Determination and Characterization of IS4Bsu1-Insertion Loci and Identification of a New Insertion Sequence Element of the IS256 Family in a Natto Starter

Keitarou Kimura and Yoshifumi Itoh

Division of Applied Microbiology, National Food Research Institute, 2-1-12 Kannondai, Tsukuba, Ibaraki 305-8642, Japan

Received April 17, 2007; Accepted June 19, 2007; Online Publication, October 7, 2007

The insertion sequence IS4Bsu1 frequently causes Bacillus subtilis starters for the production of Japanese fermented soybean pasts (natto) to lose the ability to produce poly-γ-glutamate, the viscous material characteristic of natto. Bacillus subtilis NAFM5, a derivative of a natto starter, has six IS4Bsu1 copies on its chromosome. In this study, we determined all six insertion loci of the insertion sequence (IS). One was located in the coding region of ykID, a putative gene involved in polyketide synthesis. Four were located in non-coding regions between iolR and iolA, between tuaA and lytC, between rap1 and orf1 (a potential gene of unknown function), and between ynaE and orf3 (a putative gene similar to thiF), and one resided in an intergenic region between divergent possible orf4 and orf5 genes of unknown function. Here we describe the structural features of these loci and discuss the effects of the IS4Bsu1 insertions on the functions of the target gene and the expression of the downstream genes. In addition, we found that strain NAFM5 and commercial natto starters possess eight to 10 loci of another IS of the IS256 family (designated IS256Bsu1) on their chromosomes. IS256Bsu1 appeared active in transposition, potentially causing phenotypic alterations in natto starters like those induced by IS4Bsu1.

Key words: soybean fermentation; poly-γ-glutamate production; integrative and conjugative element; polyketide synthesis; gene transfer

IS4Bsu1 is an insertion sequence of the IS4 family that was originally discovered in a Bacillus subtilis natto starter.1,2 This mobile DNA element frequently and preferentially transposes into comP, a member of the comQXPA quorum-sensing operon, which controls the expression of diverse stationary-phase specific phenotypes, including genetic competence, flagellation, degradative enzyme synthesis, and the production and degradation of poly-γ-glutamate (PGA).1–7 Since PGA confers characteristic viscous and sticky properties on natto products, the ability to produce this polypeptide is critical for natto starters. Several IS4Bsu1 sites are numerous among commercial natto starters and B. subtilis strains isolated from Asian fermented soybeans.2,8 Commercial natto starters have at least six copies of IS4Bsu1, which appear to be located on a common locus on chromosomes, thus providing genetic evidence that commercial starters share the same origin.2 Some starters possess additional copies that have transposed onto new sites.2,5 Since transposition of the IS into a new site can inactivate a receiver gene, causing phenotypic alterations, this mobile element is used as a molecular marker to ensure the genetic integrity of natto starters.9 One IS4Bsu1 locus was recently identified in an intergenic region between iolR and iolA, the former being the regulatory and the latter the first gene of a myo-inositol utilization operon (iolABCDEFGHIIJ).10 This insertion appears to influence the expression of the myo-inositol utilization operon and the regulatory gene minimally. Because none of the common IS4Bsu1 loci has been determined, whether this is a common or a novel IS4Bsu1 locus in the analyzed strain is unknown. The IS4Bsu1 insertions in common loci might cause a defect in a metabolic or cellular process of natto starters, and might provide a clue to the IS4Bsu1 sequence that was initially introduced into an ancestor of the starter strains and the mechanism that delivers this type of IS element among B. subtilis strains.

In this study, we determined the locations of all six IS4Bsu1 sequences on the chromosome of B. subtilis NAFM5, a derivative of a commercial starter. One of the
IS4Bsu1 loci was apparently included in the integrative and conjugative element (ICE) Bsu1, and another was located in a DNA region unique to the natto starter, implying that both were initially introduced into the original of the starters. Furthermore, we found eight to 10 loci of another IS element of the IS256 family on the chromosomes of strain NAFM5 and commercial starters. This second IS appeared to be capable of active transposition, and hence might alter the properties of natto starters such as IS4Bsu1.

