KOR. JOUR. MICROBIOL. September, 1986 p. 236-242 Vol. 24, No. 3

Analysis on the Nucleotide Sequence of the Signal Region of *Bacillus subitilis* Extracellular Cellulase Gene

Seo, Y.S., Y.H. Lee, U.H. Pek and H.S. Kang*

Department of Genetic Engineering, Doosan Research Lab. Seoul Yoido. P.O. Box80 Korea

**Department of Microbiology, College of Natural Sciences, Seoul National University, Seoul, Korea*

*Bacillus subtilis*로부터 분리한 cellulase 유전자의 조절부위에 대한 염기서열분석

서연수 · 이영호 · 백운화 · 강현삼*

두산연구소 유전공학연구실, *서울대학교 자연과학대학 미생물학과

Abstract: The nucleotide sequence of the genetic control site of *Bacillus subtilis* gene for (1-4) - β -D-glucan endoglucanase (cellulase) was determined according to the procedures of the dideoxy chain termination method(Sanger *et. al.*, 1977). The deduced amino acid sequence of this enzyme has a hydrophobic signal peptide at the NH_2 terminus similar to those found in fifteen other extracellualr enzymes from *Bacillus* species. This is followed by a sequence resembling the *Bacillus* ribosome binding site 14 nucleotide before the first codon of the gene. The presumptive promoter sequence was located 92 base pairs upstream from the initiation codon. The homology region in signal sequences was striking when comparing all the signal sequences of sixteen extracellular enzymes from *Bacillus* species so far compiled.

Key words: *Bacillus subtilis*, cellulase, DNA sequencing, signal sequence.

The genus *Bacillus* is an industrially important source of secreted enzymes. There have been several reports of the cloned and secreted enzymes from *Bacillus*. These include cellulase from *B. subtilis* which is an extracellular enzyme that can hydrolyze carboxymethyl cellulose, cereal β -glucan and lichenan and the molecular weight of this enzyme is known to be 33,000±1,000 daltons (Lee and Park, 1985). The purpose of this research is to elucidate the nucleotide sequence of the 5'-flanking region of the previously cloned *B. subtilis* cellulase gene (Seo *et. al.*, 1985) in order to produce this enzyme extracellularly in yeast for the degradation of barley β -glucan.

This facilitates greatly the filtration of beer and allows the use of higher proportion of unmalted cereals in the griest (Bamforth and Martin, 1981). Informations are now available on the transcriptional control sequence of *B. subtilis* and promoter sequences have been characterized for the σ^{55}-RNA polymerase of *B. subtilis* (Moran *et. al.*, 1982). DNA and amino acids sequencing data on the secreted proteins from *Bacillus* species have been accumulated enough to allow the comparison

between the signal sequences of *Bacillus* extracellular proteins. They include type I β-lactamase (Mezes *et. al.*, 1983; Sloma and Gross, 1983; Wang *et. al.*, 1985), type II β-lactamase (Hussain *et. al.*, 1985) from *B. cereus*, penicillinase (Neugebauer *et. al.*, 1981) and α-amylase (Stephens *et al.*, 1984) from *B. licheniformis*, neutral protease (Vasantha *et. al.*, 1984), subtilisin (Wells *et. al.*, 1983; Vasantha *et. al.*, 1984), α-amylase (Takkinen *et. al.*, 1983), and barnase (Paddon and Hartely, 1986) from *B. amyloliquefaciens*, neutral protease (Takaki *et. al.*, 1985) and α-amylase (Nakajima *et. al.*, 1985) from *B. stearothermophilus*, and α-amylase (Yang *et. al.*, 1983; Ohmura *et. al.*, 1983), β-glucanase (Murphy *et. al.*, 1984), subtilisin (Stahl *et. al.*, 1984; Wong *et. al.*, 1984), neutral protease (Yang *et. al.*, 1984), and levansucrase (Steinmetz *et. al.*, 1985) from *B. subtilis*. Comparison between amino acid sequences deduced from the above-listed sequencing data shows that there exist highly conserved regions of homology in the hydrophobic signal. We report here the nucleotide sequence of a 769 bp DNA fragment containing the genetic control site and the sceretion signal region of *Bacillus subtilis* for the extracellular cellulase.

