Molecular cloning and nucleotide sequence of the gene encoding an endo-1,4-β-glucanase from *Bacillus* sp. KSM-330

KATSUYA OZAKI,* NOBUYUKI SUMITOMO and SUSUMU ITO

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

(Received 14 May 1991; revised 11 July 1991; accepted 22 July 1991)

The gene encoding an acid endo-1,4-β-glucanase from *Bacillus* sp. KSM-330 was cloned into the HindIII site of pBR322 and expressed in *Escherichia coli* HB101. The recombinant plasmid contained a 3.1 kb HindIII insert, 1.8 kb of which was sufficient for the expression of endoglucanase activity in *E. coli* HB101. Nucleotide sequencing of this region (1816 bp) revealed an open reading frame of 1389 bp. The protein deduced from this sequence was composed of 463 amino acids with an M₉ of 51,882. The deduced amino acid sequence from amino acids 56 through 75 coincided with the amino-terminal sequence of the endoglucanase, Endo-K, purified from culture of *Bacillus* sp. KSM-330. The deduced amino acid sequence of Endo-K had 30% homology with that of the celA enzyme from *Clostridium thermocellum* NCIB 10682 and 25% homology with that of the enzyme from *Cellulomonas uda* CB4. However, the Endo-K protein exhibited no homology with respect to either the nucleotide or the amino acid sequences of other endoglucanases from *Bacillus* that had been previously characterized. These results indicate that the gene for Endo-K in *Bacillus* sp. KSM-330 has evolved from an ancestral gene distinct from that of other *Bacillus* endoglucanases.

Introduction

Cellulolytic enzymes have been reported in several strains of *Bacillus* (Tewari & Chahal, 1977; Dhillon et al., 1985; Fukumori et al., 1986a, b; MacKay et al., 1986; Nakamura et al., 1987; Robson & Chambliss, 1987; Au & Chan, 1987; Baird et al., 1990). However, cellulases from *Bacillus* have received very limited attention, probably because most of these enzymes hydrolyse carboxymethylcellulose (CMC) but barely hydrolyse crystalline forms of cellulose and because they have had no commercial applications until recently.

We have isolated some strains of *Bacillus* that produce alkaline endo-1,4-β-glucanase (EC 3.2.1.4), the properties of which fulfil the essential requirements for enzymes to be used in laundry detergents (Okoshi et al., 1990; Shikata et al., 1990; Yoshimatsu et al., 1990). A mutant strain of one of these isolates, *Bacillus* sp. KSM-635, is currently exploited for the industrial production of an alkaline endoglucanase for use in such detergents (Yoshimatsu et al., 1990). We have cloned and sequenced the gene for the endoglucanase from *Bacillus* sp. KSM-635 (Ozaki et al., 1990). The deduced amino acid sequence of the alkaline endoglucanase protein exhibited significant homology to the amino acid sequences of other alkaline (Fukumori et al., 1986a, b) and neutral (MacKay et al., 1986; Nakamura et al., 1987; Robson & Chambliss, 1987; Zappe et al., 1988) endoglucanases produced by strains of *Bacillus* and by *Clostridium acetobutylicum*, with some amino acid residues conserved in either the alkaline enzymes or the neutral enzymes, or in both types of enzyme.

In order to clarify the mechanism of action of endo-1,4-β-glucanases, with a special focus on differences in pH optima, we have purified and characterized an acid endo-1,4-β-glucanase, Endo-K, produced by *Bacillus* sp. KSM-330 (Ozaki & Ito, 1991). Endo-K is active over an extremely narrow range of pH values, between 4.5 and 6.5, with an optimum pH at 5.2. One or more tryptophan residues in the enzyme molecule appear to be necessary for the expression of the endoglucanase activity. This report describes the cloning and sequencing of the gene encoding Endo-K to clarify the structure of the enzyme and to allow comparisons to be made between it and other alkaline or neutral endoglucanases.

Abbreviation: CMC, carboxymethylcellulose.

