Biotechnology and Bioprocess Engineering **2006**, 11: 168-172

Transcriptional Features of the Chromosomal Region Located between the *sigH* and *rplA* Genes of *Bacillus subtilis*

Sang Min Jeong[1,2]*

[1] Department of Physiology, College of Veterinary Medicine, Konkuk University, Seoul 143-701, Korea
[2] Bio/Molecular Informatics Center, Konkuk University, Seoul 143-701, Korea

Abstract In this study, the transcriptional features of a 2.8 kb region spanning the *sigH* and *rplA* genes of *Bacillus subtilis* were clarified using synthetic oligonucleotides complementary to the transcripts of the *rpmG*, *secE*, *nusG*, and *rplK* genes. The 5' ends of three transcripts corresponding to this region were located and mapped on the chromosome via primer extension analysis. Three regions, designated Prg, Pn, and Prk, which partially share the consensus sequence recognized by σ^A RNA polymerase, were theorized to function as promoter elements. The *rpmG* and *secE* genes of *B. subtilis* were cotranscribed from the designated Prg promoter, whereas the *nusG* and *rplK* genes were transcribed separately from the Pn and Prk promoters, respectively. Accordingly, the transcriptional features, as well as the gene organization, of the region encompassing the *sigH* and *rplA* genes of *B. subtilis*, including the *rpmG-secE-nusG-rplK* genes, were determined to be distinct from those of *Escherichia coli*. Divergences in terms of gene organization and transcriptional features within the relevant region would serve as excellent criteria for the delineation of phylogenetic relationships among bacteria.

Keywords: *Bacillus subtilis*, *rpmG*, *secE*, *nusG*, *rplK*, transcription

INTRODUCTION

Genetic and sequence analyses of the region encompassing the *sigH* to *rplA* genes of *Bacillus subtilis* revealed that the *sigH* gene is followed by the *rpmG*, *secE*, *nusG*, *rplK*, and *rplA* genes [1,2]. These genomic sequences were also reported in *B. subtilis* by the International Genome Project [3]. This identical genetic arrangement has also been observed in related *Bacillus* sp. [4]. The *secE* homologue gene of *B. subtilis* encodes for a short, 59 amino acid polypeptide. This polypeptide harbors one transmembrane-spanning domain that has been implicated as an essential component of the secretory machinery for protein transport in *B. subtilis* [1]. The chromosomal region harboring the *secE* homologue gene was isolated initially via the cloning of a DNA fragment that confers thiostrepton resistance (*rplK6* mutation as a genetic marker). The gene organization from *secE* to *rplA* is identical to that of *Escherichia coli*, whereas the *rpmG* counterpart of *E. coli* is located within another ribosomal gene cluster [5,6]. Furthermore, the *E. coli* genome does not naturally harbor a gene corresponding to the *sigH* gene, which encodes for one of the earliest sporulation-specific sigma factors [7]. The *secE* and *nusG* genes in *E. coli* have been reported to be cotranscribed, and that the only one nucleotide of intergenic space separates the two genes [9,10].

The strain used in the preparation of plasmid and single strand DNA in this study was the following: *E. coli* INVαF' [*end*A 1, *rec*A 1, *hsd*R17(r_K^-, m_K^+), *sup*E44, λ^-, *thi*-1, *gyr*A, *rel*A 1, *Φ80lacZΔM15(lac*ZYA-*arg*F)]. *B. subtilis* UOT-1285(*trp*C2, *lys*-1, *apr*A3 *npr*R2, and *npr*E18) was employed in RNA extraction. For the preparation of ladder DNA, 1102 base pairs of *Eco*RI-*Eco*RV fragments, including *rpmG-secE*, were cloned into pTZ19R (TOYOBO). pTE814R [1] was used to determine the transcription start point of the *rplK* gene.

Oligodeoxyribonucleotides were synthesized with an ABI 392 DNA/RNA synthesizer (Applied Biosystems, Foster City, CA, USA). The complementary oligonucleotides for primer extension for the *rpmG*, *secE*, *nusG*, and *rplK* genes are listed in Table 1. The given nucleotide numbers correspond to those shown in Fig. 1 in the Jeong *et al.* [1]. Oligonucleotides (10 pmol) were 5' end labeled with T4 polynucleotide kinase and [γ-^{32}P] ATP [11].

