Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 4609–4613, June 1989
Genetics

Frameshifting is required for production of the transposase encoded by insertion sequence *1*

(site-directed mutagenesis/frameshift signals/cointegration)

YASUHIKO SEKINE AND EIICHI OHTSUBO

Institute of Applied Microbiology, University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113, Japan

Communicated by Takashi Sugimura, March 23, 1989 (received for review October 27, 1988)

ABSTRACT Insertion sequence IS*1* has two coding frames, *insA* and *insB*, which are essential for its transposition. Here, we show that a frameshifting event in the −1 direction from the 3′ end region of the *insA* frame to an open reading frame (B′ frame), extending from the 5′ end of the *insB* frame, is involved in production of the InsA–B′–InsB fusion protein that has IS*1* transposase activity. The frameshifting event is likely to have occurred at the sequence AAAAAC where the *insA* frame overlaps the B′ frame. Interestingly, this sequence is also present in one of the two sequences identified in retroviruses as frameshift signals for production of transframe polyproteins from the overlapping genes *gag–pro* or *gag–pro–pol*.

Insertion sequence *1* (IS*1*) is the smallest active transposable element in prokaryotes (1–3) that is involved in various genomic rearrangements, including cointegration events between two replicons (4, 5). IS*1* is present in various copy numbers in chromosomes as well as plasmids of certain bacteria belonging to the family Enterobacteriaceae (6–9). IS*1* from plasmid R100 is 768 base pairs (bp) long (10); IS*1* elements from the other sources have divergent sequences (7, 11, 12) and hence are collectively termed iso-IS*1* elements (7). IS*1* in R100 (termed IS*1*R) comprises two reading frames, *insA* and *insB*, that are essential for transposition and cointegration (13–15). These frames are 47 bp apart and have been assumed to be transcribed polycistronically from a promoter located at the defined left-terminal region of IS*1*R (16, 17).

A 126-bp reading frame, named B′ frame, ends in-frame at the 5′ end of *insB* coding frame. Seventy-nine base pairs of the B′ reading frame overlap the 3′ end of the *insA* coding frame. The B′ frame does not seem to encode a polypeptide, since several iso-IS*1* elements have no initiation codons in the overlapping region (7, 12). The following facts, however, cannot be readily accounted for in the expression of *insA* and *insB* in IS*1*R, particularly the expression of *insB* and its extended region: (*i*) A ribosome-binding sequence does not precede *insB* but does precede *insA*. The *insA* product was detected, but the protein product of *insB* has not been demonstrated (18, 19). (*ii*) Two IS*1*R mutants defective in *insA* and *insB* do not complement each other in restoring their cointegration ability (unpublished results). (*iii*) If a 5-bp insertion is placed in the B′ region between *insA* and *insB*, then IS*1*R can no longer mediate cointegration (14).

In this report, we show genetically that part of the B′ open reading frame actually encodes a polypeptide and that a frameshifting event in the −1 direction in the region where the *insA* and B′ frames overlap is involved in the production of an InsA–B′–InsB fusion protein that has IS*1* transposase activity. Thus the expression of the transposase gene of IS*1*R can help to resolve conflicting facts associated with the region preceding *insB*. We also discuss possible frameshifting mechanisms involved in production of the IS*1* transposase.

MATERIALS AND METHODS

Bacterial Strains and Plasmids. Bacterial strains used were *Escherichia coli* K12 derivatives, JM109 (20) and JE5519 (13). Plasmid pSAM3, a pUC18 derivative, carried one copy of IS*1*S derived from the *Shigella sonnei* chromosome (21). pSEK plasmids were those carrying IS*1*R mutants. Plasmid pHS1 was a derivative of pSC101 (22).

Media. Culture media used were LB broth, L-rich broth, and ϕ medium (23). L-agar plates contained 1.5% (wt/vol) Bacto-agar (Difco) in LB broth.

Enzymes. *Kpn* I, *Pst* I, *Bam*HI, *Mlu* I, T4 DNA ligase, DNA polymerase I (large fragment), and T4 polynucleotide kinase were obtained from Takara Shuzo (Kyoto, Japan). *Bst*EII was from New England Biolabs. Enzymes were used as recommended by their suppliers.