The nucleotide sequence data reported in this paper have been submitted to GenBank and have been assigned the accession number M68872.
Methods

Bacterial strains and cloning vectors. Bacillus sp. KSM-330 was used as the source of the gene for Endo-K. This micro-organism, a relative of Escherichia coli Bacillus subtilis, was isolated from a soil sample (Ozaki & Ito, 1991). Escherichia coli HB101 (F−, hsdS20 recA13 ara-14 proA2 lacY1 galK2 rpsL20 xyl-5 mtl-1 supE44 thi−1) and E. coli JM109 (recA1 Δlac− pro endA1 gyrA96 thi−1 hsdRI7 supE44 relA1 F′traD36 proAB IacPZ) were used as hosts for cloning and sequencing. Plasmid pBR322 and bacteriophages M13mp18 and M13mp19 were used as the vectors.

Media. E. coli strains harbouring plasmids were grown on LB broth or on M9CA medium (Maniatis et al., 1982), each of which contained 50 µg ampicillin ml−1 (sodium salt, Sigma). 2 × YT broth was used for cultivation of E. coli JM109 infected with M13 bacteriophage (Messing, 1983). P-medium contained (w/v) 1.0% Polypepton (Nihon Pharmaceutical), 0.5% yeast extract (Difco), 0.1% KH2PO4, 0.25% Na2HPO4, 12H2O and 0.02% MgSO4.7H2O, and was used for the propagation of Bacillus sp. KSM-330. Solid media contained 0.8 or 1.5% (w/v) Bacto agar (Difco).

Preparation of DNA. Bacillus sp. KSM-330 cells grown on P-medium at 30 °C for 24 h were harvested by centrifugation (9800 g, 10 min) at 4 °C. Genomic DNA was prepared from precipitated cells as described by Saito & Miura (1963). Plasmid DNA and the replicative form of M13 bacteriophage DNA were isolated by the alkaline extraction procedure of Birnboim & Doly (1979). Covalently closed circular DNA was purified by CsCl/ethidium bromide equilibrium density gradient centrifugation.

Cloning of the endoglucanase gene. Genomic DNA from Bacillus sp. KSM-330 and plasmid pBR322 DNA were both digested with HindIII (Boehringer Mannheim) and ligated with T4 DNA ligase (Boehringer Mannheim). Transformants that appeared on LB agar plates supplemented with ampicillin were overlaid with 1.0% (w/v) agar that contained 0.5% NaCl, 1 mg ml−1 mg lysozyme ml−1 and 50 mM citric acid/sodium citrate buffer (pH 5.2). After incubation at 37 °C for 5 h, production of the endoglucanase activity by the transformants was detected by staining the overlaid CMC with Congo red dye, as described by Teather & Wood (1982).

Assay of endoglucanase activity. E. coli carrying the recombinant plasmid was grown on LB broth supplemented with ampicillin at 37 °C for 24 h. Cells were harvested from 35 ml of the culture by centrifugation (9800 g, 10 min) at 4 °C and were resuspended in 7 ml 10 mM MOPS/NaOH buffer (pH 7.0). The suspension of cells was sonicated at 4 °C, and the cell debris was removed by centrifugation (12000 g, 15 min) at 4 °C. The endoglucanase activity of the supernatant was assayed as described previously (Ozaki & Ito, 1991). The preparation of enzyme was incubated at 40 °C for 60 min with 100 mM citrate buffer (pH 5.2) that contained 1% (w/v) CMC (degree of substitution 0.68, Sanyo Kokusaku Pulp), and the amount of reducing sugar liberated was measured by the 3,5-dinitrosalicylic acid procedure (Miller et al., 1960). One unit (U) of enzymatic activity was defined as the amount of enzyme that produced 1.0 µmol of reducing sugar as glucose per minute under the conditions of the reaction.