The oligonucleotides (10 pmol) were 5' end labeled at 37°C for 60 min with 10 units of T4 polynucleotide kinase and 30 μCi of [γ-^{32}P] ATP in 1 μL of kinase buffer (0.1 M Tris-HCl (pH 8.0), 5 mM dithiothreitol, 10 mM $MgCl_2$). The reaction was arrested via the addition of 1 μL of 0.5 M EDTA (pH 8.0) and 50 μL of DEPC-treated distilled water, and then extracted once using phenol-chloroform. The labeled oligonucleotide was purified with ethanol precipitations in the presence of 3 M sodium acetate and 5 μg of yeast tRNA.

B. subtilis preculture, which was grown in 2XSG me-

***Corresponding author**
Tel: +82-2-450-3716 Fax: +82-2-450-3037
e-mail: jsmyh1981@yahoo.co.kr

Table 1. Primer sequence for determination of start point

Primer name	Sequence	Position
RPMR	5′-TCCCGCATGTCTTGCATGCTAACGTAATC-3	423 to 450
BSEPEX	5′-ATTTCTTTCCCAACATCTTTAAAGAATTTC-3′	606 to 635
NUSG2	5′-ATCCAGCATCCCCATTGATTCAACACG-3′	1025 to 1051
RPLK1	5′-CCAGCAGGAATTTGCAATTTTACAAC-3′	1675 to 1700
RPLK2	5′-GTGATGTGGTAATAGGGTAAACCCTCCCAC-3′	1603 to 1632

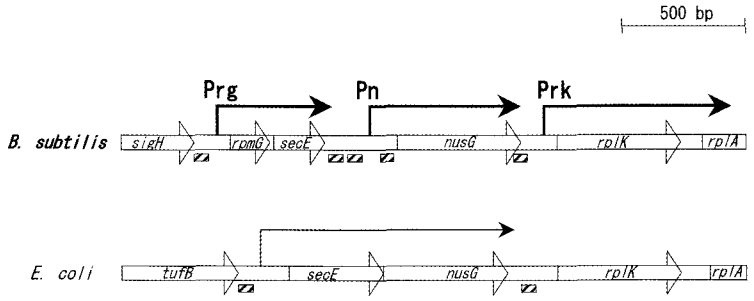

Fig. 1. Physical map and alignment of transcription units in the *secE* region. The bold solid lines indicate the transcriptional units described in this report. The solid line for *E. coli* shows the transcriptional start point determined by transcript mapping in a previous report. σ^A consensus promoters are designated for each of the genes, as follows: Prg; *rpmG-secE*, Pn; *nusG*, Prk; *rplK-rplA*. The cross-hatched bars below the map show the inverted repeat symmetry associated with termination.

dium containing 0.1% glucose at 37°C, was diluted using 15 volumes of the same fresh medium, when the optical density of the culture had reached an OD_{660}=0.5 units. During cultivation, 50 mL of the culture was collected as a vegetative cell fraction in the mid-logarithmic phase, and 14 mL of the culture was withdrawn at several time points after the end of the logarithmic growth phase. Total cellular RNA was extracted in accordance with the methods developed by Kirby *et al.*, with some modifications [12].

