Amino Acid Sequence and Stereoselective Hydrolytic Reaction of an Endo-1,4-β-glucanase from a Bacillus Strain†

Jun Hitomi, Yuji Hatada, Shunro Kawaminami, Shuji Kawai, and Susumu Ito††

Tochigi Research Laboratories of Kao Corporation, 2606 Akabane, Ichikai, Haga, Tochigi 321–34, Japan

Received May 12, 1997

Bacillus sp. KSM-522 produces three different extracellular endo-1,4-β-glucanases [EGs; Okoshi et al., Agric. Biol. Chem., 54, 83-89 (1990)]. Here, we report the molecular cloning and sequencing of the gene for the fourth EG (EG-IV) of the organism and the mechanism of its hydrolytic reaction. The structural gene contained an open reading frame of 1911 bp, corresponding to 636 amino acids, the amino acid sequence of which was very close to that of an EG of Clostridium cellulovorans, belonging to the cellulase family E2. The molecular mass of the extracellular mature enzyme (Ser126 through Lys632) was calculated to be 69,076 Da, a value close to the 69.2 kDa measured for the recombinant enzyme in Bacillus subtilis. The optimum pH and temperature for activity of the recombinant enzyme were pH 8.0 and 50°C, respectively. By 1H-NMR spectroscopy, we demonstrated that the hydrolysis of p-nitrophenyl β-D-cellobioside by EG-IV proceeded with inversion of the anomeric configuration.

Key words: Bacillus; cellulase; endoglucanase; family E; inversion

We have found and characterized many endo-1,4-β-glucanases (EG, EC 3.2.1.4) in strains of alkaliphilic and neutrophilic Bacillus, and some alkaline and alkali-tolerant EGs of them can be used as effective additives in laundry detergents under alkaline conditions.1,2) We also cloned and sequenced some of the genes for these EGs5-8) and expressed them in Bacillus subtilis in quantity, with a special focus on differences in pH optima for the activity of these enzymes.6-9)

Recently, we cloned the gene for an EG, designated EG-IV, from cells of neutrophilic Bacillus sp. KSM-522, which had originally been isolated as a multiple cellulase producer.10,11) This study had first been aimed at understanding the structure and mechanism of action of the alkaline EG, E-II, or E-III, of this organism, but one of the genes that we cloned was found to encode an unidentified EG-IV, belonging to cellulase family E (family 9 of glycosyl hydrolases).12) Many EGs of strains of the genera Clostridium, Pseudomonas, Butyrivibrio, Dictyostelium, Perseus, and Thermomonospora have been reported to be the members of cellulase family E.12-14) However, a cellulase of this class has not yet been reported for the genus Bacillus. This is the first report that provides proof of the occurrence of a family E cellulase in Bacillus by sequencing the gene and by analyzing the NMR spectra of reaction products.

Materials and Methods

Bacterial strains and propagation. The source of the gene examined in this study was Bacillus sp. KSM-522, which had previously been isolated from a soil sample in our laboratory. The organism was grown in medium III that contained 1% (w/v) maltose as the source of carbon.15) Escherichia coli HB101 (F− hsdS20 recA13 ara-14 proA2 lacY1 galK2 rpsL20 xyl-5 mtl-1 supE44 leuB6 thi-1) and pBR322 were used for cloning experiments. E. coli was grown, with shaking, in LB broth (Difco) at 37°C.

Isolation of DNA and transformation. Genomic DNA from Bacillus sp. KSM-522 was prepared as described by Saito and Miura,15) and plasmid DNA was isolated by the alkaline extraction procedure of Birnboim and Doly.16) Competent E. coli HB101 (TakaRa) and Bacillus subtilis ISW1214 (levA8 metB5 hisM1) cells were transformed with plasmids by the methods of Hanahan17) and Chang and Cohen,18) respectively. Transformed B. subtilis cells were grown at 30°C for 60 h, with shaking, in LB broth with tetracycline (15 µg/ml).

