Appl Microbiol Biotechnol (1996) 46 : 538-544

N. Kitamoto · M. Go · T. Shibayama · T. Kimura
Y. Kito · K. Ohmiya · N. Tsukagoshi

# Molecular cloning, purification and characterization of two endo-1,4-β-glucanases from *Aspergillus oryzae* KBN616

Received: 3 July 1996 / Accepted: 15 July 1996

**Abstract** Two endo-1,4-β-glucanase genes, designated *celA* and *celB*, from a shoyu koji mold *Aspergillus oryzae* KBN616, were cloned and characterized. The *celA* gene comprised 877 bp with two introns. The CelA protein consisted of 239 amino acids and was assigned to the cellulase family H. The *celB* gene comprised 1248 bp with no introns. The CelB protein consisted of 416 amino acids and was assigned to the cellulase family C. Both genes were overexpressed under the promoter of the *A. oryzae* taka-amylase A gene for purification and enzymatic characterization of CelA and CelB. CelA had a molecular mass of 31 kDa, a pH optimum of 5.0 and temperature optimum of 55 °C, whereas CelB had a molecular mass of 53 kDa, a pH optimum of 4.0 and temperature optimum of 45 °C.

## Introduction

The filamentous fungus *Aspergillus oryzae* has been used in Japan for food production, such as soy sauce or shoyu, miso and sake, and has also been used for industrial enzyme production, such as α-amylases and proteases. *A. oryzae*, when grown on a shoyu koji composed of soybean and wheat, secretes soybean-cell-wall-degrading enzymes including cellulases, pectinases, xylanases and hemicellulases. These enzymes synergistically degrade the soybean cell wall components such as cellulose, pectin and hemicellulose. The genetic and enzymatic characterization of these enzymes could elucidate their roles in soybean cell wall degradation, which would result in an increased utilization of soybean protein and a decrease in the pressed cake obtained through the press filtration of soy sauce mash.

Cellulose, which is the world's most abundant biopolymer, is the major component of plant cell walls. Biological breakdown of cellulose is catalysed by the cellulase enzyme complex, consisting of three classes of enzymes, endo-1,4-β-glucanase (EC 3.2.1.4), cellobiohydrolase (exo-1,4-β-glucanase, EC 3.2.1.91) and β-glucosidase (EC 3.2.1.21). Numerous studies on fungal cellulases have been done and several fungal genes encoding endo-1,4-β-glucanase have been isolated from *Aspergillus aculeatus* (Ooi et al. 1990), *Aspergillus kawachii* (Sakamoto et al. 1995), *Fusarium oxysporum* (Sheppard et al. 1994), *Humicola insolens* (Davis et al. 1993; Dalbøge and Heldt-Hansen 1994) and *Trichoderma reesei* (Penttilä et al. 1986; Saloheimo et al. 1988, 1994). However, little is yet known about the structures and genes of the *A. oryzae* endo-1,4-β-glucanases. Here we describe the cloning and sequence analysis of two endo-1,4-β-glucanase genes, *celA* and *celB*, from an industrial shoyu koji mold strain, *A. oryzae* KBN616. In addition, these genes were overexpressed under the promoter of the *A. oryzae* taka-amylase A gene (*Taa*, Tsukagoshi et al. 1989), purified to homogeneity and characterized for their enzymatic properties.

## Materials and methods

### Strains, plasmids, media and transformation

*A. oryzae* KBN616, obtained from Bio'c (Toyohashi, Japan), was used for isolation of DNA and grown at 30 °C in Czapex-Dox medium. *A. oryzae* KBN616-39 (*niaD*$^-$), a derivative of KBN616, was used for transformation and grown at 30 °C in the modified Czapex-Dox medium containing $NH_4Cl$ as the nitrogen source instead of $NaNO_3$. Plasmid pND300 and its derivatives, which contain the *A. oryzae niaD* gene (Kitamoto et al. 1995), were used for transformation of *A. oryzae* KBN616-39 (*niaD*$^-$). *A. oryzae* transformations were carried out as described by Ballance and

N. Kitamoto (✉) · M. Go · T. Shibayama · T. Kimura · Y. Kito
Food Research Institute, Aichi Prefectural Government,
2-1-1 Shinpukuji-cho, Nishi-ku, Nagoya 451, Japan.
Fax +81 52 532 5791

K. Ohmiya
Faculty of Bioresources, Mie University, Tsu 504, Japan

N. Tsukagoshi
Department of Applied Biological Science, School of Agriculture,
Nagoya University, Nagoya 464, Japan

Turner (1985). NiaD$^+$ transformants grew after 7–10 days at 30 °C on plates with nitrate as sole nitrogen source. Transformants were purified through conidia. Starch medium (2% starch, 1% polypeptone, 0.5% $KH_2PO_4$, 0.5% KCl, 0.1% $NH_4Cl$, 0.05% $MgSO_4$) was used for cellulase production by *A. oryzae* transformants. *Escherichia coli* DH5α and MV1184 were used for DNA manipulation. pUC plasmids were used to subclone various DNA fragments and also to prepare single-stranded plasmid DNAs.

Strains of plasmid-carrying *E. coli* were grown at 37 °C in Luria-Bertani medium with 50 μg ampicillin/ml. *E. coli* transformations were carried out as described by Hanahan (1983).