The 5′ end transcripts were analyzed via primer extension and Northern blotting [11]. Eighty μg of total cellular RNA and 1 pmol of 5′ end labeled oligonucleotide primer mixture was precipitated, washed with 70% ethanol, and dried. The mixture was subsequently completely suspended with 24 μL of formamide, after which 6 μL of hybridization buffer (200 mM PIPES (pH 6.7), 2 M NaCl, 5 mM EDTA) was added, and the suspension was heated for 10 min at 85°C. The hybridized mixture was maintained at 30°C overnight, then precipitated, washed with 70% ethanol, and air-dried. The ^{32}P end labeled primer and RNA were resuspended in 20 μL of RT buffer (120 mM Tris-HCl (pH 8.3), 96 mM KCl. 19.2 mM $MgCl_2$, and 2.4 mM DTT). Five units of reverse transcriptase (Takara Shuzo Co., Japan) containing RNase inhibitor (Pharmacia) 10 mM β-mercaptoethanol, and 1 mM of each deoxyribonucleotide, were added. Incubation then continued for an additional 2 h at 37°C. The reaction was arrested via the addition of 1 μL of 0.5 M EDTA, after which 5 μg of RNaseA was added, followed by 30 min of incubation at 37°C. The products were then extracted using phenol-chloroform, and precipitated with ethanol in the presence of 0.3 M sodium acetate and 5 μg of yeast tRNA. The pellet was dissolved in 10 μL formamide sequencing dye mix, and radioactivity was measured via Cerenkov counting. The reaction products were analyzed on 6% polyacrylamide urea sequencing gel, alongside a sequencing ladder which had been generated using an appropriate single-stranded template, coupled with the same primer as used in the primer extension step.

The single-stranded DNA generated from the appropriate plasmids was then utilized for sequencing via dideoxy chain termination [11].

A 2,782-nucleotide-long region corresponding to the carboxyl terminus of the SigH (σ^H) protein and the amino-terminus of the RplA protein are shown in Fig. 1. The 5′ end of the transcripts within this region was analyzed via primer extension analysis, using five oligonucleotides; RPMR, BSEPEX, NUSG2, RPLK2, and RPLK1 (see Table 1). The three 5′ end sites and *rpmG-secE* transcripts were identified via primer extension and Northern analysis (Figs. 2 and 3). These transcription initiation sites were shown to be located at nucleotide positions 373, 922, and 1533. The latter two positions were also verified via primer extension analysis, using different oligonucleotide primers (data not shown). These sites were shown to be preceded by promoter-like sequences recognizable by σ^A containing RNA polymerase, and were therefore designated as the Prg, Pn, and Prk promoters, respectively (Fig. 1). The putative −35 and −10 sequences of these promoters are as follows:

Prg; TTGACA--18bp--TATGTT
Pn; TTGAAG--16bp--TAAAAC
Prk; TTGAAA--17bp--TAATAT

Two detected minor signals (Fig. 3B) were suggested

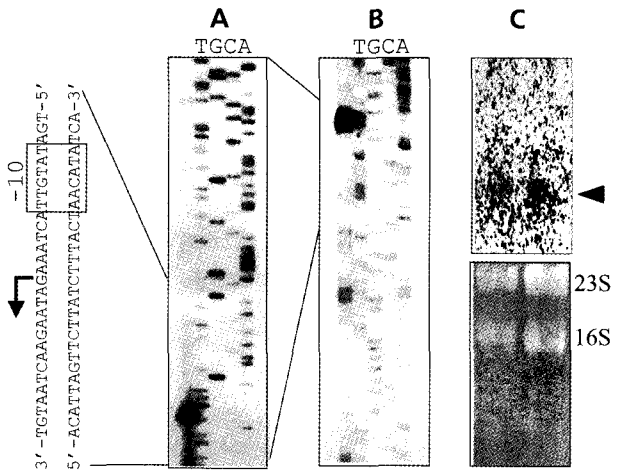

Fig. 2. Determination of *rpmG-secE* transcripts in *B. subtilis*. The primer complementary to the 5′ end of *rpmG* (panel A) and *secE* (panel B) were respectively hybridized to the total RNA (80 μg) isolated from strain UOT-1285. Lanes A, C, G, and T, all represent products of sequencing reactions obtained with the same primer. The double strand −10 sequence of the promoter is also presented. The arrows at the G base indicate the start (+1) of the mRNA of the *rpmG-secE* gene. The designated RNA transcripts are shown in panel C. The *rpmG-secE* transcript was hybridized using RPMR primers. About 500 bps of *rpmG-secE* gene transcript was indicated on the upper side, with a black arrowhead. 23S and 16S ribosomal RNA, as controls, are shown below. Lane 1; 10 μg of total RNA, lane 2; 20 μg of total RNA.