MATERIALS AND METHOD

Bacterial strains and plasmids

The plasmid pBAG10 (Fig.1), a derivative of pBAG1 containing the cellulase gene from *Bacillus subtilis* ATCC6633 was used as a source of the cellulase gene. *E. coli* HB101 was used as a host for plasmid transformation. *E. coli* JM109 (Yanisch-Perron *et. al.*, 1985) was used for M13 RF DNA transformation and M13 phage propagation.

Media and growth conditions

Bacteria were routinely grown in L broth (10g Difco tryptone, 5g Difco yeast extract, 5g NaCl/liter adjusted to pH7.2). When necessary, ampicillin (50 μg/ml) was added to L broth. The medium was solidified by the addition of Difco Bacto agar (20g/liter).

Purification of plasmid of M13 RF DNA

CsCl-ethidium bromide equilibrium centrifugation of cleared lysate (Guerry *et. al.*, 1973) was used for the large scale isolation of plasmid and M13 RF DNA. Rapid preparation was performed according to the procedure of Birnboim and Doly (1979).

Enzymes and reagents

Restriction enzymes, T4 DNA ligase, and Klenow fragment of DNA polymerase I were

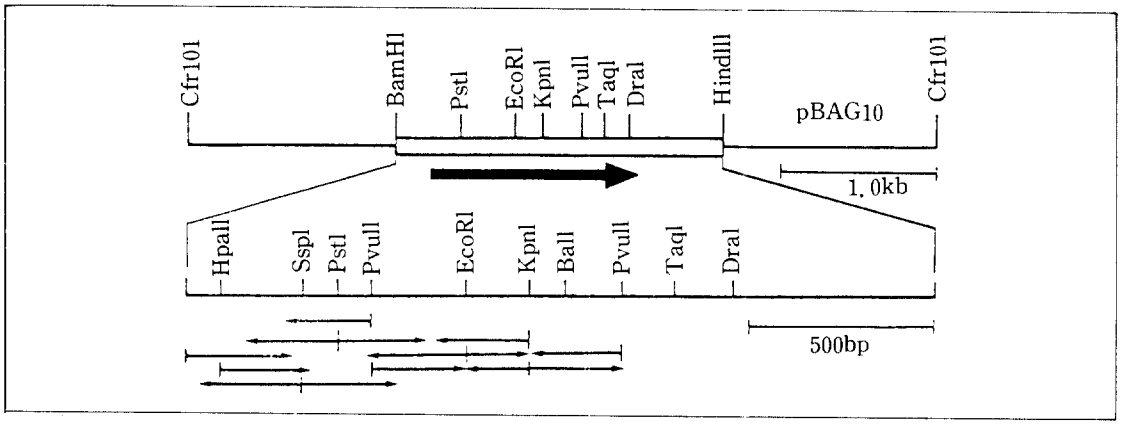

Fig. 1. *Structure of plasmid pBAG10 and sequencing strategy for the B. subtilis cellulase gene.*
The open bar represents the cloned chromosomal DNA fragment which contains the intact cellulase gene. The transcriptional direction and the deduced size of this gene are indicated by a large thick arrow. DNA fragment were subcloned into the polylinker of M13mp18 or M13mp19 RF DNA and sequences were determined by the Sanger dideoxy chain termination method. The thin arrows indicate the direction and extent of the DNA sequence analysis.

purchased from Bethesda Research Laboratories (BRL), New England Laboratories (NEB), or Takara Shuzo. M13 sequencing kit was obtained from NEB. α-^{32}P-dATP was purchased from Amersham. Each was used according to the manufacturer's specification. Major supplier of other chemicals was Sigma Chemical Company.

Transformation

Transformation of *E. coli* HB101 was performed according to Kushner (1978). After a heat pulse at 43 ℃ for 30 seconds, the cells were diluted 10-fold in L-broth and incubated at 37 ℃ for one and half hour to allow for expression before plating on L-agar plates containing the appropriate antibiotic. When selecting for recombinants of M13 RF DNA, the transformants were mixed with a 3ml overlaying agar (0.6%) containing 0.2 ml fresh grown JM109, 0.1ml of 2% X-gal (5-bromo-4chloro-3-indoyl-βD-galactoside) and 0.02ml of 100mM IPTG (isopropyl-β-D-thiogalactopyranoside) and then poured onto L-agar.