## Isolation of chromosomal DNA and construction of a genomic library

The chromosomal DNA, isolated from *A. oryzae* as described previously (Kitamoto et al. 1993), was partially digested with *Sau*3AI. Fragments of 10–23 kb were isolated by electroelution from an agarose gel and ligated into the *Bam*HI site on Charomid 9–28 cosmid DNA (Nippon Gene, Toyama, Japan). The DNA was packaged in vitro and recombinant phages were infected to *E. coli* DH5α. The genomic library contained about $5 \times 10^5$ clones with an average insert size of 18 kb. This library was screened with the digoxigenin-labelled polymerase-chain-reaction (PCR)-amplified genomic DNA fragments described below. The DNA of cosmid clones that hybridized with the probes were analysed by Southern hybridization and hybridizing fragments were subcloned into pUC plasmids.

## Amplification of *celA* and *celB* genomic DNA sequences by PCR

A genomic DNA fragment encoding a portion of the *A. oryzae celA* gene was amplified by the method of semi-nested PCR. Three oligonucleotide primers were designed, based on three amino acid sequences conserved well between *A. aculeatus* F1 carboxymethylcelullase (CMCase) and *Erwinia carotovora* celS (Saarilahti et al. 1990). The sense primer, P1 [5′-AA(C/T)AA(C/T)CT(G/C)-TGGGG(C/T)AA(A/G)GA-3′], was homologous to the sense strand for amino acid residues 37–43 of *A. aculeatus* F1-CMCase. The antisense primers, P2 [5′-CCA(A/G)ATCAT(A/G)AG(C/T)-TC(A/G)TA(G/A)TC-3′] and P3 [5′-GC(A/G)GT(A/G)AA(C/G)-AG(A/G)TC(A/G)TA(A/G)GA-3′], were complementary to the sense strand for amino acid residues 115–121 and 132–138 of *A. aculeatus* F1-CMCase respectively. A 254-bp fragment was specifically amplified from genomic DNA by combinations of three primers, the first and second rounds of PCR using P1–P3 and P1–P2 respectively. The amplified 254-bp fragment cloned on pUC118 was sequenced as described by Sanger et al. (1977) and found to contain an open-reading frame encoding 84 amino acids bearing a relatively high homology to the sequence of *A. aculeatus* F1-CMCase.

A genomic DNA fragment encoding a portion of the *A. oryzae celB* gene was also amplified by semi-nested PCR. Three oligonucleotide primers were designed, based on three amino acid sequences conserved well among the family of C cellulases such as *T. reesei* CBHI and *T. reesei* EGI. The sense primers, P4 [5′-GA(A/G)TT(T/C)AC(T/C)TT(T/C)GA(T/C)GT(C/G)GA-3′] and P5 [5′-TG(T/C)GG(T/C)CT(T/C)AA(T/C)GG(T/C)GC(T/C)(T/C)T-3′], were homologous to the sense strand for amino acid residues 143–149 and 155–161 of *T. reesei* CBHI respectively. The antisense primer, P6 [5′-GC(C/T)TCCCA(A/G)AT(A/G)TCCAT(T/C)TC-3′], was complementary to the sense strand for amino acid residues 229–235 of *T. reesei* CBHI. A 191-bp fragment was amplified specifically from genomic DNA by combinations of three primers, the first and second rounds of PCR using P4–P6 and P4–P5. The amplified 191-bp fragment was subcloned, sequenced and found to contain an open-reading frame encoding 63 amino acids bearing a relatively high homology to those of the fungal family of C cellulases.

## Construction and expression of the *A. oryzae celA* and *celB* genes under the control of the *A. oryzae* taka-amylase gene (*Taa*) promoter

To construct the *celA* and *celB* genes, expression plasmids under the control of the *A. oryzae* taka-amylase (*Taa*) promoter and the promoter regions of *celA* and *celB* were substituted for the *Taa* promoter region as follows. A 0.6-kb *Taa* promoter, extending from the 5′ end to the initiator ATG, was amplified by PCR, using the *Taa* gene as template, the M13 reverse primer and the antisense primer, TaaP (5′-CCGAGCTCATGCATAAATGCCTTCTGTGG-3′), as primers. TaaP included cleavage sites for *Eco*T22I and *Sac*I at the 5′ end. The PCR-amplified *Taa* promoter was digested with *Eco*RI and *Sac*I and cloned into pUC119 digested with *Eco*RI and *Sac*I to construct plasmid pTAP200.

In order to ligate the *Taa* promoter fragment precisely next to the *celA* and *celB* coding regions, an *Eco*T22I cleavage site was introduced just before the ATG codon of both the *celA* and *celB* genes by PCR. The 0.6-kb *Taa* promoter fragment excised from pTAP200 with *Eco*RI and *Eco*T22I digestion and the 1.3-kb *celA* gene excised from pTAE400 with *Eco*T22I and *Eco*RI digestion were ligated and cloned into the *Eco*RI site on pUC118 to create plasmid pTACA100, which contains the *celA* gene under the control of the *Taa* promoter. A 5.1-kb *Hin*dIII fragment, containing the *A. oryzae niaD* gene from plasmid pND300, was subcloned into pTACA100, resulting in plasmid pTACA200. Plasmid pTACA200 was transformed to the *niaD*-deficient *A. oryzae* strain KBN616-39.

The *Eco*T22I and *Sac*I-digested 1.9-kb *celB* gene was ligated to *Eco*T22I and *Sac*I sites on pTAP200 to construct pTACB100, which was cotransformed to *A. oryzae* KBN616-39 together with pND300.

## Enzyme assay

Endoglucanase (carboxymethylcellulase) activity was determined by evaluating the viscosity changes of a reaction mixture containing carboxymethylcellulose (CMC, degree of substitution, 0.6; molecular mass, 180 kDa; Cellogen WS-C; Daiichi Seiyaku Kogyo, Kyoto) at 40 °C and was monitored for 5 min by using a cone plate-type viscometer (Tokisangyo Co.). The reaction mixture contained CMC solution (1%) in 50 mM sodium acetate buffer (5 ml; pH 5.0 for CelA and pH 4.0 for CelB) with enzyme solution (1 ml). One unit of enzyme activity was defined as the increase of 1 fluidity unit [reciprocal of the viscosity: $mPa^{-1}\,S^{-1}$ (centipoise$^{-1}$)] of the CMC solution in 1 min (Ohmiya et al. 1987).