Materials and Methods

Strains and media. Bacillus subtilis NAFM5 (rif\textsuperscript{R}, bio) was derived from the commercial starter Miura (Natto Manufacturer Miyagino, Sendai, Japan).\textsuperscript{2) Commercial starters Naruse and Takahashi were obtained from the Naruse Fermentation Institute (Tokyo, Japan) and the Laboratory of Yuzo Takahashi (Yamagata, Japan) respectively. B. subtilis and Escherichia coli DH5\textalpha (Bethesda Research Laboratory, Bethesda, MD) cells were cultured at 37°C in Luria-Bertani (LB) medium.\textsuperscript{13) The minimal medium in which B. subtilis strains were grown contained 0.1 m MOPS (pH 7.0), 4.2 mM K\textsubscript{2}HPO\textsubscript{4}, 4.2 mM KH\textsubscript{2}PO\textsubscript{4}, 8.8 mM (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4}, 2 mM MgCl\textsubscript{2}, 0.7 mM CaCl\textsubscript{2}, 50 mM MnCl\textsubscript{2}, 1 mM ZnCl\textsubscript{2}, 5 mM FeCl\textsubscript{3}, 2 mM thiamine hydrochloride, biotin (0.5 μg/ml), and 1% (w/v) of glucose or myo-inositol.

Isolation of IS4Bsu1-flanking regions. Chromosomal DNA of B. subtilis NAFM5 was digested with HaeIII, which has no cutting site on IS4Bsu1, by Southern blotting using an IS4Bsu1 DNA probe for the IS element (see below). Fragments of DNA in agarose gels that corresponded to the hybridized bands were eluted using a DNA extraction kit (GE Healthcare Bioscience, Little Chalfont, UK) and circularized using a DNA ligation kit (Nippon Gene, Tokyo, Japan). The flanking DNA regions of the IS elements were then amplified by PCR using the circularized DNA molecules as templates and oligonucleotide primers 5’-GATGTT-ACTACCTCTATTTAAGTTC-3’ (complementary to nt 1,311–1,337 of IS4Bsu1) and 5’-GTGTTAAACTTATGCTTGC-3’ (complementary to nt 126–152 of IS4Bsu1), and KOD-plus DNA polymerase (Toyobo Biochemicals, Osaka, Japan).

Construction of genomic DNA library. A library of NAFM5 chromosomal DNA fragments was constructed as described previously.\textsuperscript{15) Briefly, NAFM5 chromosomal DNA was partially digested with SauIII and resolved by agarose gel electrophoresis. Fragments of about 20 kb were eluted from the gel as described above, and ligated with BamHI-cleaved EMBL3 (Stratagene, La Jolla, CA). The ligated DNA was packaged in vitro using Gigapack II XL Packaging Extract (Stratagene), followed by screening for λ clones harboring IS4Bsu1 by plaque hybridization using IS4Bsu1 DNA as the probe (see below).

Southern blotting and plaque-hybridization. Chromosomal DNA was digested with an appropriate restriction enzyme, resolved by electrophoresis on 1% (w/v) agarose gel, and transferred onto Hybond N+ membranes (GE Healthcare Bioscience) using a VacGene blotter (GE Healthcare Bioscience) as described previously.\textsuperscript{2) The nt 137 to 1,301 region of IS4Bsu1 and the nt 155 to 1,165 region of IS256Bsu1 were amplified by PCR using NAFM5 chromosomal DNA as the template and oligonucleotide primers: the primer set used to amplify IS4Bsu1 DNA was as described above and those for amplification of the IS256Bsu1 region were 5’-TAACAGATTACGAGGCAGATATCGAGCC-3’ (nt 155 to 171 of IS256Bsu1) and 5’-AGCTAGTGGATGCACTTCAACGC-3’ (complementary to nt +1,139 to 1,165 of IS256Bsu1). Fluorescein-labeled DNA probes were prepared according to the supplier’s instructions using the amplified DNA fragments as templates and ECL random labeling kits (GE Healthcare Bioscience), then hybridized with membrane DNA fragments or with λ clone DNA fragments containing IS4Bsu1. Hybridized DNA was visualized on X-ray films (Fuji Film, Tokyo, Japan) using ECL detection kits (GE Healthcare Bioscience).

