Nucleotide Sequence of Insertion Sequence IS3411, Which Flanks the Citrate Utilization Determinant of Transposon Tn3411

NAOTAKA ISHIGURO* AND GIHEI SATO

Department of Veterinary Public Health, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido 080, Japan

Received 10 August 1987/Accepted 22 January 1988

The nucleotide sequences of insertion sequences IS3411L (left) and IS3411R (right), present as direct terminal repeats in the citrate utilization of citrate utilization transposon Tn3411, and of IS3411 (generated by intramolecular recombination between IS3411L and IS3411R) were determined. The three IS3411 elements (IS3411L, IS3411R, and IS3411) were 1,309 base pairs long and identical in DNA sequence. IS3411 had 27-base-pair terminal inverted repeats with three bases mismatched and one long open reading frame (240 amino acids) that was proposed to be a transposase. Three polypeptides of 29,000, 27,000, and about 10,000 molecular weight, determined by IS3411, were identified in minicells. Since Tn3411 generates a 3-base-pair repeat upon integration, the nucleotide sequences of IS3411 were compared with those of IS3.

Materials and Methods

Escherichia coli K-12 derivatives, SG11 (recA and rifampin-resistant mutant of C600 thr thi leu lac) and JC1569 (leu his arg met lac mal gal man rpsL) were used as host cells for the plasmid (7). Minicell-producing E. coli M2141 (pro thi lac minA minB) (4) was used for analysis of plasmid-encoded polypeptides. The bacterial plasmids used in this study are listed in Table 1. L broth and L agar were used for routine bacterial growth (7). Simons citrate agar and minimal citrate broth were used for the determination of citrate utilization (6). To detect the transposability of IS3411, the mating system for co-integrate formation was performed as previously described (7). Large preparative isolation and purification of plasmid DNA have been previously described (9). Restriction endonucleases, the Klenow fragment of DNA polymerase I, T4 DNA ligase, and BAL 31 nuclelease were purchased from Takara Shuzo (Kyoto, Japan) or from New England BioLabs, Inc., and used as recommended by the suppliers. General procedures for cloning and DNA manipulations were as described by Maniatis et al. (12). Nucleotide sequences were determined by the chemical degradation method of Maxam and Gilbert (13), using 3'-labeled fragments, as used recently in our laboratory (8). For IS3411-encoded polypeptide analysis, minicells containing plasmids were purified as described by Jackson and Summers (10) and labeled with 20 μCi of [35S]methionine per ml (1,445 Ci/mmol; Amersham Corp.). The radioactive minicell samples were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (4).

Results and Discussion

The citrate utilization plasmid pOH2 (pBR322::Tn3411) yielded the Cit− deletion mutants by intramolecular recombination between IS3411L and IS3411R at high rates (7). To examine the functional difference between IS3411L and IS3411R for the occurrence of the deletions, we constructed the recombinant plasmids which contain directly oriented copies of IS3411L or IS3411R and examined whether they yield deletion mutants. Regardless of whether they carried either IS3411L or IS3411R, all recombinant plasmids generated deletions when they were subcultured in the broth (data not shown). From these results, we could find no significant functional difference between IS3411L and IS3411R in the formation of deletions. Such a phenomenon is remarkable in the plasmids containing direct repeats of the IS elements such as IS1 (2) and IS50 (15).

To detect polypeptides encoded by IS3411L, IS3411R, and IS3411, which is generated by intramolecular recombination between them, minicell analysis by using deletion plasmids was carried out. The physical structures of pOH7, pOH14, pOH23, pOH34, and pOH36, constructed in the previous study (7), are shown in Fig. 1A. The deletion plasmids pOH53 and pOH54 were constructed by deletion of internal HincII-digested fragments from pOH9 (Fig. 1A). Three polypeptides with molecular weights of 29,000, 27,000, and about 10,000 were produced in the minicells harboring plasmids pOH14 (lane 1), pOH23 (lane 3), and pOH34 (lane 2), slightly produced with pOH7 (lane 5), but not made at all with plasmid pOH54 (lane 8) which had an internal region of IS3411 deleted (Fig. 1B). The polypeptide with an apparent molecular weight of about 10,000 is an IS3411-coding prod-