DNA sequence determination

DNA sequencing was carried out by dideoxy chain termination method (Sanger *et. al.*, 1977) with a M13 sequencing kit. pBAG10 DNA was double digested so that all the fragments generated had non-symmetrical restriction ends. DNA fragments to be sequenced were purified from agarose by electroelution. Specific DNA fragments were cloned into the M13 mp18 or M13mp19 RF DNA (Yanisch-Perron *et. al.*, 1985). The sequence was determined on both strands and sequences including the

Fig. 2. *Nucleotide and deduced amino acid sequence of the signal region of B. subtilis cellulase gene.*

The nucleotide sequence is presented from BamHI site (nucleotide -325) to EcoRI site (nucleotide +444). The nucleotide sequence is counted from the first base of the initiation codon (ATG) as +1. The inferred amino acid sequence is shown beneath the nucleotide sequence. The hydrophobic core of the signal peptide (amino acids 1 to 29) is underlined. Presumptive "-35" and "-10" and Shine-Dalgarno sequence are indicated with thick solid lines. The putative signal cleavage site is marked (▲).

restriction sites used for cloning were resequenced in other clones to avoid potential errors due to closely spaced restriction sites.

RESULTS AND DISCUSSION

Sequence analysis and secretion signal

We have previously shown that the cellulase gene was contained in a functional state in a 2.1 kilobase insert in plasmid pBR322 and have determined its transcriptional direction (Seo. *et. al.*, 1985). This insert was recloned into pUC19 (Yanisch-Perron *et. al.*, 1985) to facilitate the isolation of fragment to be sequenced. This plasmid was designated to pBAG10 (Fig.1), which was used as source of the cellulase gene. The sequencing strategy and restriction map for those enzymes used to determine the sequence are shown in Fig.1. DNA sequencing was performed extensively in the 5'-flanking region of the cellulase gene extending from BamHI to PvuII. Only the sequence of the BamHI-EcoRI fragment (769 bp) is presented here (Fig.2). The sequence of Fig.2 was searched for an open reading frame for the cellulase. There is only one such sequence that satisfies the conditions that are required to be a gene for an extrcellular enzyme. The open reading frame is extended to the PvuII site without any interruption of termination codons (data not shown). The deduced N-terminal extension has the characteristics of a cleavage signal peptide. It has a positively charged short N-terminal segment with three residues (Met-1-Lys-Arg-3). This is followed by 26 amino acid residues that are most hydrophobic, otherwise neutral (Ser-4 to Ala -29). The last 10 residues of the signal peptide are mostly small amino acids (Gly-20-Gly -Met-Pro-Ala-Pro-Gln-Ala-Ser-Ala-29). Beginning at amino acid 27 is located a potential signal peptidase recognition site with Ala-X (Ser here)-Ala, which appears most frequently between the signal sequence and the N-terminal amino acid of the mature protein. The putative signal peptidase cleavage at the bond between Ala-29-Val-30 is also compatible with the empirical rules (e.g., -3, -1 rule; window of cleavage sites) constructed by von Heijne (1983, 1984) and Perlman and Halverson (1983). Moreover, the cellulase sequence shows homology with the other *Bacillus* signal sequence as shown in Fig.3. This kind of homology is not known yet in other organisms (Watson, 1984).

Murphy *et. al.* (1984) suggested there was some homology in signal sequences of 5 extrcellular enzymes from different *Bacillus* species. We confirmed the homology sequences of *Bacillus* extracellular enzymes so far compiled(Fig.3). The first homologous sequence of Leu-Leu or Leu-Phe occurs at the fixed position (4 or 7) relative to the end of the hydrophilic N-terminal segment in twelve out of sixteen sequences including the cellulase while there is minor variation in this position from the hydrophilic end. This sequence repeats immediately two or three amino acids after the first homolgous one. It is apparent from these considerations that cellulase has a signal sequence typical of those so far identified in *Bacilli*.

