Molecular Cloning and Nucleotide Sequence of the Alkaline Cellulase Gene from the Alkalophilic Bacillus sp. Strain 1139

By FUMIYASU FUKUMORI,* TOSHIAKI KUDO, YOSHIKO NARAHASHI AND KOKI HORIKOSHI
The Riken Institute, Wako-shi, Saitama-ken 351-01, Japan

(Received 19 November 1985; revised 10 March 1986)

The cellulase gene from the alkalophilic Bacillus sp. strain 1139 was cloned in Escherichia coli using pBR322. Plasmid pFK1 was isolated from transformants producing cellulase, and the cloned cellulase gene was found to be in a 4.6 kb HindIII fragment. The cellulase gene was subcloned in a functional state on a 2.9 kb DNA fragment and its nucleotide sequence was determined. The coding sequence showed an open reading frame encoding 800 amino acids. The pFK1-encoded cellulase had the same enzymic properties as the extracellular cellulase produced by the alkalophilic Bacillus sp. strain 1139, but its $M_r$ was slightly higher.

INTRODUCTION

Cellulases (1,4-$\beta$-D-glucan glucanohydrolase, EC 3.2.1.4) have been detected in many microorganisms. Most cellulases are multienzyme systems and are active over acidic or neutral pH ranges. The structural genes for the multienzymes of Clostridium thermocellum (Cornet et al., 1983) and Cellulomonas fimii (Gilkes et al., 1984) have been cloned in Escherichia coli. Recently, the nucleotide sequences of the cellulase gene of Clostridium thermocellum (Béguin et al., 1985) and the $\beta$-glucanase gene of Bacillus subtilis (Murphy et al., 1984) were determined.

In our laboratory, many alkalophilic bacteria which grow well at high pH (10–11) have been isolated (Horikoshi & Akiba, 1982). These bacteria produce many extracellular enzymes which have optimum pH values for activity in the alkaline region. One of these bacteria, the alkalophilic Bacillus sp. strain N-4, produced several cellulases which had broad pH activity ranges (pH 5–10–9) (Horikoshi et al., 1984), and two cellulase genes from this organism have been cloned (Sashihara et al., 1984). Recently, an alkalophilic Bacillus sp., strain 1139, which produced one alkaline cellulase (pH optimum 9.0) was isolated. This strain was not a true cellulolytic micro-organism because the enzyme was unable to hydrolyse native cellulose (Fukumori et al., 1985). We have now cloned the cellulase gene of Bacillus sp. 1139. This paper deals with the molecular cloning of the gene and the determination of its nucleotide sequence. We also describe some of the properties of the enzyme.

METHODS

Bacterial strains and plasmid. Bacillus sp. strain 1139, producer of an alkaline cellulase, was isolated from soil (Fukumori et al., 1985). Other bacterial strains were E. coli K12 strain HB101 (Pro-$\beta$-led B1 lac Y hsdR hsdM ara-l4 galK2 xyl-5 met-l supE44 F- endo1 recA Str') and JM103 (Messing et al., 1981). Plasmid pBR322 was used throughout.

Media. LB medium (Maniatis et al., 1981) was used for E. coli, and PY-CMC (Fukumori et al., 1985) for Bacillus sp. 1139.

Preparation of DNA. Bacillus sp. 1139 was grown aerobically to the early stationary phase at 37°C in PY-CMC medium. Bacterial chromosomal DNA was purified by the methods of Saito & Miura (1963). The vector pBR322 was purified by the method of Bolivar et al. (1977). Recombinant plasmids were detected by the boiling method (Davis et al., 1980).

0001-3093 © 1986 SGM
Construction of recombinant plasmids. DNAs were digested with HindIII at 37 °C for 1 h (plasmid DNA) or for 14 h (chromosomal DNA). After digestion, 1 μg of plasmid and 3 μg of bacterial chromosomal DNA were mixed and ligated with T4 DNA ligase overnight at 15 °C. This ligation mixture was used for transformation (Lederberg & Cohen, 1974).

Nick translation and hybridization of DNA digests. DNA was labelled by nick translation, using E. coli DNA polymerase I in the presence of $[^{32}P]$dATP (Rigby et al., 1977). The hybridization technique of Southern (1975) was used.

Assay of cellulase. Enzyme activity was measured as described previously (Fukumori et al., 1985).