DNA Preparation. Strain JM109 harboring a plasmid was grown in L-rich broth. The crude lysis method was used to examine and isolate a small amount of plasmid DNA from many cell cultures (24). The alkaline lysis method (25) was used to prepare plasmid DNA for nucleotide sequencing.

Nucleotide Sequencing. DNA from pUC18 derivatives was sequenced by the dideoxynucleotide method (26, 27), using a 7-DEAZA sequencing kit (Takara Shuzo), DNA polymerase I (large fragment), the 15-mer M13 primer M1 (Takara Shuzo), and a 17-mer reverse primer (Amersham). Synthetic oligonucleotides used for construction of the IS*1*R sequence were also used as primers. DNA was labeled with [α-^{32}P]dCTP (400 Ci/mmol, 10 mCi/ml; 1 Ci = 37 GBq; Amersham).

Chemical Synthesis of IS*1*R and Its Mutants. Twenty-two oligonucleotides (Figs. 1 and 2) were synthesized using an Applied Biosystems DNA synthesizer model 381A and purified on an 8% polyacrylamide/urea gel. The oligonucleotides synthesized were divided into three groups that could be displaced by the *Kpn* I–*Bst*EII, *Bst*EII–*Mlu* I, and *Mlu* I–*Bam*HI segments containing the IS*1*S sequence, an iso-IS*1* element carried by plasmid pSAM3 (Fig. 1). IS*1*S in pSAM3, which has *Bst*EII and *Mlu* I sites, as does IS*1*R, was inserted in a restriction site between the *Kpn* I and *Bam*HI sites of plasmid pUC18 (Fig. 2 and ref. 21). Several oligonucleotides were, therefore, synthesized with a cohesive end for *Kpn* I, *Bst*EII, *Mlu* I, or *Bam*HI. Also, the resultant sequence was designed to have a 9-base direct repeat at each end of the IS*1*R sequence (see Figs. 1 and 2).

To displace, for example, the *Mlu* I–*Bam*HI segment in pSAM3 with the corresponding synthetic fragment, seven oligonucleotides, named E_2, E_3, E_4, F_1, F_2, F_3, and F_4 (Fig. 1), in equal molar (5 pmol) amounts were mixed and treated with T4 polynucleotide kinase for 60 min at 37°C. Two oligonucleotides, E_1 and F_5, which have *Mlu* I and *Bam*HI cohesive ends, respectively, were then added to the phos-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: IS, insertion sequence.

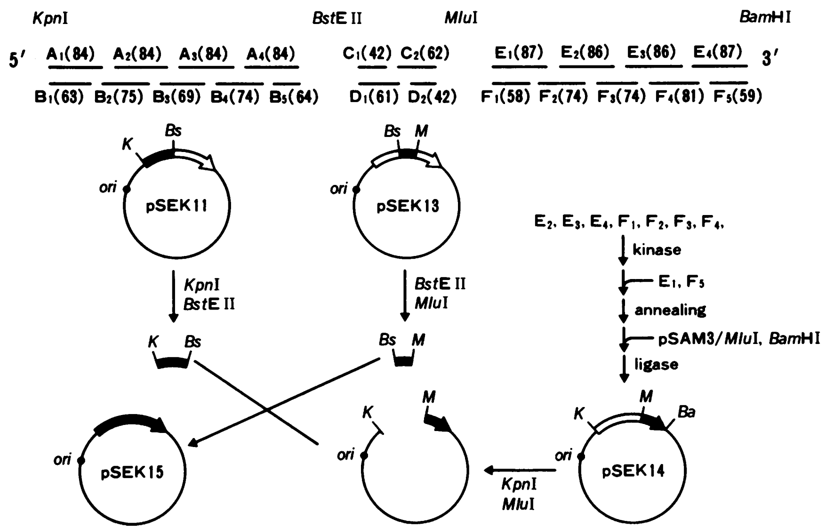

FIG. 1. Chemical synthesis of IS*1*R. Short lines represent synthetic oligonucleotides that are designated as A1, A2, etc. Numbers in parentheses are nucleotide lengths (see Fig. 2). Restriction sites used to construct IS*1*R are as follows: K, *Kpn* I; Bs, *Bst*EII; M, *Mlu* I; Ba, *Bam*HI. ori, Origin.

