



COMPARISON OF NUCLEOTIDE SEQUENCES OF ENDO-B-1,4-GLUCANASE GENES FROM *Bacillus subtilis* STRAINS

KIM YU-KYOUNG¹, OH HYUN-JEONG² AND KO YOUNG HWAN^{3*}

¹Division of Sustainable Agriculture Research, Jeju Agricultural Research and Extension Services, Jeju-do 697-828, Korea.

²Bio-Agr. Co. Ltd., Jeju-do 690-022, Korea.

³Department of Food Bioengineering, Jeju National University, Jeju-do 690-756, Korea.

*Corresponding Author: Email- yhkbl@cheju.ac.kr

Received: March 09, 2012; Accepted: April 12, 2012

Abstract- Three *Bacillus subtilis* strains were isolated from soil, compost and animal waste slurry in Jeju Island, Korea and their endo- β -1,4-glucanase genes cloned. Cellulase genes from *B. subtilis* SL9-9, C5-16 and S52-2 encoded proteins of 480, 470 and 499 amino acid residues, respectively. DNA sequences of the genes were compared with those of other known cellulase genes. The deduced amino acid sequences of the cellulase genes from the isolates matched quite well the modern concepts of multidomain cellulolytic enzymes. The enzymes were composed of three discrete domains: catalytic domains (CD) of glucosyl hydrolase family 5/A2 and interdomain linker and cellulose-binding domains (CBD) of family IIIa. Similar to the modular organization of many *Bacillus* endoglucanases, the CDs of these enzymes were located in the N-terminal region and CBDs in the C-terminal region.

Keywords- *Bacillus subtilis*, β -1,4-glucanase gene, nucleotide sequence, catalytic domain, cellulose-binding domain

Citation: Kim Yu-Kyoung, Oh Hyun-Jeong and Ko Young Hwan (2012) Comparison of Nucleotide Sequences of Endo-B-1,4-Glucanase Genes from *Bacillus subtilis* Strains. International Journal of Biotechnology Applications, ISSN: 0975-2943 & E-ISSN: 0975-9123, Volume 4, Issue 1, pp.-130-133.

Copyright: Copyright©2012 Kim Yu-Kyoung, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Introduction

Cellulose, β -1,4-glucan, is a major constituent of plant cell walls. Enzymatic hydrolysis of cellulosic materials is important for effective utilization of biomass. The complete enzymatic hydrolysis of cellulosic materials requires different types of cellulases: endo- β -1,4-glucanases (EC 3.2.1.4, endoglucanase, carboxymethyl cellulase, endocellulase), exo- β -1,4-glucanases (E.C. 3.2.1.91, cellobi-ohydrolase, avicelase, exocellulase) and β -glucosidases (EC 3.2.1.21, cellobiase) [4, see review]. These three enzyme groups work synergistically to promote the hydrolysis of crystalline cellulose [24, see review]. Endo- β -1,4-glucanases, endo-acting enzymes that hydrolyze 1,4- β -D-glucosidic linkages in cellulose, break down cellulose chains in a random fashion, thereby converting long chains to oligosaccharides. In contrast, exo- β -1,4-glucanases remove glucose or cellobiose units from the non-reducing end of the cellulose chain. Ultimately, β -glucosidases hydrolyze cellobiose to two glucose units and generally enhance cellulose degradation by relieving feedback inhibition of endo- and exo- β -1,4-glucanases by cellobiose.

Cellulolytic microbes have been investigated for many years [25]. A variety of bacterial species secrete cellulases, including strains of *B. subtilis* [11], *B. licheniformis* [2], *B. stearothermophilus* [9], *B. amyloliquefaciens* [12] and many other species [15, see review]. In addition to isolating cellulolytic microbes, reduction of enzyme production costs and enhancement of cellulase activity have become major areas of research [13].

The cellulase genes of many microorganisms have been cloned and the nucleotide sequences of endo- β -1,4-glucanase genes from several *Bacillus* strains have been determined [16, 18, 19, 21, 23]. The study of cellulolytic enzymes at the molecular level has revealed some of the modular features that contribute to their catalytic activity [22, see review]. Despite considerable diversity, sequence comparisons show that the catalytic cores of cellulases belong to a restricted number of families [1].

Here, we describe the properties and DNA gene sequences of endo- β -1,4-glucanases from *Bacillus subtilis* strains isolated from soil, compost and animal waste slurry in Jeju Island, Korea. The sequences of the genes were compared with those of other

known cellulase genes and their modular structures were defined.