* Corresponding author.
TABLE 1. Bacterial plasmids

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Phenotype*</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUC18</td>
<td>Ap*</td>
<td>14</td>
</tr>
<tr>
<td>pED100</td>
<td>Self-ligated HindIII fragment of F. trai+, f2</td>
<td>22</td>
</tr>
<tr>
<td>pOH2</td>
<td>pBR322::Tn3411, Cit+, Ap', Te*</td>
<td>9</td>
</tr>
<tr>
<td>pOH7</td>
<td>Self-cloning from pOH2 with BamHI, Cit+</td>
<td>9</td>
</tr>
<tr>
<td>pOH9</td>
<td>Spontaneous deletion in pOH2, pBR322::IS3411, Ap' Te*</td>
<td>7</td>
</tr>
<tr>
<td>pOH14</td>
<td>Spontaneous deletion in pOH13, pOH11::IS3411, Kmr</td>
<td>7</td>
</tr>
<tr>
<td>pOH23</td>
<td>Spontaneous deletion in pOH22, Kmr</td>
<td>7</td>
</tr>
<tr>
<td>pOH36</td>
<td>Spontaneous deletion in pOH35</td>
<td>7</td>
</tr>
<tr>
<td>pOH53</td>
<td>Deletion plasmid of HindII fragment D from pOH9, Ap', Te*</td>
<td>This study</td>
</tr>
<tr>
<td>pOH54</td>
<td>Deletion plasmid of HindII fragment E from pOH9, Ap', Te*</td>
<td>This study</td>
</tr>
</tbody>
</table>

* Cit+, citrate utilization; Ap’, ampicillin resistance; Te*, tetracycline resistance; Kmr, kanamycin resistance.

uct, and there were no significant differences among IS3411R (pOH7), IS3411L (pOH34), and IS3411 (pOH9 and pOH14). From the deletion analysis for IS3411 in pOH9 (Fig. 1A), it appears that this small polypeptide was encoded by the DNA region (ca. 451 base pairs [bp]) from the left end of IS3411 to the second HindII site. On the other hand, the 29,000- and 27,000-dalton polypeptides were not observed in the minicells harboring pOH7 and pOH9 and its derivatives (lanes 5 through 8 in Fig. 1B). It is possible that the presence of

β-lactamase (31,000-, 28,000-, and 25,000-molecular-weight bands) produced by the vector plasmid obscured IS3411-encoded 29,000- and 27,000-molecular-weight polypeptides. Since the two larger polypeptides were slightly produced by plasmid pOH23 but not by pOH36, they were encoded by the DNA region from the left end of IS3411 to the unique restriction site of HindIII. When plasmids pOH9, pOH14, and pOH34, which carried the intact IS3411 element, were donors, stable cointegrates between donor plasmids and pED100 were obtained at a frequency of 1 x 10^-7 to 3 x 10^-7 in the overnight mating. However, when deletion plasmids pOH23, pOH36, pOH53, and pOH54 were used as donors, no cointegrates were obtained. Therefore, three IS3411-encoded polypeptides seem to be required for the transposition of IS3411. However, it was not clear from this study whether three IS3411-encoded polypeptides are required for intramolecular recombination between IS3411 elements in plasmids.

To further characterize the structure of three IS3411 elements (IS3411L, IS3411R, and IS3411), nucleotide sequences of the three IS3411s were determined. Figure 2 presents a summary of sequencing strategy for IS3411 elements IS3411L and IS3411R (in pOH2) and IS3411 (in pOH9). The unique restriction sites for endonucleases Sall and HindIII on IS3411 in pOH2 and pOH9 were used as the labeling sites for the sequencing assay. For IS3411L in pOH2, the 479-, 363-, and ca. 650-bp HindII fragments were cloned into the HindII site of pUC18 (14), and DNA sequences were determined. On the other hand, for IS3411L, IS3411L, in pOH2 and IS3411 in pOH9, pOH2 and pOH9 linearized with NdeI enzyme were shorted by digestion with BAL 31