phorylated sample, heated at 95°C for 2 min in ligation buffer in a total volume of 10 μl, and subsequently incubated at 65°C for 15 min. After gradually cooling the sample to 37°C for annealing, 0.02 pmol of pSAM3 DNA that had been digested with *Mlu* I and *Bam*HI was added. The mixture was treated with T4 DNA ligase at 12.5°C overnight in a total volume of 30 μl and used to transform JM109 as described by Yoshioka *et al.* (23). Plasmid pSEK14 containing the synthetic *Mlu* I–*Bam*HI fragment was obtained from an ampicillin-resistant transformant of JM109. Using a strategy similar to that used to construct pSEK14, two other plasmids, pSEK11 and pSEK13, were obtained that contained the synthetic *Kpn* I–*Bst*EII fragment with oligonucleotides A_1–A_4 plus B_1–B_5 and the *Bst*EII–*Mlu* I fragment with oligonucleotides C_1, C_2, D_1, and D_2, respectively. Nucleotide sequences of the cloned fragments in these pSEK plasmids were confirmed by sequencing. To obtain plasmid pSEK15 carrying the entirely synthetic IS*1*R, both the *Kpn* I–*Bst*EII fragment of pSEK11 and the *Bst*EII–*Mlu* I fragment of pSEK13 were purified and displaced by the *Kpn* I–*Mlu* I segment in pSEK14.

Mutant IS*1*Rs were constructed by substitution of oligonucleotides that contained a specific altered sequence, and their structures were confirmed by sequencing. Mutant IS*1*Rs in pSEK plasmids have been called IS*1*Rn, where n is the plasmid designation.

Analysis of the Frequency of Cointegration. Each pSEK plasmid was introduced by transformation into strain JE5519, which harbors pHS1. Ampicillin-resistant transformants were selected at 30°C on L-agar plates containing ampicillin (50 μg/ml). Then 0.1 ml of a $1:10^6$ dilution of an overnight culture of the transformant grown at 30°C was inoculated into a 5-ml sample of L broth (15–25 samples total), and cells were grown at 30°C for 24 hr. Next, 0.2 ml of each culture was plated on agar plates containing tetracycline (10 μg/ml) and incubated at 42°C for 24 hr. The frequency of cointegration was determined from the number of tetracycline-resistant cells by a fluctuation test (30), after identification of the cells harboring the IS*1*-mediated cointegrates, in which IS*1*R or its mutant appeared as a direct repeat at both junctions of the two plasmid sequences, as described (13, 16).