Purification of cellulase. Cellulase was purified from Bacillus sp. 1139 as described by Fukumori et al. (1985).

Preparation of antiserum. Antiserum against cellulase from Bacillus sp. 1139 was prepared as described previously (Sashihara et al., 1984).

Amino acid sequence analysis. NH$_2$-terminal amino acid sequence determinations were performed with a Beckman protein/peptide sequencer, model 890-M, using programme selection 3 (cell temp. 57 °C, converter temp. 50 °C) with 0-1 m-Quadrol, and Beckman reagents and solvents (Edman & Henshen, 1975). Phenylthiohydantoin derivatives of amino acids were identified in a Beckman HPLC system: model 420 controller, 110A pumps, model 165 detector (269 and 319 nm), Ultrasphere ODS column (5 μm, 250 × 46 mm, 50 °C) with trifluoroacetate (TFA)/acetate buffer system (Hawke et al., 1982). The buffers used were slightly modified as follows: solvent A, acetonitrile and TFA/acetate buffer (18.3 mm/0.4 mM, pH 6.2 with NaOH) (1:9, v/v) and solvent B, acetonitrile/TFA buffer (3.8 mM, pH 3.5) (3:1, v/v).

DNA sequencing. This was done by the dideoxy chain termination method (Sanger et al., 1977), specific restriction fragments being cloned into mp8 or mp9 M13 vectors for dideoxy sequencing (Messing, 1983) by using [α-32P]dCTP.

Enzymes and chemicals. RNAase, lysozyme, DNAase, ampicillin (Ap) and tetracycline (Tc) were purchased from Sigma. Restriction endonucleases, DNA polymerase I and T4 ligase were obtained from Bethesda Research Laboratories. M13 sequence kit (Takara Syuzo, Kyoto, Japan) was used for DNA sequencing.

RESULTS

Cloning of the cellulase gene in E. coli. Ligation mixtures were used to transform E. coli HB101, and about 10$^4$ Ap$^+$ Tc$^+$ transformants per μg DNA were obtained (15% of the Ap$^+$ transformants were Tc$^+$). Eight transformants produced shallow craters around colonies (Fukumori et al., 1985) on LB agar plates containing carboxymethylcellulose. Plasmid pFK1 was obtained from an Ap$^+$ Cellulase$^+$ transformant which contained a 4.6 kb HindIII fragment. This plasmid could transform E. coli HB101 to Ap$^+$ Cellulase$^+$ at a high frequency.

Restriction map of pFK1. pFK1 DNA was digested with several endonucleases, and the digests were analysed by agarose gel electrophoresis. The restriction map of this plasmid is shown in Fig. 1. The 4-6 kb HindIII fragment was further digested with HincII, KpnI, PvuII and XhoI.

Homology between the cloned fragment and chromosomal DNA. To analyse the origin of the DNA inserted in pBR322, 32P-labelled pFK1 was hybridized to restriction enzyme-digested chromosomal DNAs of Bacillus sp. 1139 and E. coli HB101 which had been immobilized on nitrocellulose sheets. pFK1 hybridized to a 4-6 kb HindIII fragment. This plasmid could transform E. coli HB101 to Ap$^+$ Cellulase$^+$ at a high frequency.

Expression and enzymic properties of cellulase in E. coli.

(i) Localization. E. coli HB101 (pFK1) was grown aerobically in LB broth for 24 h at 37 °C. The extracellular, periplasmic and cellular cellulase activities (Kato et al., 1983) were 66 (11.3%), 402 (68.8%) and 116 (19.9%) mU per ml of culture, respectively. The synthesis of
cellulase in *E. coli* was constitutive and no effect of carboxymethylcellulose supplementation was observed.

(ii) The optimum pH for the cellulase activity of *E. coli* HB101(pFK1) was 9.0, as was observed for *Bacillus* sp. 1139 (Fukumori et al., 1985).

(iii) Immunological studies. The periplasmic cellulase of *E. coli* HB101(pFK1) was tested against antiserum prepared against the cellulase of *Bacillus* sp. 1139. Periplasmic cellulase of *E. coli* HB101(pFK1) gave a line of precipitation that fused with that for cellulase from *Bacillus* sp. 1139 (Fig. 4). No reaction was observed with the periplasmic fraction of *E. coli* HB101(pBR322). In the activity assay system, the addition of 15 μl antiserum caused 100% inhibition of the cellulase activity (about 200 μU) from *E. coli* HB101(pFK1).