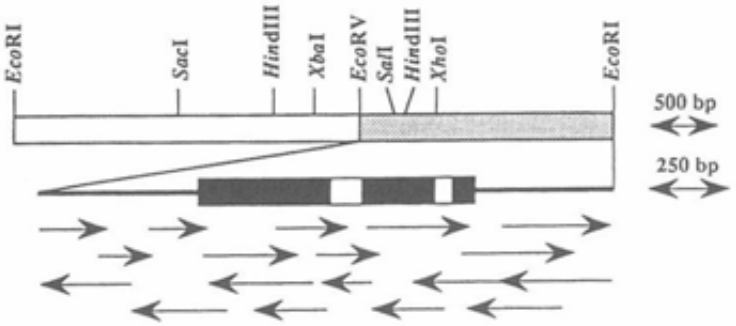
## Purification of the CelA from *A. oryzae* transformant TA31

A 610-ml sample of culture supernatants was harvested after 4 days growth and dialysed against 10 mM TRIS/HCl buffer, pH 7.5. Protein was adsorbed to a DEAE-Toyopearl 650M column ($3.2 \times 9.0$ cm) equilibrated in the same buffer and eluted by pulse elution with the same buffer containing 0.1 M NaCl. The CelA-containing fractions were diluted ten times with the same buffer and loaded on a DEAE-Toyopearl 650 M column ($3.2 \times 9.0$ cm). Protein was eluted with a linear gradient from 0 to 0.2 M NaCl.

## Purification of the CelB from *A. oryzae* transformant TB1

A 150-ml sample of culture supernatants was harvested after 4 days growth and dialysed against 10 mM sodium acetate buffer pH 4.0. Protein was adsorbed batchwise to SE-Sephadex C50 and eluted by pulse elution with 10 mM sodium acetate buffer pH 6.0. The CelB-containing fractions were diluted ten times with 10 mM sodium acetate buffer pH 4.0 and loaded on a SE-Sephadex C50 column ($3.2 \times 7.0$ cm) equilibrated in the same buffer. Protein was eluted with a linear pH gradient from 4.0 to 6.0 in 10 mM sodium acetate buffer.

A




B

```
GATATCGATGTGAGATCACGGGCTGAAGGACTATATCCGAAGACGAAGTAAGACGCCGAATAAAATATACCTAAGTCTCCTTCCCTAGGT -411
GGATATGTAGAGTTCTTGACCAAACATCTTCAAGTATAGTGATACATGTGCCGATGGATAGTTAAAGGGTAGTAAGATTAGTCGGTATTT -321
CGGGCCTGCCAACCATGCGCGAACGTATGTCAACTGCAGATCGTGAATGAATTTCATTAATTCGAAAATTTCTGTCGACGTCGGAGAGGA -231
TGCTTCGGTAGGCACGGGAAGCTTGTGGATCATAGGGACAATCCAGAAAGTAGTTTACTGACGTATGCGATACCTTTGGAGGGTGAAGAC -141
TTGGAATATGTGGGTTGTGCTAGCAACCTGTCAGCTCTGTGCTTTTAAACACGTTGCCCGCTACTTCCAGACCTGCATTATCCATCAGTC  -51

TTGCAACACTATCCATTAGGCCAAATTAATCTACCACCTATACTTTCACCATGAAGCTCTCATTGGCACTTGCTACGCTCGTGGCCACAG   40
                                                  M  K  L  S  L  A  L  A  T  L  V  A  T  A   14

CATTCAGTCAAGAGCTGTGTGCACAGTACGACAGTGCCTCGAGCCCTCCGTACTCAGTGAACAACAACCTCTGGGGACAGGACTCGGGAA  130
  F  S  Q  E  L  C  A  Q  Y  D  S  A  S  S  P  P  Y  S  V  N  N  N  L  W  G  Q  D  S  G  T   44

CCGGCTCTCAGTGCGTCTACGTTGACAACCTCTCCAGCTCAGGCGCTGCGTGGCACACTACTTGGACCTGGAACGGTGGAGAAGGCAGTG  220
  G  S  Q  C  V  Y  V  D  N  L  S  S  S  G  A  A  W  H  T  T  W  T  W  N  G  G  E  G  S  V   74
                          #
TGAAAAGCTATTCCAACTCTGCAGTCACCTTCGACAAAAAGCTCGTGAGCGATGTCCAGAGTATCCCCACCGACGTGGAGTGGAGTCAGG  310
  K  S  Y  S  N  S  A  V  T  F  D  K  K  L  V  S  D  V  Q  S  I  P  T  D  V  E  W  S  Q  D  104

ATAACACTAATGTCAACGCCGATGTGGCGTACGACCTGTTTACCGCTGCGGACCAGAACCACGTCACCTATAGCGGCGATTATGAACTGA  400
  N  T  N  V  N  A  D  V  A  Y  D  L  F  T  A  A  D  Q  N  H  V  T  Y  S  G  D  Y  E  L  M  134

TGATTTGgtacgtatgaaccgatcactcctgtaagcgatcaatcctaccctacatcggcagtgaggaaccttcaactgacataaccgggg  490
  I  W                                                                                        136

ttttccatttgcaagGCTCGCACGCTACGGTACTATTCAGCCCATCGGAACCCAAATCGATACCGCTACAGTTGAAGGCCATACCTGGGA  580
                L  A  R  Y  G  T  I  Q  P  I  G  T  Q  I  D  T  A  T  V  E  G  H  T  W  E  161

ACTGTGGTACGGCACCACCATCCAGGCAGGCGCCGAGCAGAAAACGTACAGCTTCGTCTCGGCAACCCCAATCAACACCTTCGGCGGGGA  670
 L  W  Y  G  T  T  I  Q  A  G  A  E  Q  K  T  Y  S  F  V  S  A  T  P  I  N  T  F  G  G  D  191

CATTAAAAAGTTTTTCGACTATATCACATCTAAGCACAGTTTCCCTGCTTCTGCCCAGTACTTGATCAgtaagagcccacatatccctca  760
 I  K  K  F  F  D  Y  I  T  S  K  H  S  F  P  A  S  A  Q  Y  L  I  N                          214

ccccccgtgcaagcaaagccactaacaagggcaatttatagATATGCAGTTCGGAACTGAGCCGTTTACCGGTGGCCCTGTTACTTTCACA  850
                                                  M  Q  F  G  T  E  P  F  T  G  G  P  V  T  F  T  230

GTTCCCAACTGGACTGCCAGTGTCAACTAAGAATATGAATGCCTCTTGCCCAATTGACAAATGGGGTAGGCAGTCCGTGGCTAGGTGGAG  940
V  P  N  W  T  A  S  V  N  ***                                                                239
    #
TTGGACATGGTCTGCTTAGTAACAATTTGTAGAGAGTTGTTTGATTTGCTTCGAAGTTTTGGGGGTCGCGATTGATTCTCCCCTATTAAT 1030
TCGAGTCTAGCAAGAAGACCATATAACTGTTCATCTGTTAGACGCAAGAAAATTTAGTTAGTATTTACCATTCGCGTCCGCTCATTCAAT 1120
CAGAGCTTGCTATATTTAATTTGGTTACTGGCAACCATCTGCTAAGCCGCTTACTGTTCCTGCTGCTCATCGACTCTTTGGCCTCTTCCA 1210
TGAGTGACTTCTAACTGACGACATGTATGTAATCTTCGGATGTTTGGCTCTGGAGGTCCTGGATAGAAGATAACATGCGTTCAGACGTCG 1300
AGCTTCAAGGGCCGAATTC                                                                         1319
```

