications. Briefly, the enzyme was separated by native PAGE at 2°C, and the gel was laid on a 1% agarose gel (5 mm thick) containing 0.5% CMC and 100 mM sodium acetate buffer (pH 4.5) that had been solidified in a Petri dish (11 cm × 7 cm). After incubation at 37°C for 60 min, the overlaid gel was removed, and the agarose replica gel was stained with 0.2% Congo red for 10 min and destained with 1 M NaCl. Location of the enzyme was detected in the gel as unstained bands.

Results

Internal amino acid sequence of CMCase IIIa

We first cleaved CMCase IIIa with lysyl endopeptidase to obtain peptides for sequencing, which could then be used to develop primers for PCR. We obtained the following amino acid sequences: GVQLSAVRTA (P1), ILSGVNVAGHTT DLLK (P2), and DFFNYLYQ (P3). The amino acids sequences of peptides P1 and P2 do not show significant homology with any sequence in the protein and nucleic acid data banks by homology searching with the BLAST programme (http://www.ddbj.nig.ac.jp/search/blast-j.html). However, P3 is 87% identical to a sequence in *Aspergillus kawachii* endo-β-1,4-glucanase (Sakamoto et al. 1995). Thus, a PCR primer for the amplification of the CMCase-encoding gene was constructed based on the P3 amino acid sequence.

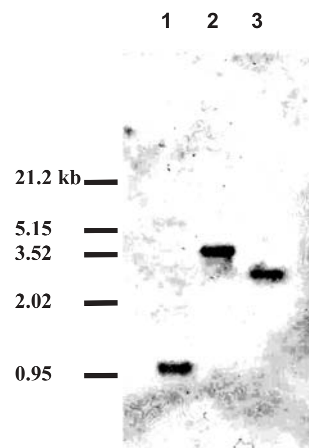

Fig. 2. Hybridization of a *cel3A* probe to *P. arcularius* genomic DNA. Genomic DNA (1 μg) from *P. arcularius* was digested with *Sal*I (*lane 1*), *Pst*I (*lane 2*), or *Xho*I (*lane 3*). The size markers consist of *Eco*RI/*Hin*dIII double digests of lambda DNA. The blot was probed with pF12–2 labeled with digoxigenin

Nucleotide sequence of *cel3A* from *P. arcularius*

The nucleotide sequence of *cel3A* is shown in Fig. 1. The respective sequences appear in DDBJ/EMBL/GenBank databases under the accession number AB187524. The coding region, from ATG to the stop codon, is 1329 bp in *cel3A* and encodes a protein of 243 amino acids. The locations of the exons and introns of the gene were determined from the nucleotide sequences of PCR products amplified by 3′- and 5′-RACE and RT-PCR. All the introns started with GT and ended with AG. The coding region was split into five exons by four introns.

Genomic Southern hybridization was performed using genomic DNA from *P. arcularius* monokaryon 69B-8 digested with *Sal*I, *Pst*I, and *Xho*I and using a probe generated from an initially amplified DNA fragment corresponding to a partial *cel3A* clone. The Southern blots showed a single band for each digest, indicating that *P. arcularius cel3A* is a single-copy gene in the haploid genome (Fig. 2).

Characterization of protein sequence of Cel3A

Three regions from the deduced protein sequence of *cel3A* contained regions similar to peptides P1, P2, and P3. Some differences were found between these peptides and the deduced internal amino acid sequence from *cel3A*, which may be due to low accuracy of the protein sequence analysis.