(iv) The *M*₉ of pFK1-encoded cellulase was estimated to be 94,000 by the SDS-PAGE method of Laemmli (1970), using an immunological assay kit (Bio-Rad).

**Determination of amino acid sequence.** The amino acid composition of the extracellular cellulase of *Bacillus* sp. 1139 is presented in Table 1. Molar ratios obtained by amino acid analysis (Fukumori et al., 1985) were closely consistent with those derived from the DNA sequence. The NH₂-terminal sequence of the purified cellulase was determined by automated Edman degradation up to the 10th residue to be Glu–Gly–Asn–Thr–Arg–Glu–Asp–Asn–Phe–Lys–. This amino acid sequence was identical to that deduced from the DNA sequence. Therefore, 29 or 30 amino acid residues [residues -29 (-30) to -1] may be a signal peptide which is removed during the secretion process.
DNA sequence of an alkaline cellulase gene

1750
GGAACAGAAGTTGAAATTCCAGTTGTTCATGATCCAAAAGGAGAAGCTGTTCTTCCTTCTGTTTTTGAAGACGGTACACGTCAAGGTTGG
GlyThrGluValGluIleProValValHisAspProLysGlyGluAlaValLeuValSerValPheGluAspGlyThrArgGlnGlyTrp

1850
GACTGGGCTGAGTCTGCGTGGAAACACGTCTTAACATTTGAAAGCAAGCAAGCCTTCCCATACGCTGCTTGGAAATGCTAACT
GluValValProSerAspValThrAlaThrAlaArgLeuAspPheProAspProValArgAlaThrGlyAlaAlaMetAsnIleAsnLeuValPheGluProProThrAsnGlyTyrTrpVal

1900
GACTGGGCTGAGTCTGCGTGGAAACACGTCTTAACATTTGAAAGCAAGCAAGCCTTCCCATACGCTGCTTGGAAATGCTAACT
GluValValProSerAspValThrAlaThrAlaArgLeuAspPheProAspProValArgAlaThrGlyAlaAlaMetAsnIleAsnLeuValPheGluProProThrAsnGlyTyrTrpVal

1950
GTTGATTCTATCTGATCTCGGCTGGAAACACGTCTTAACATTTGAAAGCAAGCAAGCCTTCCCATACGCTGCTTGGAAATGCTAACT
GluValValProSerAspValThrAlaThrAlaArgLeuAspPheProAspProValArgAlaThrGlyAlaAlaMetAsnIleAsnLeuValPheGluProProThrAsnGlyTyrTrpVal

2000
GTTGATTCTATCTGATCTCGGCTGGAAACACGTCTTAACATTTGAAAGCAAGCAAGCCTTCCCATACGCTGCTTGGAAATGCTAACT
GluValValProSerAspValThrAlaThrAlaArgLeuAspPheProAspProValArgAlaThrGlyAlaAlaMetAsnIleAsnLeuValPheGluProProThrAsnGlyTyrTrpVal

2050
GTTGATTCTATCTGATCTCGGCTGGAAACACGTCTTAACATTTGAAAGCAAGCAAGCCTTCCCATACGCTGCTTGGAAATGCTAACT
GluValValProSerAspValThrAlaThrAlaArgLeuAspPheProAspProValArgAlaThrGlyAlaAlaMetAsnIleAsnLeuValPheGluProProThrAsnGlyTyrTrpVal

2100
GTAAGAGATATTACAAACATTCAAGATGACACGTTACTACGTAACA'rGATGATCATTTTTGCAGATGTAGAAAGTGACTTTGCAGGGAGA
ValArgAspIleThrAsnIleGlnAspAspThrLeuLeuArgAsnMetMetIleIlePheAlaAspValGluSerAspPheAlaGlyArg

2150
GTAAGAGATATTACAAACATTCAAGATGACACGTTACTACGTAACA'rGATGATCATTTTTGCAGATGTAGAAAGTGACTTTGCAGGGAGA
ValArgAspIleThrAsnIleGlnAspAspThrLeuLeuArgAsnMetMetIleIlePheAlaAspValGluSerAspPheAlaGlyArg