Hybridization analysis of DNA digests. HindIII digests of Bacillus sp. KSM-330 genomic DNA (5 µg) were subjected to electrophoresis on a 0.8% (w/v) agarose gel and then electrophoretically transferred to a Zeta-Probe blotting membrane (Bio-Rad) after denaturation by soaking the gel with 0.2 M NaOH/0.5 M NaCl for 30 min at room temperature. Approximately 0.4 µg of the HindIII insert, isolated from the recombinant plasmid, was labelled with digoxigenin-11-dUTP and used as a probe for hybridization analysis of the HindIII digests of the genomic DNA on the membrane, by use of a DNA-Labelling and Detection Kit (Boehringer Mannheim).

Nucleotide sequencing. Restriction fragments were subcloned into an appropriate position of the multiple cloning site of M13mp18 or M13mp19, and ordered deletion clones for sequencing were prepared by the method of Henikoff (1984), using a Kilo-Sequence Deletion Kit (Takara Shuzo). Single-stranded DNA was prepared by the procedure of Messing (1983). The nucleotide sequence was determined by the dideoxy chain-termination method of Sanger et al. (1977), using a fluorescent dye primer (Smith et al., 1986) and a modified T7 DNA polymerase (Sequenase, United States Biochemical), on an automated DNA sequencer (Applied Biosystems). Approximately 500 bases were read from a sample in each lane, and both strands of the endoglucanase gene were completely sequenced.

Results and Discussion

Cloning of the gene for Endo-K

Among several thousand ampicillin-resistant transformants of E. coli HB101, one was found to produce endoglucanase activity, as detected by the procedure for staining with Congo red dye. This endoglucanase-expressing transformant contained a 7.7 kb recombinant plasmid (pKC3301) carrying two HindIII inserts of 3·1 and 0·2 kb in the HindIII site of pBR322 (Fig. 1). E. coli HB101 harbouring pKC3301 produced intracellular endoglucanase activity at a rate of approximately 10 U per litre of broth, when cells were grown at 37 °C for 24 h on LB broth that contained ampicillin. The plasmid-encoded endoglucanase was active from pH 4·2 to 7·0 with a maximum activity at pH 5·2, and it was active from 10 to 60 °C with a maximum activity at 45 °C (data not shown). These characteristics of the activity correspond to those of the Endo-K secreted by Bacillus sp. KSM-330 (Ozaki & Ito, 1991).

To clarify the origin of the 3·1 kb insert of pKC3301, hybridization analysis was done using the digoxigenin-labelled 3·1 kb insert of pKC3301 as a probe. The labelled insert of pKC3301 hybridized with a 3·1 kb fragment in HindIII digests of the genomic DNA from Bacillus sp. KSM-330 blotted on a membrane (data not shown). Thus, the 3·1 kb HindIII insert in pKC3301 appeared to have been derived from the genomic DNA of Bacillus sp. KSM-330.

Location of the gene for Endo-K

The 3·1 kb insert isolated from pKC3301 was subcloned into the HindIII site of pBR322 using E. coli HB101 as host strain. Two plasmids, pKC330 and pKC330R, were obtained, carrying the insert in opposite orientations. The yields of Endo-K activity of E. coli HB101 that
The nucleotide sequence of the 1.8 kb HindIII–HpaI fragment was determined. Only one large open reading frame (ORF), beginning with an ATG codon at nucleotide 226 and ending with a TAA codon at nucleotide 161, was identified in the 1816 bp nucleotide fragment. This ORF, beginning with an ATG codon at nucleotide 181, was identified in the 1816 bp nucleotide fragment. Upstream from this ORF, there is a putative ribosome-binding site with a sequence of AAGGGGAGATGA, followed five bases later by a potential initiation codon, ATG. The sequence would exhibit a free energy of binding (ΔG) of −13.0 kcal mol⁻¹ (−54.6 kJ mol⁻¹), as calculated by the method of Tinoco et al. (1973), if it bound to the 3' end of the 16S rRNA from B. subtilis. The ΔG value and the distance between the initiation codon and this putative ribosome-binding site agree reasonably well with data for other reported ribosome-binding sites from several strains of Bacillus (Hager & Rabinowitz, 1985).