Construction of recombinant plasmids. The genomic DNA from Bacillus sp. KSM-522 and pBR322 were both digested with BamHI and ligated with T4 ligase (Boehringer Mannheim). The ligation mixture was used for transformation of competent E. coli cells. Transformant cells, which appeared on LB plates with 50 µg/ml ampicillin (Sigma), were checked for EG productivity by the Congo red-CMC procedure, as described elsewhere.19) An EG-positive clone (11.9 kb) was found to encode an active enzyme (carboxymethylcellulase, CMCase) on a 7.5-kb BamHI-BamHI insert. Subcloning showed that a 3.1-kb SphI-SphI region of the insert contained the region essential for the EG activity. The SphI-SphI fragment was inserted into the SnaI site of plasmid pHY300PLK10) and the resultant recombinant plasmid was designated pBEG38 (8.0 kb). pBEG38 was then expressed both in E. coli and in B. subtilis and EG-positive transformants were detected by the Congo red-CMC procedure. The sequence of the gene for EG in pBEG38 was measured, and it is shown in Fig. 1.

Sequencing of DNA. Sequencing was done by the dye-exchange chain-termination method of Smith et al.,19) using fluorescent terminators and an automated DNA sequencer (model 377A, Perkin Elmer). Both strands of the DNA were sequenced and computer analysis was done using a GENETYX program (SDC Software Development).

Assay of enzyme activity. EG activity was measured as CMCase activity at 50°C and at pH 8.0 in 0.1 M Tris-Cl buffer, as described elsewhere.16) The reducing sugar formed was measured by the dinitrosalicylic acid procedure.20) One unit of enzymatic activity was defined as the amount of protein that produced 1 µmol of reducing sugar as glucose per min under the conditions of the assay. Protein was measured by the method of Bradford,21) with a protein assay kit (Bio-Rad) and bovine serum albumin as the standard.

Purification of the expressed enzyme. For fermentation of EG,

† The nucleotide sequence reported in this paper has been submitted to the GenBank/EMBL/DDBJ Data Bank with the accession number AJ004098.

†† To whom all correspondence should be addressed.

Abbreviations: EG, endo-1,4-β-glucanase; PAGE, polyacrylamide gel electrophoresis; NMR, nuclear magnetic resonance; ORF, open reading frame; pNPG1 through pNPG5, p-nitrophenyl derivatives of β-D-glucopyranoside through β-D-cellobioside; CMC, carboxymethylcellulose.
transformed *B. subtilis* was grown, with shaking, at 30 °C in a medium (50 ml) composed of (w/v) 12% corn steep liquor, 0.1% yeast extract (Difco), meat extract (Wako Pure Chemical), 0.1% KH₂PO₄, 0.02% MgSO₄.7H₂O, 0.05% CaCl₂, 3% maltose, and 15 µg/ml tetracycline (pH 7.4). After 2 days of incubation of cultures, cells were removed by centrifugation (12,000 × g, 15 min) at 4 °C. To the supernatant was added solid ammonium sulfate to 90% saturation. Precipitates formed were collected by centrifugation and dissolved in a small amount of 10 mM Tris–HCl buffer (pH 7.5), which was then dialyzed overnight against a large volume of the same buffer. The retentate was put on a column of DEAE-Toyopearl 650S (5 cm × 13 cm; Tosoh) that had been equilibrated with the 10 mM Tris–HCl buffer (pH 7.5) with 2 mM CaCl₂. The column was washed with 500 ml of the equilibration buffer, and proteins were eluted with 0.3 M NaCl in the equilibration buffer. The active fractions were combined and concentrated by ultrafiltration on a PM-10 membrane (10 kDa cutoff; Amicon). The concentrated sample was adsorbed to a column of hydroxyapatite (3 cm × 20 cm; Pharmacia) equilibrated with 10 mM potassium phosphate buffer (pH 7.5), and proteins were eluted with a 1.4-liter gradient of 10 mM to 0.2 M phosphate buffer (pH 7.5). EG activity was eluted at around 60 mM phosphate. The active fractions were pooled and concentrated to a small volume by ultrafiltration. The concentrate was put directly on a column (3 cm × 20 cm) of DEAE-Toyopearl 650S equilibrated with 10 mM Tris–HCl buffer (pH 7.5) plus 2 mM CaCl₂. Elution was done with a 1.4-liter gradient of 0 to 0.3 M NaCl in the equilibration buffer. EG activity was eluted at 0.15-0.18 M NaCl. The active fractions were combined and concentrated to approximately 1 ml by ultrafiltration, and the concentrate was then put on a column (1.5 cm × 96 cm) of Sephacryl S-200 (Pharmacia) equilibrated with 10 mM Tris–HCl buffer (pH 7.5) plus 2 mM CaCl₂. Elution was done with the same buffer at a flow rate of 7 ml/h and fractions of 4 ml were collected. The active fractions were combined and concentrated to a small volume. The resultant concentrate was used as the final preparation of recombinant enzyme. The purified enzyme was homogeneous, as judged by nondenaturing polyacrylamide gel electrophoresis (PAGE). The purified enzyme was used for studies of the enzymatic properties of EG, such as optimum pH, optimum temperature, and substrate specificity, and also for NMR spectroscopic analysis.