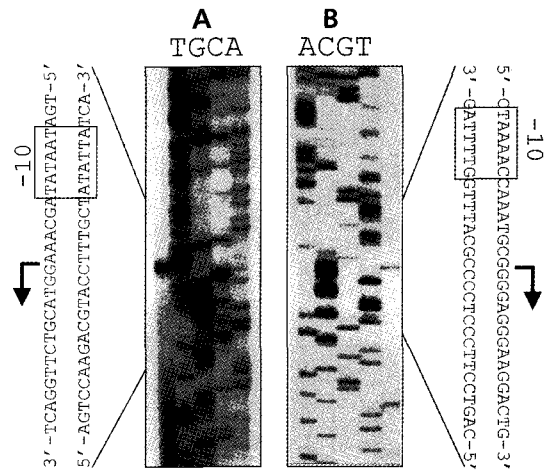

Fig. 3. Mapping of the 5′ end of mRNA for *nusG* and *rplK* via primer extension. The primer complementary to the 5′-end of *rplK* (panel A) and *musG* (panel B) were respectively hybridized to the total cellular RNA isolated from strain UOT-1285. Lanes A, C, G, and T, products of sequencing reactions obtained with the same primer. The double strand −10 sequence of the promoter is shown. The arrows at the G base indicate the start (+1) of the mRNA of each of the genes.

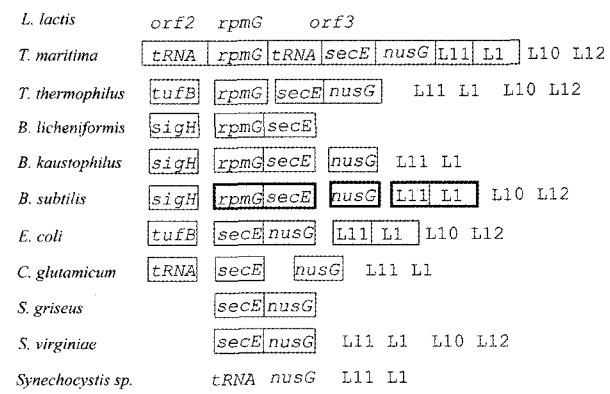

Fig. 4. Comparison of the genetic map of *rpmG-secE* and ribosomal protein gene region in a variety of microorganisms. The sporulation-specific sigma factor gene *sigH* in *Bacilli* [7] are located in the *tufB* region of *E. coli* [1]. *B. subtilis rpmG-secE* was cotranscribed (as shown in this report), in comparison with the *secE-nusG* operon of *E. coli* [8]. The genetic maps of another bacterial species were constructed from the GenBank database. The ribosomal protein genes are depicted for convenience in L11 (*rplK*), L1 (*rplA*), L10 (*rplJ*) and L12 (*rplL*), respectively. The thick boxes show the transcription units revealed by the determination of transcripts described in this paper.
L; *Lactococcus* [19], T; *Thermus* [18] or *Thermotoga* [20], B; *Bacillus* [1,4], E; *Escherichial* [8,26], C; *Corynebacterium* [21], S; *Streptomyces*. [22-24], *Synechocystis* sp. [25].

to be intermediates of the reverse transcripts, as no promoter-like sequence was detected within the upstream region. An inverted repeat sequence upstream of the Pn promoter, CTGGACAGTCCTG, may affect the expression of *nusG*, as well as its function as a transcriptional termination signal. The mechanisms underlying the regulation of the expression of the *nusG* gene, the product of which is a factor of a transcriptional antitermination system, remains an interesting target for adjunct research [9,10,13,14].

The 3′ terminal region of the transcript from Prg promoter is capable of forming a stem loop structure (ΔG = −18.3 Kcal/mL) [15] followed by a tract of T residues, which may function as a transcription termination signal. As the intergenic space between the *rpmG* and *secE* genes is only 33 nucleotides long, and because no transcript starting from this region has been detected, it was concluded that these two genes had been cotranscribed from the Prg promoter. A transcript of mRNA, approximately 500 nucleotides in length and relevant to the *rpmG-secE* gene, was detected via Northern analysis (Fig. 2C). The *E. coli rpmG* gene is located in a different region (at 82 min) of the chromosome, and was found to have been cotranscribed with the ribosomal protein gene, *rpmB* [5,6]. The *secE* and *nusG* genes of *E. coli* are similarly cotranscribed [8].