The region upstream from the ORF was scanned for sequences with homologies to various types of promoter consensus sequence recognized by RNA polymerases from B. subtilis (Doi & Wang, 1986). The sequence from nucleotides 35 to 65 resembles the consensus sequence of the sigma A-type vegetative promoters of B. subtilis (Moran et al., 1982). This sequence consists of TATAAT as the potential −35 region and TATATT as the potential −10 region, separated by 19 nucleotides. The sequence resembles the consensus sequence of sigma C-type promoters of B. subtilis (Johnson et al., 1983), found upstream from the putative ribosome-binding site of the gene from Bacillus sp. KSM-330. This sequence begins at nucleotide 181; it includes the potential −35 region of AAATT and the potential −10 region of AATTTATTTTA; and the two regions are separated by 14 nucleotides.

A large inverted-repeat sequence was found downstream from the termination codon of the ORF (from nucleotides 1691 to 1732). The ΔG value of this sequence for a stem–loop structure was calculated to be −19.8 kcal mol⁻¹ (−82.9 kJ mol⁻¹), which could be sufficient for termination of transcription. However, there is neither a GC-rich segment in the stem nor a poly(T) segment downstream from the stem, both of which are considered to be characteristics of p-independent terminators of transcription in bacterial genes (Adhya & Gottesman, 1978). Nevertheless, the combination of potential promoter and terminator sequences suggests that the gene for Endo-K may be monocistronic.

### Amino acid sequence analysis

The ORF in the nucleotide sequence encoded 463 amino acid residues, as indicated under the nucleotide sequence in Fig. 2. The molecular mass of the unprocessed protein was calculated to be 51 882 Da. There was a short, relatively basic, hydrophilic region, from amino acids 1 to 6, followed by a hydrophobic region that extended from amino acids 7 through 30. This hydrophilic–hydrophobic sequence resembles the signal peptide of extracellular proteins of Bacillus (Mézes & Lampen, 1985). The deduced sequence from amino acids 56 to 75 is identical to the amino-terminal 20 amino acid residues.
Fig. 2. Nucleotide sequence of the gene for Endo-K from Bacillus sp. KSM-330 and the deduced amino acid sequence of the encoded protein. The nucleotide sequence of the non-transcribed strand is given from 5' to 3', and the deduced amino acid sequence is shown under the nucleotide sequence. Putative -35 and -10 regions of the promoters are indicated by broken underlining. The putative ribosome-binding site is underlined with double broken underlining. Horizontal arrows indicate terminator-like, inverted-repeat sequences. The deduced amino acid sequence identical to the amino-terminal sequence of the Endo-K protein (Ozaki & Ito, 1991) is indicated by a line under the amino acid sequence.
of the Endo-K protein secreted by Bacillus sp. KSM-330 (Ozaki & Ito, 1991), and the sequence from amino acids 53 to 55 (Val-Ser-Ala) resembles the recognition site of signal peptidases (Perlman & Halvorson, 1983). The Endo-K protein in Bacillus sp. KSM-330 may possibly be cleaved by a signal peptidase at the bond between Ala 55 and Val 56 during secretion across the cytoplasmic membrane. However, the site of cleavage is far from the carboxy-terminus of the hydrophobic region and the possible signal peptide of 55 amino acid residues seems somewhat longer than other reported signal peptides of endoglucanases from Bacillus (Fukumori et al., 1986a; MacKay et al., 1986; Robson & Chambliss, 1987). The amino-terminus of Endo-K could, therefore, be processed after secretion by a proteolytic enzyme. The molecular mass of the mature enzyme was calculated to be 46090 Da, based upon the deduced amino acid sequence from amino acids 56 to 463. This size is slightly larger than the 42 kDa determined by SDS-PAGE of Endo-K that was purified from Bacillus sp. KSM-330 (Ozaki & Ito, 1991), suggesting that the carboxy-terminal region of Endo-K may be subject to proteolysis.