DNA sequencing and analysis. Nucleotides of the IS4Bsu1 flanking regions and of IS256Bsu1 were determined on both strands using an ABI310 DNA sequencer (PE Applied Biosystems, Foster City, CA) and cycle sequence reaction kits (PE Applied Biosystems). We compared homology against the NCBI protein sequence database using the BlastP program (http://www.ncbi.nlm.nih.gov/), and amino acid sequences were aligned using the Clustal W program.\textsuperscript{14) The nucleotide sequences determined in this study have been deposited in the GenBank/EMBL/DDJB databases (IS256Bsu1, accession no. AB275894; IS4Bsu1 loci, nos. AB304460 to AB304464).

Results and Discussion

Isolation and sequencing of the flanking regions of IS4Bsu1 insertion sites

We isolated the DNA sequences flanking the IS4Bsu1 copies on the B. subtilis NAFM5 chromosome using two approaches. In the first, we digested DNA with HaeIII that does not excise the IS4Bsu1 sequence and resolved the resulting HaeIII fragments by agarose gel electrophoresis. Southern blotting using the IS4Bsu1 DNA probe identified six positive DNA fragments, labeled A through F (Fig. 1), confirming that NAFM5 has six IS4Bsu1 copies on the chromosome, as does its parent Miyagino strain.\textsuperscript{2) We then extracted the DNA fragments from the agarose gels that corresponded to the hybridized DNA bands and circularized them using T4 DNA ligase.
PhrI is its cognate inhibitor.5,16,17) The C-terminal phosphatase of the aspartate-phosphatase family and that encodes the 41-amino acid RapI inhibitor. RapI is a strain 168 counterparts respectively. 15) The shared 96% and 66% amino acid identity with their phrI and rapI-4Bsu hybridization using the ISa Hae III fragments A, D, E, and F. We sequenced the flanking nucleotides by PCR using circularized DNA fragments as templates and ligase. We attempted to amplify the flanking sequences by PCR using circularized DNA fragments as templates and oligonucleotide primers with sequences corresponding to the distal regions (nt 126 to 1,337) of IS4Bsu1 (1,406 bp). Thus we isolated the DNA regions contiguous to the IS elements in HaeIII fragments A, D, E, and F.

In the second approach, we constructed and screened a λ phage library of the NAFM5 chromosome by plaque hybridization using the IS4Bsu1 probe, which resulted in five positive λ clones that carried HaeIII fragments A, B, C, D, or E. We sequenced the flanking nucleotides using PCR-amplified DNAs or recombinant λ DNAs as templates and the oligonucleotide primers described above.

The rapI and orf1 locus (fragment A)

The insertion site of IS4Bsu1 on HaeIII fragment A was determined in a non-coding region between rapI-phrI and orf1 (Fig. 2). RapI and PhrI of strain NAFM5 shared 96% and 66% amino acid identity with their strain 168 counterparts respectively.15) The rapI-coding sequence overlaps at the 3' end with the phrI sequence that encodes the 41-amino acid Rap inhibitor. RapI is a phosphatase of the aspartate-phosphatase family and PhrI is its cognate inhibitor.5,16,17) The C-terminal peptapeptide (DRVGA), which is responsible for inhibition of Rap phosphatase activity,18) was perfectly conserved between the NAFM5 and 168 PhrI peptides.