![FIG. 1. (A) Structures and restriction sites of plasmids for minicell analysis. Sections of IS3411 located in the plasmids are shown as open boxes. The dotted lines indicate the deleted DNA regions. Vertical lines indicate the positions of known cleavage sites for restriction endonucleases: E, EcoRI; H, HindIII; II, HindII; S, Sall. Ap, Ampicillin resistance; Km, kanamycin resistance; Te, tetracycline resistance; Cit*, citrate utilization. (B) Autoradiogram of [35S]methionine-labeled polypeptides synthesized in minicells harboring the plasmids. Separation was on a 15% polyacrylamide gel. Plasmids: lane 1, pOH14; lane 2, pOH34; lane 3, pOH22; lane 4, pOH36; lane 5, pOH7; lane 6, pOH9; lane 7, pOH53; lane 8, pOH54. Sizes of protein markers (cytochrome c monomer, 12,000; dimer and trimer, Oriental Yeast Co., Tokyo) are indicated on the right in kilodaltons (K). Abbreviations: BLA, β-lactamase (Ap'); TETA, polypeptide coded by tetA (Te') (I); KAN, polypeptide coded by kan (Km'). Arrows indicate IS3411-encoded polypeptides.](http://jb.asm.org/Downloaded from)
nuclease and then subsequently digested with HindIII enzyme, and resulting suitable-length fragments of IS3411 were subcloned into the polylinker regions between HindIII and HindII cleavage sites in pUC18 (14). The recombinant plasmids of pUC18 were digested with endonucleases BamHI or HindIII at sites on the vector plasmid pUC18, labeled at the 3' ends, and then analyzed for DNA sequences.

The nucleotide sequence of IS3411 elements is displayed in Fig. 3. No difference was detected among IS3411, and IS3411 \_R sequences in pOH2 and the IS3411 \_R sequence in pOH9. The complete DNA sequence of IS3411 is 1,309 bp long, and IS3411 has 27-bp terminal inverted repeats with three bases mismatched. The sequencing results for IS3411 elements in pOH2 showed that Tn3411 transposed at position 2,467 in pBR322 and generated a 3-bp (GAA) repeat upon integration (Fig. 3). This sequence duplication generated by Tn3411 was also conserved in the junction region of IS3411 in pOH9. The recognition sites for enzyme Smal were found in both terminal inverted repeats, and the other recognition sites shown in Fig. 1 and 2 were confirmed by the sequence (Fig. 3).

From minicell analysis, the 454-bp DNA fragment from the left end of IS3411 appeared to encode a small polypeptide with a molecular weight of about 10,000. The sequence of IS3411 in this region predicts three potential open reading frames (ORFs) on the same strand of DNA (ORFI, ORFII, and ORFIII in Fig. 2). The long ORFI strand extended from an ATG start codon at position 55 to a TGA stop codon at position 379 (Fig. 3), and its product would be a protein 108 amino acids long, with a molecular weight of 12,000. This ORFI is preceded by a probable RNA polymerase-binding site (marked by SD in Fig. 3) (17). Two other ORFs (II and III) are present on the same strand: ORFII, encoding 91 amino acids from nucleotide 106 (ATG) to nucleotide 379 (TAG), and ORFIII, encoding 47 amino acids from nucleotide 131 (ATG) to nucleotide 272 (TAG). However, neither of these ORFs was preceded by a probable RNA polymerase-binding site. On the other hand, ORFIV on the opposite strand in this DNA region encodes 83 amino acids from a GTG start codon at position 387 to a TGA stop codon at position 138. This ORFIV is not preceded by RNA polymerase-binding sequence.

During screening for the presence of transcriptional or translational signals, we found a long potential ORF (indicated by ORFV in Fig. 2 and 3) which was preceded by a possible RNA polymerase-binding site in the IS3411 sequence. This long ORFV has 240 amino acids from an ATG start codon at position 378 to a TAA stop codon at position 1098 (Fig. 3). The product of ORFV would be a protein with a molecular weight of 28,000. Although a possible terminator signal forming a hairpin structure was not found near the translational stop codon, three potential terminator signals were found in the DNA region behind ORFV (Fig. 3). These structures may indicate the polarity of the element.