It is difficult to understand why the *rpmG* and *secE* genes are cotranscribed in *B. subtilis*. The *secY* gene, the product of which constitutes an important part of the protein export machinery of *E. coli*, is cotranscribed as a

component of a ribosomal protein operon (*spc* operon) [16]. Some *secY* mutants evidence reduced levels of the ribosomal protein, L15 (RplO) [17]. This is suggestive of the notion that the product of the *secY* gene interacts directly with the ribosome. There is a reasonable curiosity regarding the relationship between protein synthesis and export in this case. If the product of the *secE* gene does directly interact with ribosomal protein, *rpmG-secE* might be involved with the same single transcription unit. The organizational aspects of the genes within the relevant region have been previously examined in several bacterial species. Interestingly, there appear to be substantial differences in gene arrangement between the *rpmG*, and *secE-nusG-rplK-rplA* gene clusters (Fig. 4). Spore-forming *Bacillus* strains tend to harbor a *sigH* gene in place of the *tufB* gene observed in *E. coli*. Although *Thermus thermophilus* possesses a *tufB* homologue gene in the place of the *sigH* gene [14,18], the *rpmG* gene is clustered with *secE-nusG* genes as *B. subtilis* does. Thus, the genetic organizational, and probably the transcriptional, features of the relevant regions of different bacteria are evolutionarily divergent, and not straightforward.

Acknowledgments I would like to express my gratitude to Professors Hideo Takahashi and Hirofumi Yoshikawa. I would also like to thank Dr. K. Tanaka for technical advice, and Y. Ohashi for the generous technical assistance. This research was supported by a grant-in-aid from the Ministry of Education, Science, and Culture of Japan and the Bio/Molecular Informatics Center of Konkuk University (KRF2004-F00019).

REFERENCES

[1] Jeong, S. M., H. Yoshikawa, and H. Takahashi (1993) Isolation and characterization of the *secE* homologue gene of *Bacillus subtilis*. *Mol. Microbiol.* 10: 133-142.

[2] Yasumoto, K., H. Liu, S. M. Jeong, Y. Ohashi, S. Kakinuma, K. Tanaka, F. Kawamura, H. Yoshikawa, and H. Takahashi (1996) Sequence analysis of a 50 kb region between *spo0H* and *rrnH* on the *Bacillus subtilis* chromosome. *Microbiology* 142: 3039-3046.

[3] Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, and S. Borchert, *et al.* (1997) The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* 390: 249-256.

[4] Sharp, P. M. (1994) Identification of genes encoding ribosomal protein L33 from *Bacillus licheniformis*, *Thermus thermophilus*, and *Thermotoga maritime*. *Gene* 139: 135-136.

[5] Isono, K., J. Schnier, and M. Kitakawa (1980) Genetic fine structure of the *pyrE* region containing the gencs for ribosomal proteins L28 and L33 in *Escherichia coli*. *Mol. Gen. Genet.* 179: 311-317.

[6] Lee, J. S., G. An, J. D. Friesen, and K. Isono (1981) Cloning and the nucleotide sequence of the genes for *Escherichia coli* ribosomal proteins L28 (*rpmB*) and L33 (*rpmG*). *Mol. Gen. Genet.* 184: 218-223.

[7] Dubnau, E., J. Weir, G. Nair, L. Carter, 3rd, C. Moran, Jr., and I. Smith (1988) *Bacillus* sporulation gene *spoOH* codes for σ^{30} (σ^{H}). *J. Bacteriol.* 170: 1054-1062.

[8] Downing, W. L., S. L. Sullivan, M. E. Gottesman, and P. P. Dennis (1990) Sequence and transcriptional pattern of the essential *Escherichia coli secE-nusG* operon. *J. Bacteriol.* 172: 1621-1627.

