Cloning, Sequencing, and Expression of a Thermostable Cellulase Gene of *Humicola grisea*†

Shou Takashima, Akira Nakamura, Haruhiko Masaki, and Takeshi Uozumi‡

Department of Biotechnology, Faculty of Agriculture, The University of Tokyo, Yayoi I–I–1, Bunkyo-ku, Tokyo 113, Japan

Received June 20, 1996

The egf2 gene encoding a thermostable endoglucanase (EGL2) was cloned from *Humicola grisea*. The DNA sequence of egf2 predicted two putative introns in the coding region. The deduced amino acid sequence of EGL2 was 388 amino acids in length and showed 99.5% identity with the *H. insolens* CMC 3. In addition to TATA box and CAAT motifs, putative CREA binding sites were observed in the 5′ upstream region of the egf2 gene. The egf2 gene was expressed in *Aspergillus oryzae*, and EGL2 was purified. EGL2 produced by *A. oryzae* showed a high activity toward carboxymethyl cellulose. The optimal temperature of EGL2 was 75°C, and EGL2 had more than 80% residual activity after heating up to 75°C for 10 min. This is the first report of enzymatic properties of the EGL2-type thermostable cellulase homologs from *Humicola*.

**Key words:** cellulase; *Humicola grisea*

Endoglucanases (EGL, endo-1,4-β-glucanase, EC 3.2.1.4), exo-cellulohydrolases (CBH, exo-1,4-β-glucanase, EC 3.2.1.91), and β-glucosidases (EC 3.2.1.21) are three major types of cellulolytic enzymes. Among the cellulolytic fungi, the genus *Humicola* has been known to produce several kinds of thermostable cellulases.1,2 From *H. insolens*, cDNA of several cellulase genes, including a gene encoding a cellulase that is more than 95% similar to the *H. grisea* cellulohydrolase I (CBH1),3 have been cloned, sequenced, and expressed.4 But there has been little information for *H. insolens* cellulase genes about the sequences of the promoter region and existence of introns. The enzymatic characterization of gene products also has not been reported.5 The cellulase systems of *H. grisea* and *H. insolens* seem to have many features in common, so it seems efficient to use the sequence of cellulase genes of *H. insolens* for the cloning of cellulase genes of the thermophilic fungus *H. grisea*. In this study, we have analyzed the genomic clone of a cellulase gene that is highly similar to the *H. insolens* cDNA clone, and shown enzymatic properties of the gene product.

**Materials and Methods**

Strains, plasmids, and media. *H. grisea* var. *thermoidea* IFO9854 was used for DNA isolation. The stock culture was stored on agar (1.5%) slants of MY medium (2% malt extract, 0.2% yeast extract). For chromosomal DNA preparation, the spores were inoculated in the liquid MY medium and grown for 2 days at 37°C with shaking, and mycelia were harvested by filtration. *Aspergillus oryzae* M-2-3 (argB+)10 was used as a host for expression of the cloned cellulase genes, and the stock culture was stored on DYP medium.10 Czapek-Dox medium was used for fungal transformation.10 For the expression of cellulase genes from *H. grisea*, *A. oryzae* transformants were cultivated in CD-P medium.8 Escherichia coli JM10910 was used as a host for cloning vector, pUC118.

**Genomic DNA cloning.** Fungal chromosomal DNA was prepared as described by Tonouchi et al.12 Two oligonucleotide primers (5′-ACCAT-GAAGCACAGGTCC-3′, which is identical to the nucleotide sequence 13 to 31 of the *H. insolens* endoglucanase (CMC 3) gene,13 and 5′-TATACGACACATGGCACTGA-3′, complementary to the nucleotide sequence 1171 to 1190) were synthesized, and an about 1.4-kb fragment was amplified by polymerase chain reaction (PCR) using the two primers with *H. grisea* var. *thermoidea* IFO9854 chromosomal DNA as the template. The amplified DNA was radiolabeled with [α-32P]CTP, using a random primed DNA labeling kit (Boehringer Mannheim), and used as a probe for Southern hybridization15 of *H. grisea* chromosomal DNA digested with *BamH*I, using a Hybrid-N+ nylon membrane (Amersham). The membrane was washed for 20 min, three times at room temperature in 2× SSC, 0.1% SDS and for 30 min, twice at 65°C in 0.1× SSC, 0.1% SDS. The hybridizing fragment was cloned into pUC118 using the colony hybridization technique.16

DNA sequencing. DNA sequencing was done by the dideoxy chain termination method,17 using a single strand-nested deletion kit (TaKaRa Co., Ltd.) and a RoBest sequencing kit (TaKaRa Co., Ltd.). The nucleotide sequence was obtained for both strands.