Rap-Phr systems in B. subtilis modulate phosphorylation levels and binding to the operator sites of the response regulators of two-component regulatory proteins in the signal transduction pathways that modulate genetic competence, spore formation, degradative enzyme synthesis, and motility.5,6,16,17) B. subtilis 168 possesses 11 Rap proteins (A through K). RapA, RapB, and RapE dephosphorylate the Spo0F-P response regulatory intermediate of the sporulation phosphorylation (5,17,19) while RapC, RapD, and RapF prevent binding of the ComA-P competence response regulator to its regulatory site.5,20–22) RapG and RapH interfere with ComA binding to the srfA promoter, where comS, which regulates competence development, is embedded, and the former also prevents DegU binding to the aprE promoter.23,24) The rapI-phrI cassette resides on the integrative and conjugative element Bsu1 (ICEBsu1).11) RapI activates ICEBsu1 gene expression, stimulating the excision and transfer of ICEBsu1. PhrI antagonizes the RapI function.11)

The products of orf1 and orf2 downstream of rapI showed no significant sequence similarity to those of yddM and yddN, which are positioned at the corresponding locus of the 168 chromosome15) (Fig. 2), but showed weak similarity to a hypothetical protein of Methanococcoides burtonii DSM (accession no. EAM99354) (20% identity) and to a Shewanella sp. MR-7 (EAP25066) NTPase (23% identity) respectively. Although the rapI-phrI of strain NAFM5 lacks adjacent yddK and has a downstream insertion of orf1 and orf2, conservation of the upstream yddII implies that this gene pair is also integrated in ICEBsu1, as in strain 168. Since ICEBsu1 can move among strains,11) it might have received IS4Bsu1 during transmission and might have served as a carrier of this mobile DNA element to deliver it into another strain, such as the original of natto starters. Sequencing the upstream and downstream regions of rapI-phrI should provide further insight into the structure and locus of ICEBsu1 on the NAFM5 chromosome.

The ynaE and orf3 loci (fragment B)

Another IS4Bsu1 sequence in HaeIII fragment B was identified between ynaE and orf3 (Fig. 2). The YnaE of strain NAFM5 shared 86% amino acid identity with the B. subtilis 168 counterpart of unknown function.15) However, unlike the ynaE locus of B. subtilis 168, where ynaE precedes yndD and ynaC, that in strain NAFM5 preceded orf3, which encodes a polypeptide with weak similarity (24% identity) to the ThiF proteins (molybdopterin biosynthetic proteins of the ThiF-MoeB-HesA family) of B. thiuringensis (accession no. EAO056325) and B. cereus (EAL11752, 23% identity). ThiF catalyzes the adenylation of the ThiS sulfur transfer protein in the thiamine pyrophosphate (vitamin B1) biosynthetic pathway.25) Because Orf3 has minimal similarity to B. subtilis 168 ThiF (below 20% identity) and strain NAFM5 does not require vitamin B1 for growth, this strain appears to have a thiF counterpart elsewhere.
The lytC-tuaA locus (fragment C)

IS4BSu1 on HaeIII fragment C was identified at nt −86 from the translation initiation codon of tuaA. This first gene of the teichuronic acid biosynthetic operon, tuaBCDEFGHO, encodes a membrane sugar transferase.26,27 (Fig. 2). B. subtilis NAFM5 TuaA shows 97% identity with the counterpart of B. licheniformis,28 but only 51% identity with that of B. subtilis 168. The upstream lytC encodes N-acetylmuramoyl-l-alanine amidase, a major lytic enzyme of B. subtilis, and constitutes the lytABC operon in strain 168.29 Again, LytC of strain NAFM5 closely resembles (97% identity) B. licheniformis LytC and shows less (53%) identity with B. subtilis 168 LytC.15,29 The high homology of the Lyt and Tua proteins of NAFM5 with the B. licheniformis counterparts implies that the lyt-tua locus was transferred from another bacterium, such as B. licheniformis, to an ancestor of natto starters, providing an example of natural gene acquisition by B. subtilis from closely related species.

The PhoR-PhoP two-component regulatory system is a master regulator of phosphate-dependent expression of the phosphate regulon, including the tua operon.30 Response regulator PhoP activates transcription of the tua operon at nt −43 relative to the translation start point through binding to an operator site between nt −84 and −164.26,27 The nucleotide sequences of the lytC-tuaA intergenic region, as well as the coding sequences, differ considerably between strains NAFM5 and 168 (see above), hampering assignment of the promoter and operator sequences in the NAFM5 tua operon, but B. subtilis NAFM5 proliferated even more rapidly than B. subtilis 168 under phosphate-limited conditions (≤1 mM), where the function of the tua operon appears to be indispensable for growth.30,31 This insertion thus appeared to influence expression of the tua operon minimally, and hence in the tua promoter of strain NAFM5, the PhoP-binding site should lie downstream of nt −86, the insertion point of IS4BSu1.