The endoglucanase activity of the purified Endo-K from Bacillus sp. KSM-330 is completely abolished by the presence of N-bromosuccinimide at low concentrations, and at least one Trp residue seems to be involved in the mechanism of action of this enzyme (Ozaki & Ito, 1991). The titrated number of total Trp residues in the enzyme molecule, 11.4, is very close to the number of Trp residues, 12, deduced from the nucleotide sequence of the gene.

Homology of the amino acid sequence with those of other enzymes

The deduced amino acid sequence of the Endo-K protein was compared with those of enzymes, reported to date, from various strains of Bacillus (Fukumori et al., 1986a, b; MacKay et al., 1986; Nakamura et al., 1987; Robson & Chambliss, 1987; Fukumori et al., 1989; Baird et al., 1990; Ozaki et al., 1990), Clostridium (Béguin et al., 1985; Grépinet & Béguin, 1986; Joliff et al., 1986; Hall et al., 1988; Schwarz et al., 1988; Zappe et al., 1988; Faure et al., 1989; Jauris et al., 1990), Cellulomonas (Nakamura et al., 1986; Wong et al., 1986), Ruminococcus (Ohmiya et al., 1989; Poole et al., 1990; Wang & Thomson, 1990), Butyrivibrio (Berger et al., 1989; Hazlewood et al., 1990), Pseudomonas (Hall & Gilbert, 1988), Erwinia (Guiseppi et al., 1988), Streptomyces (Nakai et al., 1988) and Trichotherma (Penttilä et al., 1986; Saloheimo et al., 1988). The sequence from amino acids 56 to 443 of Endo-K exhibited 30% homology to that of the celA endoglucanase from Clostridium thermocellum NCIB10682 (Béguin et al., 1985). In addition, 25% homology in terms of amino acid sequence was observed between Endo-K (from amino acids 130 to 273) and the endoglucanase from Cellulomonas uda CB4 (Nakamura et al., 1986). No such homology was observed between Endo-K and other endoglucanases from Bacillus. Henrissat et al. (1989) compared the amino acid sequences of cellulolytic enzymes from various origins by means of hydrophobic cluster analysis (Gaboriaud et al., 1987) and classified them into six families (family A through family F). According to their criteria, Endo-K seems to belong to family D, because the endoglucanases from C. thermocellum NCIB 10682 and Cellulomonas uda CB4 have been classified as belonging to this family. A number of genes for endoglucanases from Bacillus have been sequenced, and some of the proteins (Fukumori et al., 1986a, b; MacKay et al., 1986; Nakamura et al., 1987; Robson & Chambliss, 1987) have been classified as members of family A, and others (Fukumori et al., 1989; Baird et al., 1990; Ozaki et al., 1990) appear to belong to the same family because they are similar to the enzymes in family A. Therefore, Endo-K is the first case of an enzyme from Bacillus that belongs to family D.

The amino acid sequences from amino acids 130 to 156 (region I) and from amino acids 189 to 207 (region II) of Endo-K appear to be well conserved when suitably aligned with sequences of endoglucanases in family D, as shown in Fig. 3. This result implies that some of the conserved amino acids may play important roles in the action of these endoglucanases. Among the conserved amino acids, the Trp residue found in region II was notable, in that at least one Trp residue has been proved to be involved in the mechanism of the action of Endo-K (Ozaki & Ito, 1991). The celA endoglucanase has a carboxy-terminal domain that consists of a direct repeat of 23 amino acids (Béguin et al., 1985), as is conserved in other endoglucanases from C. thermocellum (Grépinet & Béguin, 1986; Joliff et al., 1986; Hall et al., 1988). These domains are unnecessary for both the catalytic activity
Such sequences of the reiterated amino acids and the sequences which are rich in Pro, Thr and Ser residues. pH optima between the acid and alkaline endoglucanases that this enzyme may be a single domain protein.

The aim of this study was to explain the difference in pH optima between the acid and alkaline endoglucanases by site-directed mutagenesis.

References


Acid cellulase gene from Bacillus