Fig. 3. Comparison of the deduced amino acid sequences of the *P. arcularius* Cel3A protein with those from various filamentous fungi (Ooi et al. 1990; Goedegebuur et al. 2002). Amino acid residues identical of the Cel3A protein are indicated by *dots*. Consensus boxes in fungal family 12 glycosyl hydrolases (Goedegebuur et al. 2002) are *boxed*

```
                                                                          Box1
P. arcularius    1:MHFSTLATLV AVAATASAQ- -T-------- ----LSGQYD CAPAGAYTLC QNLWGESSGV GNQNSTLIST  56
A. aculeatus     1:.KAFH.LAAL .G..V.---- -------QQ- -AQ-.CD..A TYTG.V..IN N....KDA.S .S.-C.TVNS  55
T. viride        1:.K.LQI.PTL LPV.L.---- -------QS- ----SCS..A TFSG.N.A.S N....Q.A.S .SG-CITDVS  53
F. equiseti      1:.KS-..LLAG .F.PL--.-F ---AKD---- ----.CE..G YLSSDG.S.N N.V..KD..T .D-QC.HVNW  54
H. grisea        1:.LK.A.LLGA .AVSVQ..SI P.IPANLEPR QIRS.CEL.G YWSGNG.E.L N....KDTAT SGWQC.YLDG  70

                                Box4                                              Box5
P. arcularius   57:SGNTVS-WRT QWQWQNNPNN VKSYANLLSN SA-KGVQLSA VRAAP-T-AW QWEYESKSDG IRADVSYDIW 122
A. aculeatus    56:ASSAGTS.S. K.N.SGGE.S .......SGLT FNK.-L-V.Q ISQ-IP.T.R -.S.DNTG-- .....A..LF 119
T. viride       54:L.GSAV-.S. T.D.SGGQS. ..G.P.IAL. IPN.RL-V.S ISS-MP.T.Q -.S.SGS.-- .....A..LF 117
F. equiseti     55:NNANGAG.DV E.N.SGGKD. ....P.SALL IGEDKK-TIS SITNMQST.E -.K.SGDN-- L....A..LF 120
H. grisea       71:TN.GGIQ.S. A.E..GA.D. ....PYVGKQ IQRGRK-I.D -INSMR.SVS -.T.DRTD-- ...N.A..VF 135

                             Box2
P. arcularius  123:LGTAPSGDPA SRASSYEIMI WLSGLGGIQP VGSKILSGVN VAGHTWDLWK GPNSNWQVLS FVSSTGDITD 192
A. aculeatus   120:TAADINHVTW .G--D..L.. ..ARY..V.. I..Q.ATAT- .D.Q..E..Y .A.GSQKTY. ..AP.PITSF 186
T. viride      118:TASN.NHVTY .G--D..L.. ..GKY.D... I..SQGTVN- .G.TS.N..Y ...GSM..Y. ..APGNLTNW 184
F. equiseti    121:TAAD.NHETS .G--E..L.V ..ARI..V.. I..LQT.VT- IE....E..V .M.GSMK.F. ..AP.PVN-N 186
H. grisea      136:TARD.DHPNW GG--D..L.. ..ARY...Y. I.TFHSQVN- L..R.....T .Y.G.MR.Y. .LPPS...R. 202

                         Box6                  Box3
P. arcularius  193:FNVDLKDFFN YLTQSQGVAA SQ-YVQAIQT GTEPFTGSAS LFTKAYSVAI NQ                       243
A. aculeatus   187:QGD-VN...K ....NH.FP. .SQ.LITL.F .......GPA TLSVSNWS.S V.                       237
T. viride      185:SGD-V.N.YT ..QNNK.YP. .SQ..LSY.F ...A....GT -LNNT-WT.S IN                       233
F. equiseti    187:..A.I.Q.WD ...K..NFP. DNQ.LLTF.F .......DNA K..VTNFN.H LK                       238
H. grisea      203:.SC.I...... ..ERNH.YP. REQNLIVY.V ...C...GPA R..CRDFR.D LW                       254
```

We next examined the predicted protein sorting signals. The program PSORTII (Nakai and Kanehisa 1992) predicted that the Cel3A protein was extracellular protein (66.7% probability) and that the first 18 amino acids form an N-terminal signal sequence (see Fig. 1). After removal of the signal peptide, the molecular mass of the putative mature protein would be 24 270 Da, which is nearly identical to the molecular mass of the purified *P. arcularius* 69B CMCase IIIa determined by SDS-PAGE (24.3 kDa; Ishihara et al. 2005). Based on these results, it is possible that the Cel3A protein is *P. arcularius* CMCase IIIa.

Figure 3 shows an alignment of the deduced amino acid sequence for Cel3A and homologous proteins from other filamentous fungi that belong to the family 12 glycosyl hydrolases. The Cel3A protein has 43% amino acid identity and 60% similarity with FI-CMCase of *A. aculeatus* (Ooi et al. 1990), 42% identity and 60% similarity with Cel12A from *T. viride*, 40% identity and 59% similarity with Cel12A from *F. equiseti*, and 42% identity and 59% similarity with Cel12A of *H. grisea* (Goedegebuur et al. 2002). Six consensus boxes in fungal family 12 glycosyl hydrolases are present in the Cel3A protein. Thus, the Cel3A protein appears to be an endoglucanase and a member of the family 12 glycosyl hydrolases (Goedegebuur et al. 2002).

Expression of *cel3A* cDNA in *E. coli*

A full-length synthetic *P. arcularius cel3A* cDNA was constructed and cloned into pET-26b (+) to create expression construct pET-*cel3A*. We tried to express recombinant Cel3A (rCel3A) without a C-terminal His tag (encoded by pET-*cel3A*) in *E. coli* BL21 (DE3). When the *E. coli* were induced under standard conditions (for 4 h, 37°C with 1 mM IPTG), we found no CMCase activity in the crude extract using the colorimetric assay. Thus, we attempted to optimize the induction temperature for improved activity. We found maximum CMCase activity (0.33 U/mg protein) in the colorimetric CMCase assay following induction for 4 h at 25°C with 1 mM IPTG (Table 2).