Bacterial strains and media-

Hundreds of cellulolytic bacteria were screened and isolated from soil, compost and animal waste slurry. Among the isolates, three strains, SL9-9, C5-16 and S52-2, were chosen and identified as *Bacillus subtilis* [10] and deposited in the Korean Agricultural Culture Collection (KACC, Rural Development Administration, Korea). They were maintained and subcultured in agar medium containing carboxymethyl cellulose (CMC), 10.0; yeast extract, 1.0; (NH₄)₂SO₄, 2.5; K₂HPO₄.3H₂O, 0.25; NaCl, 0.1; MgSO₄.7H₂O, 0.125; FeSO₄.7H₂O, 0.0025; MnSO₄.4H₂O, 0.025; agar, 10 (g/L, each).

Cloning and DNA sequence analysis of β -1,4-glucanase gene-

The full-length β -1,4-glucanase genes from the three bacterial strains were cloned from the genomic DNAs of each bacterial strain by polymerase chain reaction (PCR) using primers based on the DNA sequences of previously cloned *Bacillus* cellulase genes. The primers used were BS cel up : 5'-ATG AAA CGG TCA ATC TCT ATT TTT-3' as a forward primer and BS cel dn : 5'-ACT AAT TTG GTT CTG TTC CCC AAA-3' as a reverse primer.

Molecular sizes of the resulting PCR products were analyzed on 1.0% agarose gel to confirm the 1.2 kb β -1,4-glucanase gene. These PCR products were purified with a DNA purification kit (QIAGEN, Valencia, California, USA), after which the purified PCR products were cloned into pGEM-T vector and sequenced. The nucleotide sequences of the cellulase genes were confirmed and compared through a BLAST nucleotide search provided by the National Center for Biotechnology Information (NCBI) Gene bank (U.S. National Library of Medicine, Bethesda, Maryland, USA).

Nucleotide and amino acid sequence analysis of β -1,4-glucanase gene-

The nucleotide sequences of the cloned genes were compared with those of the cellulase genes registered in the BLAST database using the Gen doc program (Table 1).

The *Bacillus subtilis* SL9-9 gene showed the highest identity with the *B. subtilis* strain C-36 endo- β -1,4-glucanase gene with 99% identity. The *Bacillus subtilis* C5-16 gene showed the highest identity with the *B. subtilis* strain AH18 cellulase gene with 99% identity. The *Bacillus subtilis* S52-2 gene showed the highest identity with the *B. subtilis* WRD-2 cellulase gene with 99% identity. The *Bacillus subtilis* SL9-9 gene showed 93% identity with the *Bacillus subtilis* C5-16 gene and 96% identity with the *Bacillus subtilis* S52-2 gene. Further, the C5-16 and S52-2 genes shared 93% identity. Based on these results, it is reasonable to conclude that all three of the cloned genes from our *Bacillus* strains code for cellulase protein (Table 1).

We compared the deduced amino acid sequences of the cellulases from the three *B. subtilis* isolates with those of *B. subtilis* PAP115 (GenBank accession number CAA28392) [16], *B. subtilis* DLG (GenBank accession number P07983) [23] and *B. subtilis* C-36 (GenBank accession number ABG78039) [5].

Amino acid sequences of cellulase proteins from different sources were aligned and compared using the COBALT (Constraint-based Multiple Alignment Tool) program provided by NCBI. The high degree of sequence similarity observed between the endoglucanases from the different *B. subtilis* strains revealed a close evolutionary relationship between the isolates (Fig. 1).

lutionary relationship between the isolates (Fig. 1).

Table 1- Homology among nucleotide sequences of cellulase genes from *Bacillus* strains

Bacterial isolate	Homology (%)	Source of cellulase gene	GenBank accession No.
<i>B. subtilis</i> SL9-9	99%	<i>B. subtilis</i> strain C-36 endo- β -1,4-glucanase	ABG78039
	93%	<i>B. subtilis</i> C5-16	JQ346089
	96%	<i>B. subtilis</i> S52-2	JQ346091
<i>B. subtilis</i> C5-16	99%	<i>B. subtilis</i> strain AH18 cellulase	ABK63475
	93%	<i>B. subtilis</i> SL9-9	JQ346090
	93%	<i>B. subtilis</i> S52-2	JQ346091
<i>B. subtilis</i> S52-2	99%	<i>Bacillus</i> . sp. WRD-2 cellulase	AAX54913
	96%	<i>B. subtilis</i> SL9-9	JQ346090
	93%	<i>B. subtilis</i> C5-16	JQ346089