Transcriptional and translational signals

Potential transcription initiation and translation signals can be discerned from the nucleotide sequence. This 769 base pair DNA sequence has been scanned for the sequences related to the -35(TTGACA) and -10(TATAAT) consensus sequences which are usually separated by 17-18 base pairs. There is only one sequence which was in agreement with the consensus and was located 92 base pairs upstream from the ATG of the cellulase gene. It corresponds closely to the consensus -35 and -10 which are separated by 18 base pair (Fig. 2). Moreover 20 out of 30 of the base pairs upstream from the -35 region are AT which is typical of σ^{55} promoters (Moran *et. al.*, 1982). Transcription usually starts at an A or G, six or seven nucleotides downstream from the -10 sequence which suggests that the cellulase transcript starts at the G at -85 which is seven nucleotides distant from the -10 sequence. The putative promoter is followed by

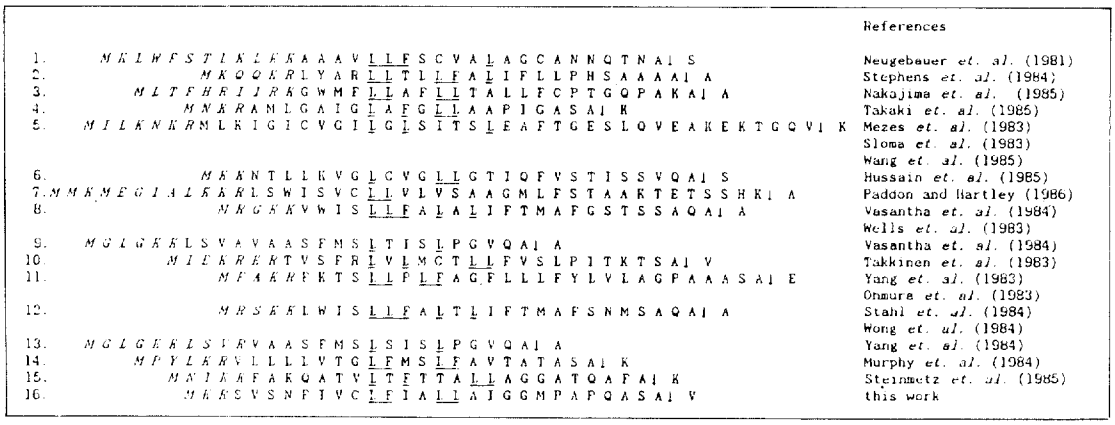

Fig. 3. *N-terminal amino acid sequences of precusors of extracellular enzymes of Bacilli.*

The underlines show the highly conserved regions of signal sequences in *Bacilli* which is evident when the sequences are aligned at the first homologous regions. The hydrophilic regions are italicized. The known or predicted cleavage sites are indicated by an arrow(↓). Amino acids are symbolized by the one-letter code, i.e. A=Ala, C=Cys, D=Asp, E=Glu, F=Phe, G=Gly, H=His, I=Ile, K=Lys, L=Leu, M=Met, N=Asn, P=Pro, Q=Gln, R=Arg, S=Ser, T=Thr, V=Val, W=Trp, Y=Tyr. Enzymes and *Bacillus* strains are as follows; 1 : β-lactamase, 2 : α-amylase, *B. licheniformis;* 3 : *α-amylase,* 4 : *pro-protease (neutral), B. stearothermophilus;* 5 : β-lactamase I, 9 : pro-protease (neutral), 10 : α-amylase, *B. amyloliquefaciens;* 11 : pro-α-amylase, 12 : pro-subtilisin, 13 : pro-protease (neutral), 14 : β-glucanase, 15 : levansucrase, 16 : cellulase, *B. subtilis.*

a Shine-Dalgarno sequence (AAAGGAGG) which is complementary to the sequence of UUUCCUCC found close to the 3' terminus of the 16S rRNA of *Bacillus subtilis.* This sequence is in turn followed by an ATG fourteen nucleotides away, which is a little longer than the typical one but is in the range observed for other Shine-Dalgarno sequences. There is another sequence similar to the Shine-Dalgarno sequence at 62 base pairs from the ATG. If this is the true Shine-Dalgarno sequence, the GTG or the TTG (Wang *et. al.,* 1985) which is 6 and 11 nucleotide away respectively, might function as the initiation codon. Such a case seemed to be more spacially favorable than the former one. But the open reading frame was not formed beginning from either GTG or TTG.