Table 2. Production of the recombinant Cel3A in *Escherichia coli* BL21(DE3)

Temperature (°C)	Specific activity (U/mg protein)[a]
20	0.069
25	0.33
30	0.019
37	n.d.

[a] CMCase activity was determined by the colorimetric enzymatic assay; n.d., no activity detected

To confirm the expression of CMCase IIIa protein, we performed zymography following native PAGE (Fig. 4). A clear zone was observed in the lane containing crude cell extract from *E. coli* carrying pET-F12 and was not observed in the lane containing crude cell extract from *E. coli* carrying vector only (pET26b (+)). These results confirmed that rCel3A was expressed as an active form in *E. coli*.

Enhancement of CMCase activity with cellobiose

Previous studies showed that the activity of purified *P. arcularius* CMCase IIIa is enhanced approximately fivefold by the addition of enzyme reaction products such as cellobiose (Ishihara et al. 2005). To confirm that CMCase activity of rCel3A expressed in *E. coli* was enhanced by reaction products, we therefore examined the effect of 20 mM cellobiose on the activity of rCel3A in the viscometric assay (Table 3). This assay showed that the activity of rCel3A was enhanced 4.47 fold by 20 mM cellobiose. Therefore, rCel3A has nearly the same enzymatic characteristics as native CMCase IIIa.

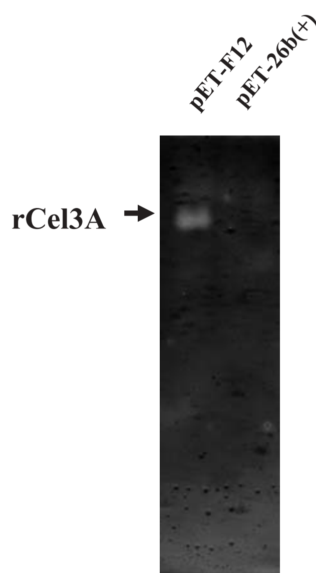

Fig. 4. Zymography of recombinant Cel3A. Ten micrograms (μg) of the purified enzyme was separated by native polyacrylamide gel electrophorcsis (PAGE) and analyzed by activity staining with CMC as the substrate

Table 3. Enhancement of native and the recombinant Cel3A activity from *Polyporus arcularius* by cellobiose

Enzyme	Cellobiose (mM)	Relative activity (%)[a]
CMCase IIIa	0	100
(from *P. arcularius*)	20	522
Recombinant Cel3A	0	100
(from *E. coli*)	20	447

[a] CMCase activity was determined by the viscometric method

Discussion

Previous studies have examined the effect of carbohydrate oligomers on native fungal cellulase, but the effects on recombinant enzyme have not been assessed (Denman et al. 1996). Therefore, in the current studies, we produced recombinant Cel3A in *E. coli* to directly determine whether the enzyme activity of Cel3A is affected by addition of carbohydrate polymers to the protein molecule. Preliminary experiments showed that the optimal induction temperature for producing active Cel3A was 25°C. Significant CMCase activity was not found in the cell extract from the *E. coli* induced at 30° or 37°C, probably because the protein may not fold correctly due to an excessive rate of transcription and translation. On the other hand, at 20°C, *E. coli* growth might be too low. Viscometry showed that the activity of Cel3A was enhanced 4.47 fold by the addition of cellobiose, which is very similar to the behavior of native CMCase IIIa (Ishihara et al. 2005). These results indicate that glycosylation and other postranslational modifications, such as proteolytic processing, are not required for Cel3A activity.

In this study, we found that approximately 30% of the rCel3A activity was in the *E. coli* culture broth and that this could degrade CMC in solid agar medium (data not shown). Therefore, the recombinant *E. coli* expressing *P. arcularius cel3A* cDNA might be better for the industrial decomposition of cellulose.

One potential N-glycosylation site (Asn^{50}) was present in the Cel3A sequence, as predicted from the tripeptide sequence (Asn-Xaa-Ser/Thr, where Xaa is any residue), and potential O-glycosylation sites were not found (Denman et al. 1996). In addition, the molecular mass of the putative mature protein after proteolytic removal of the signal peptide was predicted to be 24270 Da, which is identical to the mass of the purified *P. arcularius* 69B CMCase IIIa (24.3 kDa) as determined by SDS-PAGE (Ishihara et al. 2005). This result supports the idea that native *P. arcularius* CMCase IIIa is not glycosylated or subjected to other posttranslational modifications, such as proteolytic processing.