**Fig. 1A,B** Restriction map and the nucleotide sequence of the *celA* gene. **A** Restriction map of the *celA* gene. The sequencing strategy for the *celA* gene is outlined below the restriction map by arrows. The sequencing region is indicated by a *hatched box*. The coding region is indicated by solid (exons) and open (introns) boxes. **B** Nucleotide and deduced amino acid sequences of the *celA* gene. Putative intron sequences are in lower-case letters. *** The translation stop codon. Numbers on the right refer to nucleotide sequence (negative numbers refer to nucleotides upstream from the *celA* ATG) and amino acid sequence. # Two potential *N*-glycosylation sites. A putative termination sequence [TAG...TA(T)GT...TTT] is underlined. This sequence has been deposited in the DDBJ, EMBL and GenBank nucleotide sequence databases under the accession number D83731

## Results

### Isolation and characterization of the *A. oryzae celA* gene

A genomic library of *A. oryzae* KBN616, constructed in Charomid 9–28, was screened with the 254-bp genomic DNA fragment described above. Of $5.3 \times 10^4$ colonies screened, five positive clones were found to contain a 4.3-kb *Eco*RI fragment, which was hybridized to the probe on Southern blot analysis and was further characterized in detail. The *celA* gene was present on a 1.8-kb *Eco*RV-*Eco*RI restriction fragment, which was sequenced in both directions (Fig. 1A). The coding region consisted of 877 bp and was interrupted by two introns of 98 bp and 62 bp (Fig. 1B). These two introns had characteristics typical of filamentous fungal genes, such as consensus 5′ and 3′ splice-junction sequences and putative lariat-formation internal sequences. The positions of two introns are the same as those of the *A. aculeatus* F1-CMCase gene. The coding sequence comprised 239 amino acids with a molecular mass of 26 096Da. Two potential *N*-glycosylation sites (Asn-Xaa-Thr/Ser-Zaa, where Zaa is not Pro) were found, one of which ($Asn^{233}$-Trp-$Thr^{235}$) existed at the same position as that of the *A. aculeatus* F1-CMCase and *A. kawachii* CMCase I.

### Isolation and characterization of the *A. oryzae celB* gene

A genomic library of *A. oryzae* KBN616 was also screened with the 191-bp genomic DNA fragment described above. Of $3.6 \times 10^4$ colonies screened, seven positive clones were found to contain a 9.2-kb *Hin*dIII fragment, which was hybridized to the probe on Southern blot analysis and was further analysed in detail. The *celB* gene was present on a 2.3-kb *Sma*I-*Hin*dIII restriction fragment, which was sequenced in both directions (Fig. 2A). The coding region consisted of 1248 bp and contained no introns (Fig. 2B). The coding sequence comprised 416 amino acids with a molecular mass of 42 623Da. Eight potential *N*-glycosylation sites were found and some of them appeared to be glycosylated as judged by sodium dodecyl sulphate/polyacrylamide gel electrophoresis (SDS-PAGE), since CelB formed a rather broad band (Fig. 3).

### Comparison of the *A. oryzae* KBN616 CelA and CelB with other cellulases and xylanases

The deduced amino acid sequence of the *A. oryzae* KBN616 CelA was compared to that of *A. aculeatus* F1-CMCase and appeared to contain an N-terminal extension of 16 amino acids before the N terminus ($Gln^{17}$) of the mature protein. The deduced amino acid sequence of the *A. oryzae* KBN616 CelA revealed a significant homology to those of the family H cellulases (Henrissat and Bairoch 1993) such as *A. aculeatus* F1-CMCase (63%), *A. kawachii* CMCase I (77%) and *E. carotovora* CelS (31%).