2200
GTAAGAGATATTACAAACATTCAAGATGACACGTTACTACGTAACA'rGATGATCATTTTTGCAGATGTAGAAAGTGACTTTGCAGGGAGA
ValArgAspIleThrAsnIleGlnAspAspThrLeuLeuArgAsnMetMetIleIlePheAlaAspValGluSerAspPheAlaGlyArg

2250
GTAAGAGATATTACAAACATTCAAGATGACACGTTACTACGTAACA'rGATGATCATTTTTGCAGATGTAGAAAGTGACTTTGCAGGGAGA
ValArgAspIleThrAsnIleGlnAspAspThrLeuLeuArgAsnMetMetIleIlePheAlaAspValGluSerAspPheAlaGlyArg

2300
GTAAGAGATATTACAAACATTCAAGATGACACGTTACTACGTAACA'rGATGATCATTTTTGCAGATGTAGAAAGTGACTTTGCAGGGAGA
ValArgAspIleThrAsnIleGlnAspAspThrLeuLeuArgAsnMetMetIleIlePheAlaAspValGluSerAspPheAlaGlyArg

2350
GTAAGAGATATTACAAACATTCAAGATGACACGTTACTACGTAACA'rGATGATCATTTTTGCAGATGTAGAAAGTGACTTTGCAGGGAGA
ValArgAspIleThrAsnIleGlnAspAspThrLeuLeuArgAsnMetMetIleIlePheAlaAspValGluSerAspPheAlaGlyArg

2400
GTAAGAGATATTACAAACATTCAAGATGACACGTTACTACGTAACA'rGATGATCATTTTTGCAGATGTAGAAAGTGACTTTGCAGGGAGA
ValArgAspIleThrAsnIleGlnAspAspThrLeuLeuArgAsnMetMetIleIlePheAlaAspValGluSerAspPheAlaGlyArg

2450
GTAAGAGATATTACAAACATTCAAGATGACACGTTACTACGTAACA'rGATGATCATTTTTGCAGATGTAGAAAGTGACTTTGCAGGGAGA
ValArgAspIleThrAsnIleGlnAspAspThrLeuLeuArgAsnMetMetIleIlePheAlaAspValGluSerAspPheAlaGlyArg

2500
GTAAGAGATATTACAAACATTCAAGATGACACGTTACTACGTAACA'rGATGATCATTTTTGCAGATGTAGAAAGTGACTTTGCAGGGAGA
ValArgAspIleThrAsnIleGlnAspAspThrLeuLeuArgAsnMetMetIleIlePheAlaAspValGluSerAspPheAlaGlyArg

Fig. 3. Nucleotide sequence of the cellulase gene from Bacillus sp. 1139. The DNA sequence of the coding strand is given from 5' to 3', numbered from nucleotide 1 at the putative initiation site. The proposed ribosomal binding site (SD) is underlined with a dashed line. The predicted amino acid sequence is given below the DNA sequence. The amino acids are numbered taking the NH₂-terminal amino acid of the matured protein as 1. Underlined amino acids have been determined by automated Edman sequencing of the purified cellulase. The hairpin loop of the putative rho-independent terminator site is underlined, and the poly-T region is overlined.

Fig. 4. Ouchterlony double-diffusion analysis. Well AB contained antibody to the purified cellulase from Bacillus sp. 1139. Samples were added to the outer wells as follows: well 1, periplasm fraction of E. coli HB101(pFK1) (270 µM, 200 µg protein); well 2, purified enzyme of Bacillus sp. 1139 (270 µM, 4.5 µg protein); well 3, periplasm fraction of E. coli HB101(pBR322) (200 µg protein).
Table 1. Amino acid composition of the extracellular cellulase from the alkalophilic Bacillus sp. 1139