적 요

고초균의 (1-4)-β-D-glucan-endoglucanase (cellulase) 유전자의 유전 조절부위에 대한 염기서열이 Sanger 의 사슬종결법으로 결정이 되었다. 이 염기서열에서 유추된 아미노산의 배열 N-말단에 여러 고초균속에서 분비되는 15개의 효소에 존재하는 분비신호와 아주 유사한 소수성 아미노산 잔기로 된 signal peptide가 존재하였다. 이 유전자의 개시 codon 앞에는 14 염기 상류에 16 S rRNA의 3' 말단과 상보성이 아주 큰 Shine-Dalgarno 서열이 존재하였으며 또한 promoter로 추정되는 −35 (TAGACA)와 −10 (TACAAT)이 개시 codon에서 상류 92 base 위치에 18 염기를 사이에 두고 존재하였다.

고초균의 cellulase signal peptide와 지금까지 알려진 15개의 다른 고초균의 signal peptide를 비교하여 본 결과 다른 생물에는 찾아볼 수 없는 homology region이 확인되었다.

REFERENCES

1. Bamforth, C.W. and H.L. Martin, 1981. β-Glucan and β-glucan solubilase in malting and mashing. *J. Inst. Brewing.* **87** 365-371.
2. Birnboim, H.C. and J. Doly, 1979. A rapid alkaline extraction procedure fpr screening recombinant plasmid DNA. *Nucleic Acids Res.* **7** 1513-1523.
3. Guerry,P., D.J. LeBlanc and S. Falkow, 1973. General method for the isolation of plasmid deoxyribonucleic acid. *J. Bacteriol.* **116**, 1064-1066.
4. Hussain, M., Carlino, A., M.J. Madonna and J.O.Lampen, 1985. Cloning and sequencing of the metallothioprotein β-lactamase II gene of *Bacillus cereus* 569/H in *Escherichia coli. J. Bacteriol.* **164**, 223-229.
5. Kushner, S.R. 1978. An improved method for transformation of *Escherichia coli* with ColEl-derived plasmids, in Boyer, H.B. and S. Nicosia(Eds.), *Genetic Engineering* Elsevier, Amsterdam, p17-23.
6. Lee, D.S. and M.Y. Park, 1985. Personal Communication.
7. Mezes, P.S.F., Yang, Y.Q., M. Hussain and J.O. Lampen, 1983. *Bacillus cereus* 569/H β-lactamas I: cloning in *Escherichia coli* and signal sequence determination. *FEBS Lett.* **161**, 195-200.
8. Moran Jr, C.P.,Lang, N., LeGrice, S.F.J, G. Lee and M. Stephens, 1982. Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis. Mol. Gen. Genet.* **186**, 339-346.
9. Murphy, N., D.J. McConnell and B.A. Cantwell, 1984. the DNA sequence of the gene and genetic control sites for the excreted *B. subtilis* enzyme β-glucanase. *Nucleic Acids Res.* **12**, 5355-5367.
10. Nakajima, R., T. Imanaka and S. Aiba, 1985. Nucleotide sequence of the *Bacillus stearothermophilus* α-amylase gene. *J. Bacteriol.* **163**, 401-406.
11. Neugebauer, K., R. Sprengel and H. Schaller, 1981. Penicillinase from *Bacillus licheniformis:* nucleotide sequence of the gene and implications for the biosynthesis of a secretory protein in a Gram-positive bacterium. *Nucleic Acids Res.* **11**, 2577-2588.
12. Ohmura, K., Yamazaki, H., Takeichi, Y., Nakayama, A., Otozai, K., Yamane, K., M.Yamasaki and G. Tamura, 1983. Nucleotide sequence of the promoter and NH_2 -terminal signal peptide region of *Bacillus subtilis* α-amylase gene cloned in pUB110. *Biochem. Biophys. Res. Commun.* **112**, 678-739.
13. Paddon, C.J. and R.W. Hartley, 1986. Cloning, sequencing, and transcription of an inactivated copy of *Bacillus amyloliquefaciens* extracellular ribonuclease (barnase). *Gene* **40**, 231-239.
14. Perlman, D. and D.O. Halvorson, 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptide. *J. Mol. Biol.* **167**, 391-409.
15. Sanger, F., S. Nicklen and A.R. Coulson, 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463-5467.
16. Seo, Y.S., Y.H. Lee and U.H. Pek, 1985. Molecular cloning of β-glucanase gene from *Bacillus subtilis* and its expression in *Escherichia coli. Korean Biochem. J.* **18**, 367-376.
17. Sloma, A. and M. Gross, 1983. Molecular cloning and nucleotide sequence of the type I β-lactamase from *Bacillus cereus. Nucleic Acids Res.* **11**, 4997-5004.
18. Stahl, M.L. and E. Ferrari, 1984. Replacement of of the *Bacillus subtilis* subtilisin structural gene with an *in vitro* -derived deletion mutation. *J. Bacteriol.* **158**, 411-418.
19. Stephens, M., Ortlepp, S.A., J.F. Oillington and D.J. McConnell, 1984. Nucleotide sequence of the 5'-region of the *Bacillus liqueniformis* α-amylase gene;