This is the first report characterizing a gene encoding a mushroom family 12 glycosyl hydrolase. The Cel3A protein contains six consensus boxes of the fungal family 12 glycosyl hydrolases, but the whole protein sequence was less than 50% identical to family 12 glycosyl hydrolases from other filamentous fungi. This result suggests that the consensus boxes are related to the active site in fungal family 12 glycosyl hydrolases and that other nonhomologous regions in Cel3A are related to the enhancement of enzyme activity by cellobiose. Thus, in future studies, we will use site-directed mutagenesis to examine the role of the nonhomologous regions in the enhancement of enzyme activity by reaction products.

References

Béguin P (1983) Detection of cellulase activity in polyacrylamide gels using Congo red-stained agar replicas. Anal Biochem 131:333–336

Dalboge H, Heldt-Hansen HP, Rasmussen G (1993) A method of cloning proteins in yeast and cellulase from *Humicola insolens*. Patent WO 93/11249

Davis BJ (1964) Disk electrophoresis. II. Method and application to human serum proteins. Ann NY Acad Soc 121:404–413

Denman S, Xue GP, Patel B (1996) Characterization of a *Neocallimastix patriciarum* cellulase cDNA (*celA*) homologous to *Trichoderma reesei* cellobiohydrolase II. Appl Environ Microbiol 62:1889–1896

Enokibara S, Mori N, Kitamoto Y (1992) Purification and some properties of a carboxymethyl cellulase from *Favolus arcularius.* J Ferment Bioeng 73:230–232

Goedegebuur F, Fowler T, Phillips J, van der Kley P, van Solingen P, Dankmeyer L, Power SD (2002) Cloning and relational analysis of 15 novel fungal endoglucanases from family 12 glycosyl hydrolase. Curr Genet 41:89–98

Gupta R, Jung E, Gooley AA, Williams KL, Brunak S, Hansen, J (1999) Scanning the available *Dictyostelium discoideum* proteome for O-linked GlcNAc glycosylation sites using neural networks. Glycobiology 9:1009–1022

Gupta R, Jung E, Brunak S (2002) Prediction of N-glycosylation sites in human proteins. (in preparation)

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, Davies G (1997) Structural and sequence-based classification of glycoside hydrolases. Curr Opin Struct Biol 7:637–644

Ishihara H, Imamura K, Kita M, Aimi T, Kitamoto Y (2005) Enhancement of the viscometric endocellulase activity of *Polyporus*

arcularius CMCase IIIa by cellobiose and cellooligosaccharides. Mycoscience 46:148–153

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (Lond) 227:680–685

Lecellier G, Silar P (1994) Rapid methods for nucleic acids extraction from Petri dish-grown mycelia. Curr Genet 25:122–123

Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951) Protein measurement with the Folin phenol reagent. J Biol Chem 193:265–275

Murai T, Ueda M, Kawaguchi T, Arai M, Tanaka A (1998) Assimilation of cellooligosaccharides by a cell surface-engineered yeast expressing beta-glucosidase and carboxymethyl cellulase from *Aspergillus aculeatus*. Appl Environ Microbiol 64:4857–4861

Nakai K, Kanehisa M (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics 14:897–911

Nelson N (1944) A photometric adaptation of Somogyi method for the determination of glucose. J Biol Chem 153:375–380

Parker JD, Rabinovitch P, Burmer GC (1991) Targeted gene walking polymerase chain reaction. Nucleic Acids Res 19:3055–3060

Ooi T, Shinmyo A, Okada H, Hara S, Ikenaka T, Murao S, Arai M (1990) Cloning and sequence analysis of a cDNA for cellulase (FI-CMCase) from *Aspergillus aculeatus*. Curr Genet 18:217–222

Sakamoto S, Tamura G, Ito K, Ishikawa T, Iwano K, Nishiya N (1995) Cloning and sequencing of cellulase cDNA from *Aspergillus kawachii* and its expression in *Saccharomyces cerevisiae*. Curr Genet 27:435–439

Somogyi M (1959) Notes on sugar determination. J Biol Chem 195:19–23

Triglia T, Peterson MG, Kemp DJ (1988) A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequence. Nucleic Acids Res 16:8166

Ward M, Wu S, Dauberman J, Larenas E, Bower B, Rey M, Clarkson K, Bott R (1993) Cloning sequencing and preliminary structural analysis of a small high pI endoglucanase (EGIII) from *Trichoderma reesei*. In: Souminen P, Reinikainen T (eds) Proceedings, 2nd TRICEL symposium on *Trichoderma reesei* cellulases and other hydrolases. Foundation of Biotechnical and Industrial Fermentation Research, Espoo, pp. 153–158

Wood BE, Ingram LO (1992) Ethanol production from cellobiose, amorphous cellulose, and crystalline cellulose by recombinant *Klebsiella oxytoca* containing chromosomally integrated *Zymomonas mobilis* genes for ethanol production and plasmids expressing thermostable cellulase genes from *Clostridium thermocellum*. Appl Environ Microbiol 58:2103–2110