**Fungal transformation.** Transformation of *A. oryzae* was done according to the method of Gomi et al.18

**Enzyme assay.** The enzyme activities toward carboxymethyl cellulose, Avicol, and xylan were measured by incubating 0.05 ml of 1% substrate solution in 50 mM sodium acetate buffer (pH 5.0) with 0.01 ml of the enzyme solution at 50°C for an appropriate time. Reducing sugars produced by these reactions were measured by the dinitrosalicylic acid method.19 For cellobiohydrosacharides, 0.1% substrate solutions were used and the enzyme activities were measured as described above. For p-nitrophenol derivatives, the enzyme activities were measured as described previously.20

For the optimal temperature, the enzyme activities toward carboxymethyl cellulose were measured as described above at various temperatures for 5 min. For the thermal stability, the enzyme solutions were treated at various temperatures for 10 min at pH 5.0 without substrate, then the remaining activities toward carboxymethyl cellulose were measured as described above at 50°C for 5 min. For the optimal pH, the enzyme activities toward carboxymethyl cellulose were measured as described above at various pHs and 50°C for 5 min. For the pH stability, the enzyme solutions were treated at various pHs at 4°C for 20 h, then the enzyme solutions were adjusted at pH 5.0 and the remaining activities toward carboxymethyl cellulose were measured as described above at 50°C for 5 min.

One unit of the enzyme was defined as the activity producing 1 μmol of reducing sugars in glucose or xylose equivalents or p-nitrophenol per min under these assay conditions.

**Protein purification.** To purify EGL2 from *A. oryzae* transformants, they were cultivated for 4 days in CD-P medium and the culture supernatants

† DDBJ/EMBL/GenBank accession number: D84470
‡ Corresponding author.

were obtained by filtration. The filtrate was buffered by adding 1/10 volume of 100 mM Tris-HCl buffer, pH 7.5, and gently mixed for 1 h at room temperature with SuperQ Toyopearl 650 m (Tosoh Co., Ltd.) previously equilibrated with 10 mM Tris-HCl buffer, pH 7.5. After centrifugation, the supernatant was ultrafiltered with Centricon 30 (Millipore) and the purified EGL2 was obtained. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was done as described by Laemmli, and stained with Coomassie Brilliant Blue R-250. Periodic acid Schiff (PAS) staining was done using the periodic acid Schiff staining system (Sigma) according to the manufacturer’s instructions, except final hematoxylin solution treatment was replaced by 0.05% HCl-0.5% NaHSO$_4$ solution treatment. Protein content was measured by a dye-binding assay kit (Protein assay kit, Bio-Rad Co., Ltd.) using γ-globulin as the standard.

Results and Discussion
Cloning of the egl2 gene
From H. insolens, a cellulase gene encoding a cellulase which is more than 95% similar to the H. grisea CBH1 has been cloned. It seems that other cellulase genes of

\[ \text{H. insolens} \] also show high sequence similarity with some of cellulase genes of \( \text{H. grisea} \). So we synthesized two PCR primers which are specific to the \( \text{H. insolens} \) CMC 3 gene and PCR was done using these primers with \( \text{H. grisea} \) chromosomal DNA instead of \( \text{H. insolens} \) chromosomal DNA as the template. As we expected, an about 1.4-kb fragment was amplified and we used it as a probe for the cloning of a cellulase gene of \( \text{H. grisea} \). Southern hybridization of \( \text{H. grisea} \) chromosomal DNA digested with \( \text{BamHI} \) was done, and a 4.2-kb fragment was found to uniquely hybridize. The hybridizing fragment was cloned into pUC118 using the colony hybridization technique. Restriction and partial sequence analyses of this clone showed that the cloned \( \text{BamHI} \) fragment contained a gene similar in sequence to the \( \text{H. insolens} \) CMC 3 gene, and this fragment was sequenced on both strands by the dideoxy sequencing method.

\[ \text{H. grisea} \]