comparison with *B. amyloliquefaciens* gene. *J. Bacteriol.* **158**, 369-372.

20. Steinmetz, M., Coq, D.L., Atmerich, S., G. Gonzy-Treboul and P. Gay, 1985. The DNA sequence of the gene for the secreted *Bacillus subtilis* enzyme levansucrase and its genetic control sites. *Mol. Gen. Genet.* **200**, 220-228.
21. Takaki, M., T. Imanaka and S. Aiba, 1985. Nucleotide sequence and promoter region of the neutral protease gene from *Bacillus stearothermophilus*. *J. Bacteriol.* **163**, 824-831.
22. Takkinen, K., Pettersson, R.F., Kalkkinen, N., Palva, I., H.Soderlund and L. Kaariainen, 1983. Amino acid sequence of α-amylase from *Bacillus amyloliquefaciens* deduced from the nucleotide sequence of the cloned gene. *J. Biol. Chem.* **258**, 1007-1013.
23. Vasantha, N., Thompson, L.D., Rhodes, C., Banner, C., J. Nagle and D. Filpula, 1984. Genes for alkaline protease and neutral protease *Bacillus amyloliquefaciens* contain a large open reading frame between the regions coding for signal sequence and mature protein. *J. Bacteriol.* **159**, 811-819.
24. Von Heijne, G. 1983. Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.* **133**, 17-21.
25. Von Heijne, G. 1984. How signal sequences maintain cleavage specificity. *J. Mol. Biol.* **173**, 243-251.
26. Wang, W., Mezes, P.F.S., Yang, Y.Q., R. W. Balcher and J.O. Lampen, 1985. Cloning and sequencing of of β-lactamase I gene of *Bacillus cereus* 5/B and its expression *Bacillus subtilis*. *J. Bacteriol,* **163**, 487-492.
27. Watson M.E.E. 1984. Compilation of published signal sequences. *Nucleic Acids Res.* **12**, 5145-5163.
28. Wells, J.A., Ferrari. E., Henner, D.J., D. A. Estell and E.Y. Chen, 1983. Cloning, sequencing, and secretion of *Bacillus amyloliquefaciens* subtilisin in *Bacillus subtilis*. *Nucleic Acids Res.* **11**, 7911-7925.
29. Wong, S., Price C.W., D.S. Glodfarb and R.H. Doy, 1984. The subtilin E gene of *Bacillus subtilis* is transcribed from a σ^{37} promoter in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1184-1188.
30. Yang, M.Y., A. Galizzi and D. Henner, 1983. Nucleotide sequence of the amylase from *Bacillus subtilis*. *Nucleic Acids Res.* **11**, 237-249.
31. Yang, M.Y., E. Ferrari and D.J. Henner, 1984. Cloning of the neutral protease of *Bacillus subtilis* and the use of the cloned gene to create an *in vitro*-derived deletion mutation *J. Bacteriol.* **160**, 15-21.
32. Yanisch-Perron, C., J. Vieira and J. Messing, 1985. Improved M13 phage cloning vectors and host strains:nucleotide sequences of the M13mp18 and pUC vectors. *Gene* **39**, 103-119.

(Received Aug. 6, 1986)