[9] Mason, S. W. and J. Greenblatt (1991) Assembly of transcription elongation complexes containing the N protein of phage 1 and the *Escherichia coli* elongation factors NusA, NusB, NusG, and S10. *Genes Dev.* 5: 1504-1512.

[10] Sullivan, S. L. and M. E. Gottesman (1992) Requirement for *E. coli* NusG protein in factor- dependent transcription termination. *Cell* 68: 989-994.

[11] Sambrook, J., E. F. Fritsch, and T. Maniatis (1989) *Molecular Cloning: A Laboratory Manual.* 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.

[12] Kirby, K. S., E. Fox-Carter, and M. Guest (1967) Isolation of deoxyribonucleic acid and ribosomal ribonucleic acid from bacteria. *Biochem. J.* 104: 258-262.

[13] Greenblatt, J., J. R. Nodwell, and S. W. Mason (1993) Transcriptional antitermination. *Nature* 364: 401-406.

[14] Heinrich, T., W. Schroder, V. A. Erdmann, and R. K. Hartmann (1992) Identification of the encoding transcription factor *NusG* of *Thermus thermophilus*. *J. Bacteriol.* 999: 7859-7863.

[15] Kyte, J. and R. F. Doolittle (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157: 105-132.

[16] Cerretti, D. P., D. Dean, G. R. Davis, D. M. Bedwell, and M. Nomura (1983) The *spc* ribosomal protein operon of *Escherichia coli*: sequence and cotranscripton of the ribosomal protein genes and a protein export gene. *Nucleic Acids Res.* 11: 2599-2616.

[17] Ito, K., M. Wittekind, M. Nomura, K. Shiba, T. Yura, A. Miura, and H. A Nashimoto (1983) Temperature-sensitive mutant of *E. coli* exhibiting slow processing of exported proteins. *Cell* 32: 789-797.

[18] Satoh, M., T. Tanaka, A. Kushiro, T. Hakoshima, and K. Tomita (1991) Molecular cloning, nucleotide sequence and expression of the *tufB* gene encoding elongation factor Tu from *Thermus thermophilus* HB8. *FEBS Lett.* 288: 98-100.

[19] Koivula, T. and H. Hemila (1992) Sequence encoding ribosomal protein L33 of *Lactococcus lactis*. *Gene* 119: 145-146.

[20] Liao, D. and P. P. Dennis (1992) The Organization and expression of essential transcription translation component genes in the extremely thermophilic eubacterium *Thermotoga maritime*. *J. Biol. Chem.* 267: 22787-22797.

[21] Barreiro, C., E. Gonzaález-Lavado, and J. F. Martín (2001) Organization and transcriptional analysis of a six-gene cluster around the *rplK-rplA* operon of *Corynebacterium glutamicum* encoding the ribosomal proteins L11 and L1. *Appl. Environ. Microbiol.* 67: 2183-2190.

[22] Miyake, K., H. Onaka, S. Horinouchi, and T. Beppu (1994) Organization and nucleotide sequence of the *secE-nusG* region of *Streptomyces griseus*. *Biochim. Biophys. Acta* 1217: 97-100.

[23] Katayama, M., Y. Sakai, S. Okamoto, F. Ihara, T. Nihira, and Y. Yamada (1996) Gene organization in the *ada-rplL*

region of *Streptomyces virginiae*. *Gene* 171: 135-136.
[24] Puttikhunt, C., T. Nihira, and Y. Yamada (1995) Cloning, nucleotide sequence, and transcriptional analysis of the *nusG* gene of *Streptomyces coelicolor* A3(2), which encodes a putative transcriptional antiterminator. *Mol. Gen. Genet*. 247: 118-122.
[25] Schmidt, J. and A. R. Subramanian (1993) Sequence of the cyanobacterial tRNA gene in *Synechocystis* PCC 6803: requirement of enzymatic 3′ CCA attachment to the acceptor stem. *Nucleic Acids Res*. 21: 2519.
[26] Chow, J. and P. P. Dennis (1994) Coupling between mRNA synthesis and mRNA stability in *Escherichia coli*. *Mol. Microbiol*. 11: 919-931.

[Received January 5, 2006; accepted April 4, 2006]