Several sequences highly conserved among the family H cellulases might be involved in the catalytic reaction. On the basis of the three-dimensional structure and site-directed mutagenesis, two totally conserved glutamate residues have been identified as catalytic residues of the *Bacillus pumilus* xylanase (Okada et al. 1989; Ko et al. 1992). The three-dimensional structural similarity between the *A. aculeatus* F1-CMCase and the *B. pumilus* xylanase indicates that two glutamate residues ($Glu^{134}$ and $Glu^{218}$) of the *A. aculeatus* F1-CMCase are involved in the enzymatic activity of this enzyme (Törrönen et al. 1993). As two glutamate residues are conserved among all the family H cellulases, the corresponding residues ($Glu^{132}$ and $Glu^{220}$) of the *A. oryzae* KBN616 CelA should also be involved in the catalytic reaction.

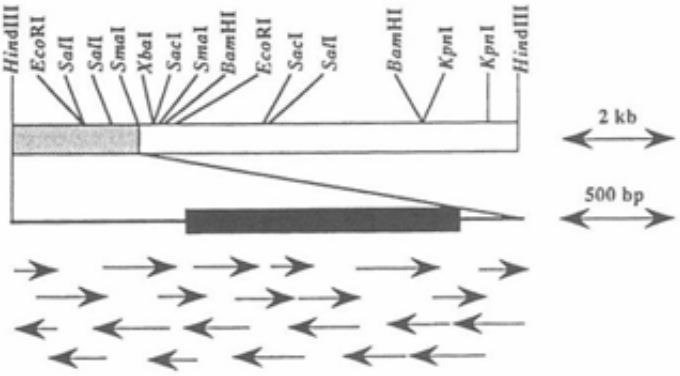
The deduced amino acid sequence of the *A. oryzae* KBN616 CelB was compared to those of the *T. reesei* CBHI (Shoemaker et al. 1983) and the *T. reesei* EGI and appeared to contain an N-terminal extension of 17 amino acids before the N terminus ($Gln^{18}$) of the mature protein. The deduced amino acid sequence of the *A. oryzae* KBN616 CelB showed a significant homology to those of the family C cellulases such as *F. oxysporum* Cfam1 (47%), *F. oxysporum* Cfam2 (39%). *Humicola grisea* CBHI (Azevedo et al. 1990; 40%), *Penicillium janthinellum* CBHI (Koch et al. 1993; 42%), *Phanerochaete chrysosporium* CBHI (Sims et al. 1988; 42%), *T. reesei* CBHI (39%) and *T. reesei* EGI (50%). Most noteworthy, CelB lacked both the hinge and cellulose-binding domains at the C terminus, a common feature of the family C cellulases.

### Overexpression, purification and characterization of *A. oryzae* CelA and CelB

In order to obtain high-level expression of *celA* and *celB* genes in *A. oryzae*, *A. oryzae* KBN616-39 (*niaD*$^-$) was either transformed with pTACA200, which contained the *celA* gene under the control of the *A. oryzae Taa* promoter and the *A. oryzae niaD* gene, or cotransformed with pTACB100 containing the *celB* gene under the control of the *A. oryzae Taa* promoter and pND300 carrying the *A. oryzae niaD* gene. Transformants grown with starch were assayed for cellulase activity to select the high-CelA- or high-CelB-producing strain for fur-

**Fig. 2A,B** Restriction map and the nucleotide sequence of the *celB* gene. **A** Restriction map of the *celB* gene. The sequencing strategy for the *celB* gene is outlined below the restriction map by arrows. The sequencing region is indicated by a hatched box. The coding region is indicated by a solid box. **B** Nucleotide and deduced amino acid sequences of the *celB* gene. *** The translation stop codon. Numbers on the right refer to nucleotide sequence (negative numbers refer to nucleotides upstream from the *celB* ATG) and amino acid sequence. # Eight potential *N*-glycosylation sites. A putative termination sequence [TAG...TA(T)GT...TTT] is underlined. This sequence has been deposited in the DDBJ, EMBL and GenBank nucleotide sequence databases under the accession number D83732