The yktD locus

HaeIII fragment D had an IS4BSu1 sequence in the coding region (nt +902) of yktD, which is followed by ykcC and is ahead of the neutral protease gene (nprE) (Fig. 2).32 The YktID protein was almost identical (99% identity) to the counterpart of B. subtilis 168, which has the O-methyltransferase motif (Pafm:OmtN) at the N-terminal (residues 1 to 120). This motif has been identified in an enzyme of Streptomyces glaucescens that catalyzes tetracenomycin E methylation at the C-9 carboxy group to yield the polyketide tetracenomycin A.33 B. subtilis 168 produces a polyketide34 and has no other yktD ortholog, making this gene a good candidate for the polyketide synthetic gene. Since nprE has its own promoter,35 this insert does not influence nprE expression. Indeed, strain NAFM5 actively produced NprE (data not shown).

The iolR-iolA locus

IS4BSu1 on HaeIII fragment E was determined at nt −74 relative to the transcription start of iolA, the protein product of which had 98% identity with B. subtilis 168

---

**Fig. 2.** Insertion Sites and Direction of IS4BSu1 Determined at Six Loci of B. subtilis NAFM5 Chromosome.

Gene names of B. subtilis 16845 were adopted. Putative genes of strain NAFM5 that appear to be absent in strain 168 are designated orf1 through orf5. Closed and open arrows indicate tsp of IS4BSu1 and chromosomal genes respectively. Percentages denote amino-acid sequence identities between gene products of strain NAFM5 and strain 168 or B. licheniformis DSM13. Sequence data of B. subtilis 168 and B. licheniformis DSM13 were taken from References 14 and 27.
IolA. This locus was identified by Morinaga et al. The iolA is the first locus of the iol operon genes (A through J) for myo-inositol catabolism and specifies methylmalonate-semialdehyde dehydrogenase (Fig. 2). The upstream iolR encodes a transcriptional repressor of the DeoR-family, which prevents expression of this operon. The IolR of strain NAFM5 had 99% identity with the B. subtilis 168 counterpart. A cis-acting operator site of IolR is located 45 bp upstream from the translation initiation codon of iolA. B. subtilis NAFM5 efficiently utilized myo-inositol as the carbon source (data not shown), confirming that this insertion does not affect iol operon expression.

**NAFM5 specific locus**

A sixth IS4Bsu1 sequence was found in a non-coding region between orf4, which encodes a 73 amino-acid protein with no significant similarity to known proteins, and orf5 on HaeIII fragment F. Part of Orf5 (46 amino acids) was weakly similar (28% identity) to the N-terminal region of hypothetical protein pXO1-37, comprising 89 amino acid residues (accession no. AF065404), encoded by a gene on B. anthracis plasmid pXO1. Strain 168 has no homologs of orf4 or orf5. Both orf4 and orf5 thus appear to be specific to natto starter strains. This IS4Bsu1 sequence might have been brought from an unknown source into the original of natto starters together with this locus.

**Target sequences of IS4Bsu1**

IS4Bsu1 recognizes a 9-bp target sequence that duplicates during an insertion process to flank the IS element. The 9-bp duplicates on HaeIII fragments A, C, D, E, and F had identical sequences, but that on HaeIII fragment B (the yne locus) differed by two bases, 5'-ATTTTATT-3' (left end) and 5'-ATATTA-TAT-3' (right end) (Fig. 3), presumably due to misincorporation during duplication. The target sequences are AT-rich and share nucleotides, 5'-AYTWATWYW-3' (W = A or T and Y = T or C), which closely resemble the 5'-ATNTWWWYW-3' (N = A, C, G or T) sequence that has been identified in comP as the consensus sequence for IS4Bsu1 transposition (Fig. 3). IS4Bsu1 appears to have a preferred, but not restricted, sequence for transposition, and thus is capable of moving onto many chromosome sites.