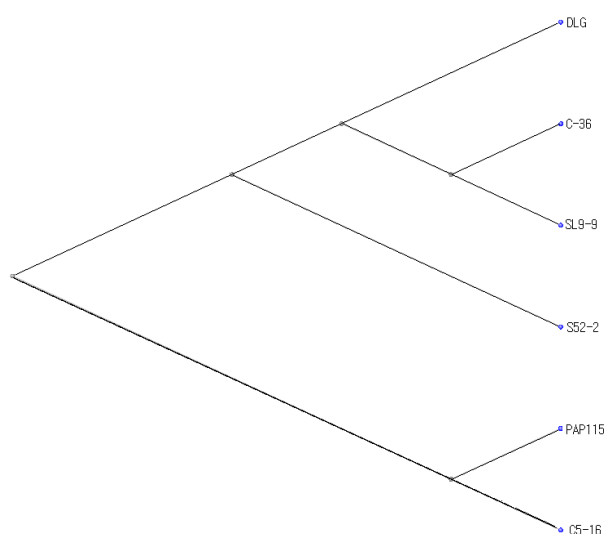


Fig. 1- Phylogenetic tree resulting from complete sequencing of cellulase genes.

According to the modern concept of cellulases [22, see review], most cellulolytic enzymes are comprised of modular multidomain proteins containing at least three separate structural elements with different functions, i.e., catalytic domain (CD), interdomain linker and cellulose binding domain (CBD). Currently, CDs of polysaccharases are grouped into at least 15 of the more than 80 known glycosyl hydrolase families, whereas CBDs fall into at least 13 families.

The cellulase genes identified in the nucleotide sequences of *B. subtilis* SL9-9, C5-16 and S52-2 encoded proteins of 480, 470 and 499 amino acid residues, respectively (Fig. 2). Computer analysis of the deduced amino acid sequences of the cellulases from our three isolates using the CDART program of NCBI revealed a modular enzyme composed of two discrete domains in the following order: CD (Q-48 through S-301) of the glucosyl hydrolase family 5/A2 (endoglucanase, EC 3.2.1.4) and CBD (V-356 through G-437) of family IIIa. Similar to the modular organization of many *Bacillus* endoglucanases, the CDs of these enzymes were located in the N-terminal region and CBDs in the C-terminal region.

When the homologous sequences were suitably aligned, a total of 45 amino acid substitutions among the six *Bacillus* strains were identified (Fig. 2).

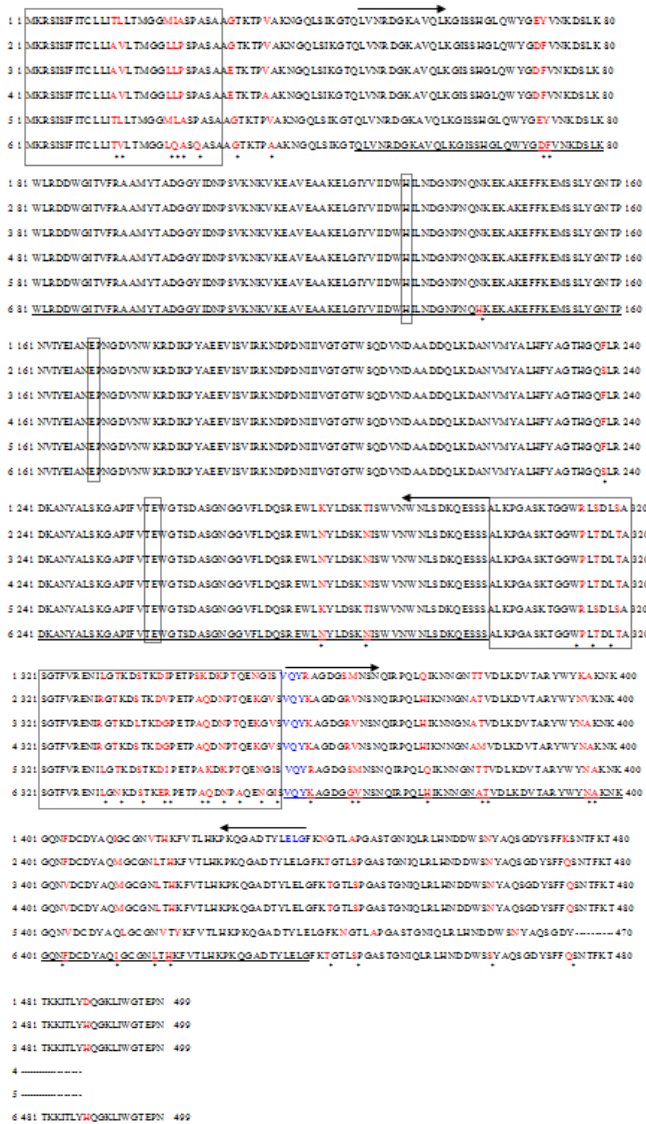


Fig. 2- Alignment of deduced amino acid sequences of endo-glucanases from the three isolates with those from other *Bacillus subtilis* strains.