**Estimation of molecular mass.** Molecular mass was estimated by SDS-PAGE [10% (w/v) acrylamide, 90 mm × 90 mm, 2.0 mm thickness], according to Laemmli, with low-range molecular mass standards (Bio-Rad), which included rabbit muscle phosphorylase b (97.4 kDa), bovine serum albumin (66.2 kDa), hen egg white ovalbumin (45 kDa), bovine carbonic anhydrase (31 kDa), soybean trypsin inhibitor (21.5 kDa), and hen egg white lysozyme (14.4 kDa). Proteins were stained with Coomassie Brilliant Blue dye.

**Chromatographic analysis of the products of hydrolysis.** Each cello-oligosaccharide or its p-nitrophenyl derivative (pNPG1-pNPG5) was dissolved in 10 ml of 50 mM Tri–HCl buffer (pH 7.0), at a final concentration of 0.05% (w/v). A sample of the purified preparation of enzyme was added to the reaction mixture, which was then incubated at 25°C for 24 h. After completion of the reaction, 10-µl samples were withdrawn and chromatographed on thin-layer plates of silica gel 60 (10 cm × 20 cm; Merck) in a solvent system composed of ethylacetate, acetic acid, and water (3 : 3 : 1, v/v). Chromatograms were developed by spraying the aconit–sulfuric acid reagent.

**Sequencing of N-terminal region.** The sample was blotted onto a polyvinylidene difluoride membrane in a Prosortb cartridge (Perkin Elmer). The N-terminal amino acid sequence of the purified enzyme was measured directly in a protein sequencer (model 476A; Perkin Elmer) that was connected to an on-line PTH amino acid analyzer.

**NMR spectroscopy.** Spectra were recorded on a JNM A-500 spectrometer (JEOL) operated at 30 °C and at 500 MHz for protons in deuterated sodium phosphate buffer (10 mM ²H₂O 7.4). The isotopic enrichment was 99% for ²H₂O (Merck). Spectral width, data point, and the number of accumulations were 6500 Hz, 16 K, and 256, respectively. The water resonance was suppressed by selective irradiation. Chemical shifts were measured relative to the calibrated resonance of internal sodium 3-(trimethylsilyl)-1-propane sulfonate (Merck). The substrate for this experiment was p-nitrophenyl β-D-cellobioside (Seikagaku Corp.) at a final concentration of 1.0 mg/ml, and the reaction conditions were the same as those described elsewhere.

---

**Fig. 1.** Thin-layer Chromatograms of the Products of Hydrolysis of Cello-oligosaccharides (A) and Their p-Nitrophenyl Derivatives (B) by the Recombinant Enzyme. The experimental procedures are described in Materials and Methods. The minor products generated from each substrate are indicated by dotted circles.
Results and Discussion

Enzymatic properties of the recombinant EG

The pBEG38-encoded EG was expressed in B. subtilis cells and was purified to homogeneity, as judged by non-denaturing PAGE (see Materials and Methods). The specific activity for CMC of the recombinant enzyme was approximately 50 units/mg protein. The N-terminal amino acid sequence was Ser-Asp-Tyr-Asn-Tyr-Val-Glu-Val-Leu-Gln. The molecular mass of the enzyme was 69.2 kDa, as judged by SDS-PAGE, a value not identical with those of E-I (78 kDa), E-II (61 kDa), or E-III (61 kDa) secreted by Bacillus sp. KSM-522.11 The optimum temperature for activity of the recombinant EG was around 50°C to 60°C. However, the pH profile of the recombinant enzyme was simply bell-shaped with an optimum pH at 8.0, while the maximum activities of both E-I and E-III were detected over a broad pH range, from 7 to 10.