A




B

```
AAGCTTCTCCCACATTTGCCGTAGTGATATACTTACTTGGCATGCAGTGAACAAAAAGACACTTCTTAAGGAGGTCTTTGGCATTTAATC -709
GCATATGGACTGTCAGCACTGTACTCTATACTCGACTGTATCTATATCATCATTTTGATAAATTTCGCATCTTATTACCTTGGAATTATC -618
TTCACCCGCCTTTAAAATTAAGGGCCCTGTGGCTGCGCACTGCAATTGACTTTAGCGCGCGACAGAATAGGTATTGTCTATGCTACAAGA -528
GTTTGGGGCAACGGGAATTTCTTCAGCTGGACGGGATAGCATCTACCTTGGCAGCGGGGGTTCCTGATCTAAAGCAATCTGCAGCGGCTA -438
AATCCGGAGATCCACTCCGCACCCGTTACGGCGCTCAAGCACCGGCAAAACATACGACTTTTGCAAGATTGCAGCACAAATGTTTCGTTT -348
ATCCTCTACCAGTCTCAGTTTGACGCTACTGTCGCCATTTGAAGCCGTTGATCAGTCCCCTTCATGTCCTTAAGTCCGGCAACATCGCCG -258
ACGCGCATCCTTCGTGCCGGCTGATCTTAATAAAACTAAAACCGGATCATATTGGAGATGAAACGTCCGGGAAAAAAGTGCGTGAATCAG -168
ATCGGACTGTCGGGCCGGAATGTTTGCAGTCCCATATCTGGCACTCTGAAGATTCCACAAGCCTTGATCCGCTGTCTGTACGTGATGCAT  -78

AAATAACAGCTACTTCCCGGTGGAGAGTCGAGAGCACACATACTCTTTATACACTCTAGTAGCAAGAGGACCTCGATATGATCTGGACAC   13
                                                                                  M  I  W  T  L    5

TCGCTCCCTTTGTGGCACTCCTGCCACTGGTAACTGCCCAGCAGGTGGGAACTACAGCGGACGCCCATCCCAGACTCACCACGTATAAAT  103
  A  P  F  V  A  L  L  P  L  V  T  A  Q  Q  V  G  T  T  A  D  A  H  P  R  L  T  T  Y  K  C   35

GTACTTCACAGAACGGTTGCACGAGGCAGAACACCTCACTCGTCCTTGATGCAGCAACCCATTTTATCCACAAGAAAGGAACACAAACAT  193
  T  S  Q  N  G  C  T  R  Q  N  T  S  L  V  L  D  A  A  T  H  F  I  H  K  K  G  T  Q  T  S   65
                          #
CCTGCACCAACAGCAACGGCTTAGACACTGCCATTTGTCCGGACAAACAGACCTGCGCGGACAACTGTGTCGTTGATGGGATCACGGACT  283
  C  T  N  S  N  G  L  D  T  A  I  C  P  D  K  Q  T  C  A  D  N  C  V  V  D  G  I  T  D  Y   95

ACGCTAGCTACGGCGTCCAGACGAAGAATGACACGTTGACCCTTCAACAATACCTGCAAACTGGGAATGCCACAAAGTCCCTGTCACCGC  373
  A  S  Y  G  V  Q  T  K  N  D  T  L  T  L  Q  Q  Y  L  Q  T  G  N  A  T  K  S  L  S  P  R  125
                       #                                          #
GCGTCTACCTCCTCGCTGAAGACGGAGAGAACTATTCCATGCTGAAACTCCTGAATCAGGAATTCACCTTCGATGTCGACGCCTCCACCC  463
  V  Y  L  L  A  E  D  G  E  N  Y  S  M  L  K  L  L  N  Q  E  F  T  F  D  V  D  A  S  T  L  155
                          #
TCGTCTGCGGCATGAATGGTGCTCTATATCTCTCTGAAATGGAGGCTTCTGGCGGAAAGAGTTCCCTAAATCAAGCCGGAGCCAAATACC  553
  V  C  G  M  N  G  A  L  Y  L  S  E  M  E  A  S  G  G  K  S  S  L  N  Q  A  G  A  K  Y  G  185

GAACCGGTTACTGTGATGCCCAATGCTACACCACGCCTTGGATCAACGGCGAAGGCAACACCGAGAGTGTCGGTTCCTGCTGTCAGGAAA  643
  T  G  Y  C  D  A  Q  C  Y  T  T  P  W  I  N  G  E  G  N  T  E  S  V  G  S  C  C  Q  E  M  215

TGGATATTTGGGAAGCCAACGCCCGAGCAACAGGGCTTACACCACACCCTTGCAACACAACCGGTCTGTACGAGTGCAGCGGCTCAGGAT  733
  D  I  W  E  A  N  A  R  A  T  G  L  T  P  H  P  C  N  T  T  G  L  Y  E  C  S  G  S  G  C  245
                                                          #
GCGGAGACTCCGGGGTCTGTGACAAGGCCGGCTGTGGATTCAATCCATATGGCCTAGGCGCAAAGGACTACTACGGTTACGGTCTCAAGG  823
  G  D  S  G  V  C  D  K  A  G  C  G  F  N  P  Y  G  L  G  A  K  D  Y  Y  G  Y  G  L  K  V  275

TCAACACCAACGAGACATTCACTGTCGTAACTCAGTTCCTCACAAACGATAACACAACTTCGGGCCAGCTCAGCGAAATCCGCCGTCTCT  913
  N  T  N  E  T  F  T  V  V  T  Q  F  L  T  N  D  N  T  T  S  G  Q  L  S  E  I  R  R  L  Y  305
        #                                                   #
ATATCCAGAACGGCCAGGTCATTCAAAATGCTGCCGTTACCTCTGGAGGAAAAACTGTCGACTCAATCACAAAGGACTTCTGCAGCGGCG 1003
  I  Q  N  G  Q  V  I  Q  N  A  A  V  T  S  G  G  K  T  V  D  S  I  T  K  D  F  C  S  G  E  335

AAGGAAGTGCCTTCAACCGACTTGGCGGCCTCGAGGAAATGGGCCACGCCTTGGGCCGCGGCATGGTTCTTGCGCTCAGTATCTGGAACG 1093
  G  S  A  F  N  R  L  G  G  L  E  E  M  G  H  A  L  G  R  G  M  V  L  A  L  S  I  W  N  D  365

ATGCAGGCTCATTTATGCAATGGCTTGATGGTGGCAGTGCCGGACCGTGCAACGCAACGGAGGGAAACCCGGCGTTGATCGAGAAGTTGT 1183
  A  G  S  F  M  Q  W  L  D  G  G  S  A  G  P  C  N  A  T  E  G  N  P  A  L  I  E  K  L  Y  395
                                                          #
ATCCGGATACTCATGTGAAGTTTTCCAAGATTCGGTGGGGAGATATTGGATCTACCTACAGGCATTAGAATGTGGGATGAATCATCTACG 1273
  P  D  T  H  V  K  F  S  K  I  R  W  G  D  I  G  S  T  Y  R  H  ***                          416

GTTGTGTGTGTATTATATTCTGCAGCTGAATGTTATGTTTTGCGGATTGTAGTGAGGTTGTTCCATAAATATCTACATCCCGACTCTTTC 1363
ATACATACTTTCAATAGGTTCCATCCTATACGCCAAAGAAGTAAAATACTAGAGCCACAGCAATACCCTCACTTACTCGGCCACTCAAGA 1453
TCCCTCCATGGCCTCCCCTGAGTTTCCCTCGCCAACTGCTCCAAAGACTTCCACTTCCAAACCCGCTCCCGGG                  1526
```