- 1- *B. subtilis* PAP115 endo-β-1,4-glucanase gene,
 - 2- *B. subtilis* DLG endo-β-1,4-glucanase gene,
 - 3- *B. subtilis* C-36 endo-β-1,4-glucanase gene,
 - 4- *B. subtilis* SL9-9,
 - 5- *B. subtilis* C5-16,
 - 6- *B. subtilis* S52-2.
- Region 1-29, signal peptide; Region 48-301, cellulase catalytic domain (glucosyl hydrolase family 5/A2); Region 302-355, linker region; Region 356-437, cellulose binding domain (CBD family IIIa); Site 131-H, 169-E and 257-E, active sites (proton donor or nucleophile).

There were 5 to 6 amino acid residue substitutions in the signal peptide region (1 through 29) and 23 amino acid residues were conserved among all of the cellulases, resulting in 79% homology. The leader region (30 through 47) consisted of 18 residues and 16 amino acid residues among them were conserved among the

cellulases, resulting in 89% homology. The CD consisted of 254 residues (48 through 301) and 248 amino acid residues were conserved among the cellulases, resulting in 97.6% homology. In the linker region (302 through 355) consisting of 54 residues, 40 amino acid residues were conserved among the cellulases, resulting in 74% homology. In the CBD (356 through 437) consisting of 83 residues, 71 amino acid residues were conserved among the cellulases, resulting in 85.5% homology. The CD showed the highest 97.6% homology while the regions of signal, leader, linker and CBD regions showed low homology.

The endo-β-1,4-glucanase gene encoded a signal peptide at the N-terminal end of the protein. Two basic residues at positions 2 and 3 (lysine and arginine) in the hydrophilic leader region were followed by a hydrophobic core of 18 amino acid residues rich in leucine and isoleucine, which is in agreement with the findings of Mezes and Lampen [17]. They found that *Bacillus* signal peptides are characterized by a short, hydrophilic, basic region along with a subsequent, long, hydrophobic region. The N-terminal residues of the endoglucanase enzymes from the isolates indicate that the cleavage sites for signal peptide removal were between amino acids A-29 and A-30. This is in agreement with the findings of Mackay et al. [16] and Robson and Chambliss [23] regarding the enzymes from *B. subtilis* PAP115 and *B. subtilis* DLG, respectively.

Henrisatt et al. [8] classified fungal and bacterial cellulases into six families on the basis of hydrophobic cluster analysis. Specifically, amino acid residues His-131 and Glu-169 of endoglucanase were predicted to form the active site in the CD domain. When Park et al. [20] changed these amino acids by site-directed mutagenesis, a mutation at His-131 resulted in complete abolishment of enzyme activity, whereas a mutation at Glu-169 resulted in significant loss of enzyme activity. Robson and Chambliss [23] also reported that there are two active sites in the endo-β-1,4-glucanase from *B. subtilis* DLG. One is located at Glu-169 as a proton donor and the other at Glu-257 as a nucleophile. We confirmed that three of the amino acid residues, H-131(histidine), E-169(glutamic acid) and E-257(glutamic acid), were conserved at the same CD sites in our *Bacillus* species.

Hamamoto et al. [6] suggested that synergistic function of the NH2-terminus and COOH-terminus of the endoglucanase from *C. cellulovorans* is essential for hydrolysis of crystalline cellulose. The truncated endoglucanase from *B. subtilis* is 4-fold more active on CMC but 5-fold less active on insoluble cellulose upon C-terminal mutation [7]. In a previous study, when the CBD of cellobiohydrolase I from *T. reesei* was removed by protease, the activity towards Avicel was completely lost and decreased adsorption of the enzyme onto this microcrystalline cellulose was observed, whereas activity towards chromogenic oligosaccharides remained unchanged [26]. Park et al. [20] reported that the cellulose-binding ability of *B. subtilis* endoglucanase is independent of its catalytic activity to hydrolyze CMC, as a mutation in the CD active site does not affect its cellulose-binding ability. All of these reports strongly support that cellulases are composed of physically separate domains and that CBDs play crucial roles in the hydrolysis of crystalline cellulose by binding to the cellulose surface [14]. Usually, linkers comprise flexible disordered chains rich in proline and hydroxy amino acid residues (serine and threonine), as well as glycine and alanine [3]. Their lengths can vary from 5-6 to 100

residues, although most often limited to a range of 20 to 50 residues. They are believed to provide flexibility and spatial separation of CDs from CBDs, which allows their autonomous function on the surfaces of insoluble substrates [22]. In this study, there were many G(glycine), S(serine), T(threonine), K(lysine) and P (proline) residues in the linker region as shown in Fig. 2.