**Fig. 1.** Nucleotide Sequence of the \( \text{H. grisea} \) egl2 Gene and the Deduced Amino Acid Sequence.

Numbers of the nucleotides begin with the initiation codon. The putative intron sequences are shown in lowercase letters. The TATA box is underlined. The CAAT sequences are double-underlined. The CREB binding consensus sequences are shown by the broken underlines. The cellulose-binding domain is shown by the wavy underline. The nucleotide sequence of the egl2 gene has been deposited in the DDBJ, GenBank, and EMBL data bases under accession No. D64470.
The nucleotide and deduced amino acid sequences of the endoglucanase gene (designated as eg12) are shown in Fig. 1. The sequence contained a coding region of 1326 nucleotides, and comparison with the cDNA sequence of the H. insolens CMC 3 gene suggested that there are two putative introns of 78-bp and 84-bp in this coding region. The comparison also showed that there are 26 nucleotide differences with two amino acid substitutions in the eg12 coding region: GGT (Gly 8 of H. insolens CMC 3) → AGT (Ser), ACC (Thr 340 of H. insolens CMC 3) → AAC (Asn). Alignment with the nucleotide sequence of the H. grisea eg12 gene with the H. insolens CMC 3 gene shows 97.7% identity and alignment of the amino acid sequence of the H. grisea EGL2 with that of the H. insolens CMC 3 shows 99.5% identity. Compared with other reported cellulases by homology searching, the H. grisea EGL2 had a significant degree of similarity with EGL1 of a phytopathogenic fungus, Macrophomina phaseolina (55.6% identity), EGL of a phytopathogenic bacterium, Pseudomonas solanacearum (43.8% identity), and EGL3 of a cellulolytic fungus, Trichoderma reesei (32.4% identity) (Fig. 2). The translation product of the coding region should have 388 amino acids in length with a molecular mass of 42605 Da. The conserved residues of the fungal cellulose-binding domain is observed in the N-terminal region followed by the hinge region which is

Fig. 2. Sequence Comparison of the H. grisea EGL2 with Other Cellulases.

HGEGL2, H. grisea EGL2; HICMC3, H. insolens CMC 3; MPEGGL1, M. phaseolina EGL1; PSEGL1, P. solanacearum EGL1; TREG3, T. reesei EGL3. Asterisks denote identical residues to those of the H. grisea EGL2.
rich in Thr, Pro, and Ala residues.

The frequencies of T, C, A, and G at the third position of codons of the egl2 gene are 55T, 177C, 25A, and 131G, respectively. There is a strong bias for the use of cytidine and guanine at the third position of codons. The GC content of the egl2 gene is 59.5%.

In addition to the coding region, we have sequenced the promoter region of the egl2 gene. A putative TATA box is found at nt -126, and putative CAAT motifs are found at positions 581-bp or further upstream from the TATA box. Within the determined upstream sequence, four CREA protein (a carbon catabolite repressor protein of Aspergillus nidulans) binding consensus sequences (5'-SYGGRG-3') 6 are observed, suggesting that some of these sequences are related to carbon catabolite repression, and the expression of the egl2 gene may be regulated by the CREA protein of H. grisea.

Expression of the egl2 gene

Although the H. insolens CMC 3 gene has been cloned, sequenced, and expressed, 4 its enzymatic properties have not been reported. To characterize the enzymatic properties of the H. grisea EGL2, we tried to overexpress the egl2 gene in H. grisea by using the transformation system of H. grisea. 13 But we failed in transforming H. grisea for reasons we do not understand. H. insolens cellulase genes have been expressed in A. oryzae, 3 so we also used the expression system of A. oryzae. 5,6,10 The expression vector of the egl2 gene, pAMYB-EGL2 was constructed by inserting the amyB (the Taka-amylase gene of Aspergillus oryzae) promoter from pMAR518 in the vector pUC118 (designated as pAMYB118), then inserting the 2.9-kb PstI (nt -38)-BamHI fragment of the egl2 gene in pAMYB118. Thus the egl2 gene would be expressed under the amyB promoter regulation. The amyB gene is highly expressed when A. oryzae is grown on maltose as a main carbon source, thus the egl2 gene also would be highly expressed when transformants are grown on maltose. In these growth conditions, production of endogenous cellulases of A. oryzae is very low, so the activity of EGL2 produced by transformants would be detected easily.

The expression vector (20 μg) was introduced into A. oryzae M-2-3, an arginine-auxotroph, by co-transformation with 20 μg of an argB-containing plasmid, pSal23. 3 The arg7 transformants were isolated and grown in 10 ml CD-P medium 3 (containing maltose as a carbon source) for 4 days at 30°C and assayed for enzyme activity. The transformants showed a 25- to 40-fold higher endoglucanase activity than the untransformed strain. Southern hybridization of genomic DNA obtained from the transformant showing the highest activity (designated as AOEGL2-1) showed that several copies of the egl2 gene had been integrated into the genome of the transformant (Fig. 3). The culture supernatants of transformants were also analyzed by SDS-PAGE (Fig. 4a). The major protein produced by the untransformed strain and the control strain transformed with pAMYB118 and pSal23 was Taka-amylose (50-kDa) 3 only. But transformants showing endoglucanase activity produced a 55-kDa protein as well as Taka-amyrase. From the culture supernatants of transformants showing endoglucanase activity, we purified a 55-kDa protein using SuperQ Toyopearl 650M (Toso-

![Fig. 3. Southern Hybridization Analysis of the A. oryzae M-2-3 Transformants.](image)

and Centriprep 30 (Millipore) (Fig. 4b). The H. insolens CMC 3 gene also has been expressed in A. oryzae and CMC 3 has been detected by SDS–PAGE as a 55-kDa protein, which was identified as CMC 3 by western blotting. 4 PAS staining showed the existence of glycosylation in the purified 55-kDa protein (Fig. 4c). The difference between the molecular mass from the deduced amino acid sequence of EGL2 (42.6 kDa) and that of purified protein estimated by SDS–PAGE (55 kDa) seemed to be a result of the glycosylation of EGL2.