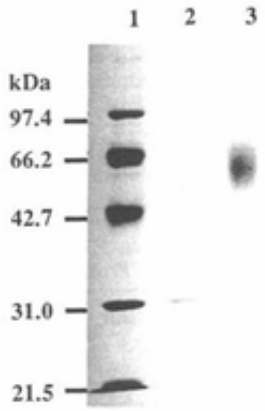



**Fig. 3** Sodium dodecyl sulphate/polyacrylamide gel electrophoretic profiles of two endo-1,4-β-glucanases purified from *Aspergillus oryzae* transformants. The gel was stained with Coomassie brilliant blue. *Lane 1* molecular size markers: (rabbit muscle phosphorylase *b*, 97.4 kDa; bovine serum albumin, 66.2 kDa; hen egg-white ovalbumin, 42.7 kDa; bovine carbonic anhydrase, 31.0 kDa; soybean trypsin inhibitor, 21.5 kDa). *Lane 2* purified CelA. *Lane 3* purified CelB

ther study. CelA and CelB were purified to homogeneity from culture supernatants of high-production transformants by anion-exchange chromatography for CelA and cation-exchange chromatography for CelB, as described in Materials and methods. CelA was purified 60.3-fold, with recovery of 51% of the initial activity. CelB was purified 29.5-fold, with recovery of 43% of the initial activity. CelA had a molecular mass of 31 kDa, which was approximately 7 kDa larger than the calculated value of 24 kDa (Fig. 3, lane 2). The molecular mass of CelB was also at least 10 kDa larger than the calculated value of 40 kDa (Fig. 3, lane 3).

The enzymatic features of CelA and CelB: pH and temperature optima and pH and thermal stabilities, were determined with the purified proteins (Table 1). The pH optima of both enzymes were determined by incubating them for 5 min at 40 °C in 50 mM sodium acetate buffers of various pH values (3.0–7.0). Their temperature optima were determined by incubation at various temperatures (35 °C to 65 °C) in 50 mM sodium acetate buffer of optimal pH. CelA had a pH optimum of 5.0 and temperature optimum of 55 °C, whereas CelB had a pH optimum of 4.0 and temperature optimum of 45 °C. The pH and temperature optima for CelA were similar to the values reported for the *A. aculeatus* FI-CMCase (Murao et al. 1988). The thermal and pH stabilities were determined by incubating the enzymes at various temperatures (35–65 °C) for 10 min and at various pH values (3.0–7.0) for 20 h at 30 °C respectively. Both enzymes were stable in the wide pH range of 3.0–7.0, but CelB was rather unstable compared to CelA at pH 3.0 and 7.0. The CelA was stable up to 55 °C and became inactivate sharply above 60 °C, whereas the CelB was stable up to 50 °C and became inactivate rather sharply above 55 °C. The thermal stability of CelA was superior to that of the *A. aculeatus* FI-CMCase (Murao et al. 1988).

**Table 1** Enzymatic characteristics of the purified endo-1,4-β-glucanases, CelA and CelB

| Characteristics | CelA | CelB |
|---|---|---|
| Molecular mass (kDa) | 31 | 53 |
| pH optimum | 5.0 | 4.0 |
| pH stability | 3.0–7.0 | 3.0–7.0 |
| Temperature optimum(°C) | 55 | 45 |
| Temperature stability (°C) | < 55 | < 50 |

## Discussion

This is the first report of the isolation, sequencing and characterization of the genes bearing a high homology to the families C and H cellulases from *A. oryzae*. The *celA* gene, encoding the family H cellulases comprised 877 bp coding for 239 amino acids. The *celB* gene, encoding the family C cellulases comprised 1248 bp coding for 416 amino acids. The DNA sequences surrounding the ATG translation start sites of *celA* (CACCATGAA) and *celB* (CGATATGAT) closely resemble the Kozak consensus sequence for filamentous fungi (Kozak 1986). The 5′ and 3′ non-coding regions of the *celA* and *celB* genes were screened for various consensus sequences. Although no typical eukaryotic promoter with a TATA box existed, one CCAAT sequence, which is often observed in fungal promoter regions, was found 192 and 453 nucleotides upstream from the translation initiation codon of *celA* and *celB* genes. In common with some other fungal genes, the typical polyadenylation signal, AATAAA, was not present within the 3′ non-coding regions of either gene. However, the sequence involved in transcription termination in *Saccharomyces cerevisiae*, defined by Zaret and Sherman (1982): TAG...TA(T)-GT...TTT, was present at nucleotides 932–966 and 1249–1314 of the *celA* and *celB* genes respectively.