The most striking difference between the recombinant EG and the two known EGs, E-II and E-III, was observed in terms of patterns of action on cello-oligosaccharides. The recombinant enzyme acted on cellotetraose (G4), cello-pentaose (G5), and cellohexaose (G6), all yielding cellobiose (G2) and cellotriose (G3) as the major end products, as confirmed by thin-layer chromatography (Fig. IA). It hydrolyzed G3 very slightly to yield glucose (G1) and G2. By contrast, E-II and E-III are both known to hydrolyze G5 and G6 to generate G1, G2, and G4 (no G3), but they...
do not hydrolyze G3 and G4.11) Thus, we concluded that the recombinant EG was different from E-II and E-III (and also E-I) and was the fourth EG designated EG-IV hereafter, found in Bacillus sp. KSM-522. EG-IV efficiently hydrolyzed CMC, pNPG3, pNPG4, and pNPG5 were cleaved by the enzyme to generate G2 and pNPG1, G3 and pNPG1 (with minor G1 and G2), and G3 and pNPG1 (with minor G1, G2, and pNPG2), respectively, but pNPG1 and pNPG2 were not hydrolyzable at all, as judged by thin-layer chromatography (Fig. 1B). The concomitant formation of G2 and G3 from G4 or G6 suggests that EG-IV has a transglucosidase activity.

The entire nucleotide sequence of the gene for EG-IV

The SphI-SspI fragment in pBEG38 hybridized with the BamHI digest of the genomic DNA from Bacillus sp. KSM-522. The nucleotide sequence of the gene for EG-IV cloned in the recombinant plasmid, extending from the 5’-terminal SmaI site (nucleotide 1) to the 3’-terminus at nucleotide 2854, was analyzed and is shown in Fig. 2. Starting from an ATG initiation codon at nucleotide 467, there is a long open reading frame (ORF) of 1116 bp that ends in a TAA stop codon at nucleotide 2377. Upstream from this ORF, we could find neither a putative ribosome-binding site (Shine-Dalgarno sequence) nor a consensus-like sequence of sigma A-type vegetative promoters of B. subtilis.27,28) A possible inverted-repeat sequence was found downstream of the stop codon of the ORF (nucleotides 2548–2564). Taking in account the ΔG for a stem-loop structure (−12.5 kJ/mol), however, this sequence could not act as a terminator of transcription.
Amino acid sequence analysis

The ORF in the nucleotide sequence encoded 636 amino acid residues, as shown under the nucleotide sequence in Fig. 2. The first N-terminal 25 amino acid sequence may resemble the signal peptides of Bacillus, but a typical recognition site of a signal peptidase was not found immediately downstream of this sequence. However, a deduced sequence that was identical to the N-terminal 10 amino acid residues of the recombinant EG-IV was found at amino acids 26–35. The predicted molecular mass of the extracellular mature enzyme (from Ser26 to Lys636) would be 69,076 Da, a value close to the 69,2kDa determined by SDS-PAGE of the recombinant EG-IV.