Enzymatic properties of EGL2

Using carboxymethyl cellulose as a substrate, the optimal temperature, the thermal stability, the optimal pH, and the pH stability of EGL2 produced by A. oryzae were measured (Fig. 5). The optimal temperature for the reaction of EGL2 was 75°C, and EGL2 showed more than 80% residual activity against heating up to 75°C for 10 min. The optimal pH of EGL2 was 5.0, and EGL2 was stable at least in a range of pH 4.0–11.0 at 4°C for 20 h. EGL2 seemed a thermostable endoglucanase when it was compared with other fungal endoglucanases. 19,20 As we have not purified native EGL2 from H. grisea, the enzymatic properties of native EGL2 are unknown (endoglucanase 2 purified from H. grisea previously 20 does not correspond to EGL2). However, the thermal stability of cloned EGL2 seems to reflect that of native EGL2, as the genus Humicola has been known to produce several kind of thermostable cellulases. 1,2

EGL2 produced by A. oryzae was tested for substrate specificities (Table). EGL2 showed high activity toward carboxymethyl cellulose and significant activity toward Avicel and xylan. For cellobiohexaosaccharides, EGL2 pre-
Fig. 4. SDS-PAGE Analysis of EGL2 Produced by A. oryzae.
(a) Culture supernatants of the transformants and A. oryzae. Lane M, molecular mass standard; lane 1, culture supernatant of A. oryzae M-2-3; lane 2, culture supernatant of the control transformant, which was co-transformed with pAMYB118 and pSal23; lane 3, culture supernatant of the transformant AOEGL2-1, which was co-transformed with pAMYB-EGL2 and pSal23.
(b) Purified EGL2; lane M, molecular mass standard.
(c) PAS staining of purified EGL2.

Fig. 5. Effects of Temperature and pH on the Endoglucanase Activity.
(a) Effects of temperature. Open circles, optimal temperature; closed circles, thermal stability.
(b) Effects of pH. Open symbols, optimal pH; closed symbols, pH stability. △, ◇, sodium citrate buffer; ◇, ●, sodium acetate buffer; ◇, ●, sodium phosphate buffer; △, A, Tris–HCl buffer; □, ■, glycine–NaOH buffer; ×, NaOH–phosphate buffer.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>U/mg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carboxymethyl cellulose</td>
<td>31.9</td>
</tr>
<tr>
<td>Avicel</td>
<td>0.0812</td>
</tr>
<tr>
<td>Xylan</td>
<td>1.01</td>
</tr>
<tr>
<td>Cellobiose</td>
<td>0.117</td>
</tr>
<tr>
<td>Cellobiose</td>
<td>0.119</td>
</tr>
<tr>
<td>Cellotriose</td>
<td>0.196</td>
</tr>
<tr>
<td>Cellotetraose</td>
<td>0.227</td>
</tr>
<tr>
<td>Cellopentose</td>
<td>0.261</td>
</tr>
<tr>
<td>Celiohexaose</td>
<td>0.261</td>
</tr>
<tr>
<td>p-Nitrophospholyl-β-D-glucoside</td>
<td>nd</td>
</tr>
<tr>
<td>p-Nitrophospholyl-β-D-cellobioside</td>
<td>nd</td>
</tr>
</tbody>
</table>

nd, not detected.

ferred to hydrolyze longer species of cellooligosaccharides. The activity toward p-nitrophospholyl-β-D-glucoside and p-nitrophospholyl-β-D-cellobioside could not be detected.

In this study we have shown that the H. insolens CMC 3 equivalent exists in H. grisea and the cloned H. grisea EGL2 produced by A. oryzae is a thermostable endoglucanase with high activity toward carboxymethyl cellulose. Generally, cellulase complexes produced by celluloytic fungi show strong activity through the synergism among the cellulase components. But the mechanism of the synergism among the cellulase components is complicated. Analysis of each of the cellulase components would be useful for understanding of the mechanism of the synergism among them.
Acknowledgments. We wish to thank Dr. K. Kitamoto for generously providing the expression system of *A. oryzae*. This work was done using the facilities of the Biotechnology Research Center, The University of Tokyo.

References