Our comparison of the deduced amino acid sequences of cellulase genes cloned from the three isolates, *B. subtilis* SL9-9, C5-16 and S52-2, found that although these enzymes were not extraordinary, they matched quite well with the modern concepts of multidomain cellulolytic enzymes.

References

- [1] Beguin P. and Aubert J.P. (1994) *FEMS Microbiol.* 13, 25-58.
- [2] Dhillon N., Chhibber S., Saxena M., Pajni S. and Vadehra D. (1985) *J. Biotechnol. Lett.* 7, 695-697.
- [3] Gilkes N.R., Henrissat B., Kilburn D.G., Miller Jr R.C. and Warren R.A. (1991) *Microbiol.* 55, 303-315.
- [4] Gilvert H.J. and Hazlewood G.P. (1993) *J. Gen. Microbiol.* 139, 187-194.
- [5] Guan X., Wu Z., Wu Q., Xu B. and Chen H. (2009) *Chin. J. Bioprocess Eng.* 3, 68-72.
- [6] Hamamoto T., Foong F., Shoseyov O. and Doi R.H. (1992) *Mol. Gen. Genet.* 231, 472-479.
- [7] Hefford M.A., Laderoute K., Willick G.E., Yaguchi M. and Seligy V.L. (1992) *Protein Eng.* 5, 433-439.
- [8] Henrissat B., Claeysens M., Tomme P., Lemesle L. and Moron J.P. (1989) *Gene.* 81, 83-95.
- [9] Kim S.W., Cho S.G. and Choi Y.J. (1997) *J. Microbiol. Biotechnol.* 7, 305-309.
- [10] Kim Y., Lee S., Cho Y., Oh H. and Ko Y.H. (2012) *ISRN Microbiology*, 2012, Article ID 650563, 9 pages. doi:10.5402/2012/650563.
- [11] Koide Y., Nakamura A., Uozumi T. and Beppu T. (1986) *Agric. Biol. Chem.* 50, 233-237.
- [12] Lee Y.J., Kim B.K., Lee B.H., Jo K. I., Lee N.K., Chung C.H., Lee Y.C. and Lee J.W. (2008) *Bioresour. Technol.* 99, 378-386.
- [13] Li X., Yang H., Roy B., Wang D., Yue W., Jiang L., Park E.Y. and Miao Y. (2009) *Afr. J. Biotechnol.* 8, 2418-2422.
- [14] Liu Y., Baker J.O., Zeng Y., Himmel M.E., Haas T. and Ding S. (2011) *J. Biol. Chem.* 286, 11195-11201.
- [15] Lynd L.R., Weimer P.J., van Zyl W.H. and Pretorius I. (2002) *Microbiol. Mol. Biol. Rev.* 66, 506-577.
- [16] Mackay R.M., Lo A., Willick G., Zuker M., Baird S., Dove M., Moranelli F. and Seligy V. (1986) *Nucleic Acids Res.* 14, 9159-9170.
- [17] Mezes P.S. F. and Lampen J.O. (1985) *The Molecular Biology of the Bacilli*, vol. II, 151-183.
- [18] Nakamura A., Uozumi T. and Beppu T. (1987) *Eur. J. Biochem.* 164, 317-320.
- [19] Ozaki K., Sumitomo N. and Ito S. (1991) *J. Gen. Microbiol.* 137, 2299-2305.
- [20] Park J.S., Nakamura A., Horinouchi S. and Beppu T. (1993) *Biosci. Biotechnol. Biochem.* 57, 260-264.
- [21] Park S.H., Kim H.K. and Pack M.Y. (1991) *Agric. Biol. Chem.* 55, 441-448.
- [22] Rabinovich M.L., Melnick M.S. and Bolobova A.V. (2002) *Biochemistry(Moscow)*, 67, 850-871.
- [23] Robson L.M. and Chambliss G.H. (1987) *J. Bacteriol.* 169, 2017-2025.
- [24] Robson L.M. and Chambliss G.H. (1989) *Enzyme Microb. Technol.* 11, 626-644.
- [25] Sukumaran R.K., Singhanian R.R. and Pandey A. (2005) *J. Sci. Ind. Res.* 64, 832-834.
- [26] Tilbeurgh V.H., Tomme P., Claeysens M., Bhikhabhai R. and Pettersson G. (1986) *FEBS Lett.* 204, 223-227.