The *celA* and *celB* genes were expressed under the taka-amylase A gene promoter in *A. oryzae*. Among transformants, TA31, a transformant carrying the *celA* gene, and TB1, a transformant carrying the *celB* gene, produced approximately 70 and 500 times more enzyme respectively than the host strain. CelA and CelB proteins were easily purified to homogeneity from the respective transformants. Although their enzymatic properties were found to be essentially the same as those of other fungal glucanases, CelA was rather stable to heat treatment compared to other mesophilic fungal enzymes. The apparent molecular masses of CelA and CelB, determined by SDS-PAGE, were approximately 7 kDa and 10 kDa larger than the respective calculated values (Fig. 3, lanes 2, 3). These discrepancies between apparent and calculated molecular masses of the enzymes may be due to the attachment of a significant amount of glycan to the enzymes, which is in accordance with reports from other extracellular fungal enzymes, since two and eight potential *N*-glycosylation sites are present in

the primary structure of CelA and CelB respectively (Figs. 1B, 2B).

In addition to the *celA* and *celB* genes, two more cellulolytic enzyme-encoding genes, *celC* and *celD*, bearing a homology to cellobiohydrolases, have been isolated from the same strain. We are now trying to improve the utilization of soybean cell wall components in the production of soy sauce by expressing all these genes in the shoyu koji mold.

## References


Azevedo MdeO, Felipe MSS, Astolifi-Filho S, Radford A (1990) Cloning, sequencing and homologies of the *cbh-1* (exoglucanase) gene of *Humicola grisea var. thermoidea*. J Gen Microbiol 136:2569–2576

Ballance DJ, Turner G (1985) Development of a high-frequency transforming vector for *Aspergillus nidulans*. Gene 36:321–331

Dalbøge H, Heldt-Hansen H (1994) A novel method for efficient expression cloning of fungal enzyme genes. Mol Gen Genet 243:253–260

Davis GJ, Dodson GG, Hubbard RE, Tolley SP, Dauter Z, Wilson KS, Hjort C, Mikkelsen J-M, Rasmussen G, Schulein M (1993) Structure and function of endoglucanase V. Nature 365:362–364

Hanahan D (1983) Studies of transformation of *Escherichia coli* with plasmids. J Mol Biol 166:557–580

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

Kitamoto N, Kimura T, Kito Y, Ohmiya K, Tsukagoshi N (1993) Structural features of a polygalacturonase gene cloned from *Aspergillus oryzae* KBN616. FEMS Microb Lett 111:37–42

Kitamoto N, Kimura T, Kito Y, Ohmiya K, Tsukagoshi N (1995) The nitrate reductase gene from a shoyu koji mold, *Aspergillus oryzae* KBN616. Biosci Biotechnol Biochem 59:1795–1797

Ko EP, Akatsuka H, Moriyama H, Shinmyo A, Hata Y, Katsube Y, Urabe I, Okada H (1992) Site-directed mutagenesis at aspartate and glutamate residues of xylanase from *Bacillus pumilus*. Biochem J 288:117–121

Koch A, Weigel CTO, Schulz G (1993) Cloning, sequencing, and heterologous expression of a cellulase-encoding cDNA (*cbh1*) from *Penicillium janthinellum*. Gene 124:57–65

Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44:283–292

Murao S, Sakamoto R, Arai M (1988) Cellulases of *Aspergillus aculeatus*. Methods Enzymol 160:247–299

Ohmiya K, Maeda K, Shimizu S (1987) Purification and properties of endo-(1→4)-β-D-glucanase from *Ruminococcus albus*. Carbohydr Res 166:145–155

Okada H (1989) Tertiary structure of xylanase and estimation of active sites by site directed mutagenesis. GBF Monogr 12:81–86

Ooi T, Shinmyo A, Okada H, Murao S, Kawaguchi T, Arai M (1990) Complete nucleotide sequence of a gene coding for *Aspergillus aculeatus* cellulase (F1-CMCase). Nucleic Acids Res 18:5884

Penttilä M, Lehtovaara P, Nevalainen H, Bhikhabhai R, Knowles J (1986) Homology between cellulase genes of *Trichoderma reesei*: complete nucleotide sequence of the endoglucanase I gene. Gene 45:253–263

Saarilahti HT, Henrissat B, Palva ET (1990) CelS: a novel endoglucanase identified from *Erwinia carotovora subsp. carotovora*. Gene 90:9–14

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

Saloheimo M, Lehtovaara P, Penttilä M, Teeri TT, Ståhlberg J, Johansson G, Pettersson G, Claeyssens M, Tomme P, Knowles JKC (1988) EGIII, a new endoglucanase from *Trichoderma reesei*: the characterization of both gene and enzyme. Gene 63:11–21

Saloheimo A, Henrissat B, Hoffren 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

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Sheppard PO, Grant FJ, Oort PJ, Sprecher CA, Foster DC, Hagen FS, Upshall AU, McKnight GL, O'Hara PJ (1994) The use of conserved cellulase family-specific sequences to clone cellulase homologue cDNAs from *Fusarium oxysporum*. Gene 150:163–167

Shoemaker S, Schweickart V, Ladner M, Gelfand D, Kwok S, Myambo K, Innis M (1983) Molecular cloning of exo-cellobiohydrolase I derived from *Trichoderma reesei* strain L27. Biotechnology 1:691–696

Sims P, James C, Broad P (1988) The identification, molecular cloning and characterization of a gene from *Phanerochaete chrysosporium* that shows strong homology to the exo-cellobiohydrolase I gene from *Trichoderma reesei*. Gene 74:411–422

Törrönen A, Kubicek CP, Henrissat B (1993) Amino acid sequence similarities between low molecular weight endo-1,4-β-xylanases and family H cellulases revealed by clustering analysis. FEBS Lett 321:135–139

Tsukagoshi N, Furukawa M, Nagaba H, Kirita N, Tsuboi A, Udaka S (1989) Isolation of a cDNA encoding *Aspergillus oryzae* taka-amylase A: evidence for multiple related genes. Gene 84:319–327

Zaret KS, Sherman F (1982) DNA sequence required for efficient transcription termination in yeast. Cell 28:563–573