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>No.</th>
<th>Mol %</th>
<th>Amino acid</th>
<th>No.</th>
<th>Mol %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lys</td>
<td>31</td>
<td>4.0</td>
<td>Gly</td>
<td>58</td>
<td>7.5</td>
</tr>
<tr>
<td>His</td>
<td>13</td>
<td>1.8</td>
<td>Ala</td>
<td>64</td>
<td>8.3</td>
</tr>
<tr>
<td>Arg</td>
<td>24</td>
<td>3.2</td>
<td>Cys</td>
<td>0</td>
<td>ND</td>
</tr>
<tr>
<td>Trp</td>
<td>20</td>
<td>2.6</td>
<td>Val</td>
<td>60</td>
<td>7.8</td>
</tr>
<tr>
<td>Asx</td>
<td>126</td>
<td>16.4</td>
<td>Met</td>
<td>13</td>
<td>1.7</td>
</tr>
<tr>
<td>Thr</td>
<td>40</td>
<td>5.2</td>
<td>Ile</td>
<td>44</td>
<td>5.7</td>
</tr>
<tr>
<td>Ser</td>
<td>41</td>
<td>5.3</td>
<td>Leu</td>
<td>46</td>
<td>6.0</td>
</tr>
<tr>
<td>Glx</td>
<td>91</td>
<td>11.8</td>
<td>Tyr</td>
<td>25</td>
<td>3.2</td>
</tr>
<tr>
<td>Pro</td>
<td>48</td>
<td>6.2</td>
<td>Phe</td>
<td>26</td>
<td>3.4</td>
</tr>
</tbody>
</table>

ND, Not detected.

DISCUSSION

Most cellulases are active at acidic or neutral pH. The genes for two of these cellulases have been cloned and sequenced (Béguin et al., 1985; Murphy et al., 1984). When the DNA sequence of the cellulase gene from the alkalophilic Bacillus sp. 1139 is compared with the sequences of the genes from Clostridium thermocellum (Béguin et al., 1985) and B. subtilis (Murphy et al., 1984), no distinct homologies in either amino acid sequence or nucleotide sequences are observed. A DNA sequence coding for a direct repeat of a 23 amino acid sequence was found in the cellulase gene of C. thermocellum but was not seen in the cellulase gene analysed here. Perhaps this direct repeat sequence is not essential for cellulase activity.

We have cloned two cellulase genes (pNK1 and pNK2) of another alkalophilic strain, Bacillus sp. N-4 (Sashihara et al., 1984). These cellulases had broad optimum pH ranges (pH 5-10) compared with the cellulase (pH optimum 9) from Bacillus sp. 1139. The nucleotide sequence of one of the cellulase genes from alkalophilic Bacillus sp. N-4 (unpublished data) exhibits partial homology with the middle region of the DNA sequence of the pFK1-encoded cellulase gene. The sequence AGGAGG (McLaughlin et al., 1981), which is highly complementary to the 3' end of B. subtilis 16S rRNA (Murray & Rabinowitz, 1982), was observed upstream of the open reading frame; the free energy of the site was $-18.8 \text{kcal mol}^{-1}$ ($-78.7 \text{kJ mol}^{-1}$) according to Tinoco et al. (1973). Although the precise translation initiation site has not been determined, cellulase must be translated from ATG codons located 3 or 6 bp downstream of this putative ribosome-binding site. The cellulase gene of Bacillus sp. 1139 is followed by a typical prokaryotic rho-independent transcription terminator sequence (Rosenberg & Court, 1979). The $\Delta G'$ calculated for the corresponding RNA hairpin structure (Tinoco et al., 1973) is $-18.4 \text{kcal mol}^{-1}$ ($-78.7 \text{kJ mol}^{-1}$). The putative signal sequence was 29 or 30 amino acid residues long, with four basic amino acids near the NH$_2$-terminal end followed by a hydrophobic region and alanine at the COOH-terminal. These observations are consistent with the signal peptides from other Gram-positive bacteria (Béguin et al., 1985; Kato et al., 1985; Murphy et al., 1984; Palva et al., 1981; Stephens et al., 1984; Yang et al., 1983).

The plasmid pFK1-encoded cellulase had the same properties as the cellulase from the alkalophilic Bacillus sp. 1139, but its $M_r$ was slightly higher (94000 vs 92000, as estimated by SDS-PAGE). This difference might be due to processing of the protein. It is possible that the cellulase secreted into the periplasmic space of E. coli may be processed at a different site(s) during the secretion process.

We thank Dr W. D. Grant for reading the manuscript and for many useful discussions, and Zeria Pharmaceutical Co. Ltd. for preparation of antiserum.

This research was partially supported by grants from Life Science Promotion, The Riken Institute and from the Ministry of Education, Science, and Culture of Japan.
DNA sequence of an alkaline cellulase gene

REFERENCES