The deduced amino acid sequence of EG-IV was found to be very similar to that of Clostridium cellulovorans (EngC) (the GenBank accession No. M95180; Shoseyov et al.30), belonging to cellulase family E2. Especially, the amino acid residues from Ser26 to Phe529 are completely identical in the two sequences. The amino acid sequence of the C-terminal end of EngC is Cys-Glu-Gly-Ile-His, which corresponds to the aligned internal sequence, 532Ala-Lys-Gly-Phe-Thr534, of EG-IV (see Fig. 2). The nucleotide sequence of EngC has an insertion of one nucleotide, T, corresponding to the position between ATA (Ile524) and TTT (Phe529) of EG-IV. As a result of the single insertion at this position, the nucleotide sequence around this region in EngC is read as Cys-Glu-Gly-Ile-His, instead of Ala-Lys-Gly-Phe-Thr. and the translation is ended immediately downstream of this nucleotide sequence by generation of a stop codon, TAA. However, the molecular mass of EngC expressed in E. coli, was shown to be approximately 7kDa30, a value very close to the mature EG-IV. If the single insertion of the T was not correct for the nucleotide sequence of EngC gene, its deduced amino acid sequence is almost identical to that of the mature EG-IV. According to the taxonomic study of Sleat et al.,31 the G + C content of the anaerobic C. cellulovorans is characteristically very low (26–27 mol%). Therefore, it is very interesting that the G+C content is 42.6 mol% for both EngC and EG-IV genes in the ORFs, a value which is almost identical to that of the genomic DNA (approximately 42 mol%) of Bacillus sp. KSM-522, which resembles B. pumilus.30

When suitably aligned, the deduced amino acid sequence of EG-IV (Met1-Lys836) had 53.3, 51.9, 51.7, 50.0, and 48.2% identity with those of EGI from C. thermocellum, EGF from C. thermocellum, CelZ from Cellulomonas fimi, CelB from C. fimi, and CelCGG from C. cellobiyticum, respectively, all of which are members of cellulase family E.

Identification of stereoselectivity of the reaction by 1H-NMR spectroscopy

It has been demonstrated by Gebler et al.37 that members of the same cellulase family have the same stereoselectivity due to similar active site topology. The stereochemical course of hydrolysis by cellulases in family E is suggested to proceed by an inversion mechanism. We therefore examined the stereoselective hydrolysis reaction catalyzed by the recombinant EG-IV, with pNPG3 as substrate.

Figure 3A shows a 1H-NMR spectrum of the anomeric proton region of a pNPG3 in deuterated phosphate buffer (10 mM, p7H 7.4). A doublet at 5.31 ppm (J=7.9 Hz) is due to the β-anomeric proton of pNPG3 and two doublets at 4.57 ppm (J=7.9 Hz) and 4.52 ppm (J=7.9 Hz) are assigned to the β-anomeric protons on the second and the third sugar of the substrate, respectively.25

In a spectrum recorded 4h after addition of EG-IV (Fig. 3B), an α-anomeric proton signal of α-cellbiose increased for the first time at 5.22 ppm (doublet, J=3.8 Hz) and the anomeric proton at 4.57 ppm decreased concomitantly. In addition, the resonance of the anomeric proton of pNPG1 was observed at 5.29 ppm (doublet). A doublet peak was also observed at 4.51 ppm and was assigned to the anomeric proton of the second sugar moiety of α-cellbiose. The generation of α-cellbiose and pNPG1 indicates clearly that the bond between the first and the second sugar of pNPG3 was cleaved by the enzyme with inversion of the anomeric configuration.

When the 1H-NMR spectrum was recorded after 9h of enzymatic reaction, a new doublet peak (J=7.9 Hz) was detected at 4.66 ppm (Fig. 3C). This signal was due to the anomeric proton of β-cellbiose that was generated by mutarotation. Figure 3D shows a 1H-NMR spectrum which was recorded after 23h of enzymatic reaction, where a doublet peak at 4.57 ppm completely disappeared, indicating completion of the reaction.
In this study, we showed the existence of a family E cellulase in *Bacillus* for the first time and demonstrated unequivocally that the enzyme catalyzes hydrolysis with inversion of the anomeric configuration. We have not yet detected the native EG-IV in cultures of *Bacillus* sp. KSM-522, and we have no idea at present how the cloned gene for the enzyme is actually expressed in *Bacillus* sp. KSM-522 cells and whether the expressed enzyme is really secreted extracellularly. The genes for E-II and E-III in this organism are also being cloned with a special focus on the classification of their cellulase class and the stereoselective course of the hydrolytic reactions.

References