Thus 240-amino-acid protein predicted from the DNA sequence shows low polarity (28% polar, 72% nonpolar). The ratio of basic (lysine, arginine, and histidine) to acidic (glutamate and aspartate) residues is 3.2, indicating a basic protein. Although the molecular weight of the product for ORFV in IS3411 coincided with those of two larger polypeptides found in sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Fig. 1B), it was not clear whether the product of ORFV corresponds to that of either of them. An IS3411-coding 240-amino-acid polypeptide is likely required for transposition of IS3411, because the internal deleted plasmid pOH3 (Fig. 1) did not generate cointegrate structure with pED100 in the mating experiments. We presume that ORFV, of 240 amino acids, is the IS3411 transposase.

The length of duplication in the integration site seems to be element specific, and procaryotic IS elements were classified into many groups on the basis of the length of target DNA duplication (3). The terminal inverted-repeat sequences of IS3411 were compared with those of an IS element, IS3, reported to duplicate to 4 bp upon integration (3, 20). The termini of IS3411 were related to those of IS3; a 7-bp sequence (nucleotides 1 to 10 from both ends) of IS3411 was especially homologous to that of IS3 (21). The termini are thought to be of importance because they are presumably recognized by proteins involved in transposition. Furthermore, to know the structural relationship between Tn3411 and IS3, we compared the DNA sequences of Tn3411 and IS3, using a computer program (SDC-Genetyx, Tokyo). Sequence homologies were found to be at least 30.8% in the entire DNA sequence (Fig. 3; 21). Especially high homology was found in the middle region where the putative transposase was encoded in both Tn3411 and IS3 (Fig. 3; 21). The same program was used to compare the amino acid sequences of the putative transposases of two IS elements. Some homology was found in Tn3411 and IS3 (Fig. 4). Although it is not clear how this homologous region found in
FIG. 3. Nucleotide sequence of IS3411. The IS3411 sequence is 1,309 bp long and is printed with both strands. Four ATG codons of the ORFs (I, II, III, and V) at positions 55, 106, 131, and 378 on the same strand and one GTG codon at position 385 on the reverse strand are over- and underlined, respectively. The possible -10, -35, and Shine-Dalgarno (SD) sequences (17) preceding the ORFs are indicated by the conventional symbols on the lines. The terminal codons are indicated by the broken lines. The heavy lines at the ends of IS3411 indicate the extent of the 27-bp terminal inverted repeat, with the three mismatched bases indicated by gaps in the lines. The 3-bp direct repeats are underlined. The arrows between the two strands indicate the other inverted repeat sequences. The open boxes show the recognition sequences for restriction enzymes, indicated by conventional symbols.
two IS elements is involved to the transposable functions of the IS, Tn3411 appears to be related to IS3.

Since the nucleotide sequences for IS3411L and IS3411R were exactly identical, it appears that the difference in transposition frequencies between inverse and direct transposition is not responsible for IS3411-coding functions. The terminal inverted repeats at the ends of most IS elements are thought to contribute important information such as the recognition sites for IS-encoded transposases (5, 11). IS3411 has a 27-bp terminal inverted repeat with three bases mismatched. The three mismatched DNA positions may include important recognition sites for DNA-protein interactions.

It is not yet concluded which ORF of four potential ORFs (ORF1 to ORFIV) could express in cells and correspond with this small polypeptide. Most IS elements, such as IS2 (5), IS5 (16), and IS10 (18), have ORFs which could encode a protein of 100 amino acids on the opposite strand against the putative transposase ORF. The small ORFs found on the reverse strand, in frame with the proposed transposase ORF, are considered to be involved in the regulation of transposition for IS elements (19). From this viewpoint, the ORFIV may be a reasonable candidate for the small polypeptide of 10,000 molecular weight detected in sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

The IS3411 element was widely distributed in Cit+ plasmids showing different genetic properties (unpublished data). The ubiquitous presence of this sequence in the Cit+ plasmids can explain the spread of the Cit+ determinant among plasmids of different incompatibility groups (6). Recombination and deletion promoted by IS3411 elements seem likely to reflect the evolutionary divergence of plasmid-mediated citrate utilization.

ACKNOWLEDGMENTS

This work was supported in part by grants 57480077, 57770322, and 58760211 from the Scientific Research Fund of the Ministry of Education, Science and Culture of Japan.

LITERATURE CITED