**Identification of a second IS in strain NAFM5**

Previously we mutagenized B. subtilis NAF12, a derivative of a natto starter using transposon Tn917-LTV to identify the genes essential for PGA synthesis. When we recovered EcoRI fragments that contained pBR322 rep, bla, cat, and a contiguous chromosomal DNA as plasmids from PGA-negative mutants induced by Tn917-LTV and analyzed them using restriction enzymes, some of the plasmids had a 1.4-kb insert in the pBR322 sequence. We also found that plasmids extracted from B. subtilis NAFM5 cells often contained an extra DNA segment in a vector region. Sequencing such segments revealed that they were mostly (about 85%) IS4Bsu1, and that the rest encoded a putative polypeptide (accession no. AB275894) homologous to transposases (Tnp) of the IS256 family, suggesting that B. subtilis NAFM5 has a second IS of the IS256 family in addition to IS4Bsu1. Hence we designated this new IS in strain NAFM5, IS256Bsu1 (1,369 bp). The Tnp (415 residues) of this IS shared 61, 42, 40, 39, and 26% amino-acid sequence identity with those of the IS256 family elements of Enterococcus faecalis (accession number AA081615), Clostridium thermocellum (EAM45915), Alkaliphilus metalliredigens (accession no. EAO082617), Thermoanaerobacter ethanolicus (EAO065699), and B. halodulans (AB126503) respectively. At both termini, IS256Bsu1 had imperfect 26-bp inverted repeats (IR), with nine identical inner bases (Fig. 4). The right IR (IRR) was located 70-bp upstream of the tnp coding region, and the left IR (IRL) overlapped by 1 bp with the termination codon (TAA) of tnp.

**Occurrence of IS256Bsu1 among natto starters**

Southern blotting revealed that NAFM5 and its parent Miyagino strain had nine IS256Bsu1 copies on their chromosomes (Fig. 4). This IS was also detected in the chromosomes of the other commercial starters, Naruse and Takahashi (Fig. 4), which had eight and 10 EcoRI segments respectively harboring IS256Bsu1. Of these, six fragments were the same size as those of strains NAFM5 and Miura, implying that they are located on common loci among the natto starters. Further Southern blot analysis with other restriction enzymes is required to establish the chromosomal locations of these fragments. Two EcoRI segments of strains NAFM5 and Miura were the same sizes as those of the Takahashi strain, but the remaining one was the same size as an EcoRI fragment of the Naruse strain. The Naruse and Takahashi strains had one and two additional EcoRI fragments respec-

### Table 1

<table>
<thead>
<tr>
<th>HaeIII fragments</th>
<th>Target sequences (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>ACTAAATTC</td>
</tr>
<tr>
<td>B</td>
<td>ATTTATTT (IRL)</td>
</tr>
<tr>
<td>C</td>
<td>ATAAATTT (IRR)</td>
</tr>
<tr>
<td>D</td>
<td>ACGCATGA</td>
</tr>
<tr>
<td>E</td>
<td>GATTATAT</td>
</tr>
<tr>
<td>F</td>
<td>ACTAAATTC</td>
</tr>
</tbody>
</table>

**Consensus**

AYTWATWYW

**Fig. 3.** Target Sequences of IS4Bsu1. HaeIII fragments A through F correspond to those of Fig. I. W = A or T; Y = T or C.
tively containing IS256Bsul, which were absent in the NAFMS and Miyagino strains. That the IS256Bsul copies were located on the same EcoRI fragments (perhaps the same loci) of the natto starter chromosomes substantiated the notion that natto starters originated from a common strain. On the other hand, the distribution of some IS256Bsul copies at unique loci depending on the starters and frequent translocation of IS256Bsul onto plasmid DNA indicate that this IS is active in transposition, like IS4Bsul, and that it has the potential to alter the properties of natto starters.

Acknowledgments

This study was supported in part by a grant-in-aid from the Japan Society for the Promotion of Science (no. 17380057).

References