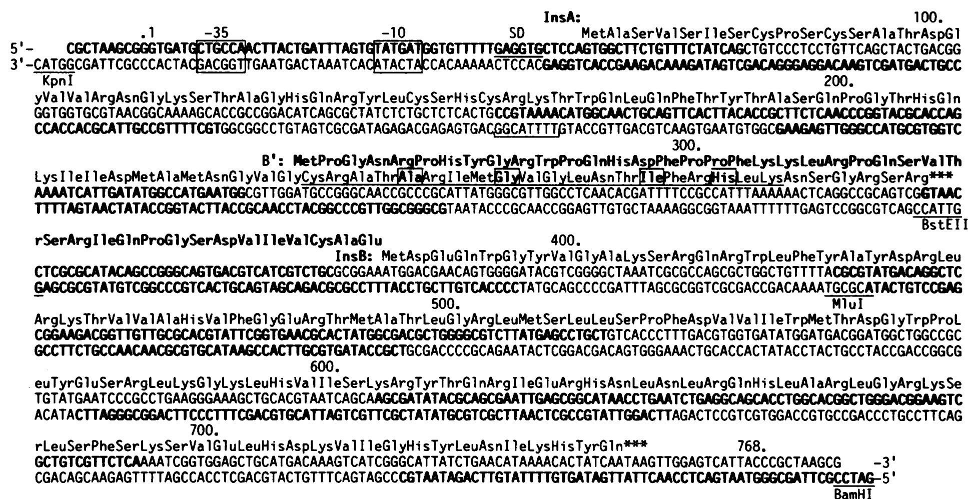

FIG. 2. Nucleotide sequence of synthetic IS*1*R. The IS*1*R sequence is numbered from 1 to 768 (10). Twenty-two synthetic oligonucleotides, which were used to construct the DNA, are distinguished by alternating between boldface and lightface letters. The DNA made has additional 9-bp direct repeats in the region flanking IS*1*R and has *Kpn* I and *Bam*HI cohesive sequences at the 5′ and 3′ ends, respectively. The 9-bp direct repeat sequence is a consensus sequence among several target sequences (28) used for the insertion of IS*1*. Putative amino acid sequences of *insA*, *insB*, and B′ coding regions are shown above the nucleotide sequences. The segment of the amino acid sequence underlined from Cys-66 to Lys-85 is likely to be responsible for DNA binding and contains consensus amino acids (shown in boxes) seen in many DNA-binding proteins (29). The rectangular boxes show a possible promoter for *insA*. A possible Shine–Dalgarno sequence is also indicated by a line between strands in the region preceding the *insA* coding frame. The region of deletion in IS*1*R16 (positions at 160–167) is bracketed.

RESULTS

Chemical Synthesis of IS*1*R and Its Mutants. To facilitate site-directed mutagenesis of the IS*1*R sequence, we synthesized 22 oligonucleotides that covered the entire IS*1*R sequence (see Fig. 2) and cloned them into pUC18 plasmids.

We then examined cointegration between two plasmids: the donor carrying a synthetic IS*1*R and the recipient pHS1, a temperature-sensitive replication mutant of a tetracycline-resistant plasmid, pSC101. Cells harboring cointegrates between the two plasmids were selected in the presence of tetracycline at 42°C. The synthetic IS*1*R was able to mediate cointegration at a frequency of 1.9×10^{-8} per division cycle, which was almost the same frequency reported previously (21). During the chemical synthesis of IS*1*R, we obtained a mutant (IS*1*R16) that had an 8-bp deletion at positions 160–167 in the *insA* coding region (positions 56–328) (see Fig. 2). As expected, this mutant mediated cointegration at a greatly reduced frequency (1.4×10^{-9}) due to the defect in the *insA* frame, which is essential for cointegration mediated by IS*1*.

Analyses of IS*1*R Mutants with Substitution Mutations. *(a) Determination of the coding region in the B′ frame essential for IS1-mediated cointegration.* As described above, previous results suggested that the B′ region is important for expression of the IS*1* transposase. Does the B′ reading frame actually encode a polypeptide essential for IS*1*-mediated cointegration? Fig. 3 shows a portion of the B′ frame nucleotide sequence from IS*1*R and from the iso-IS*1* elements IS*1*S, IS*1*D, and IS*1*F, for comparison. Base substitutions in the region of overlap between the B′ and *insA* frames of iso-IS*1* elements changed amino acids in the region encoded by the B′ frame (Fig. 3). [One substitution in IS*1*D at position 281 even introduced a termination codon into the B′ frame (Fig. 3)]. These substitutions were, however, silent or coded for amino acids that were related to ones in the *insA* frame. This suggests that the region upstream from position 304 that contains codons with silent mutations in iso-IS*1* elements codes for the InsA protein. On the other hand, the region upstream of position 304 that contains altered codons in the B′ frame may not be important.

To determine whether the B′ frame codes for a polypeptide and to further define the essential part of the B′ region, we first examined the ability of two mutants, IS*1*R21 and IS*1*R22, having nonsense mutations in the B′ reading frame between *insA* and *insB* to cointegrate (Fig. 3). Neither IS*1*R21 nor IS*1*R22 mediated cointegration in the host JE5519 (Fig. 3), demonstrating that the B′ reading frame between *insA* and *insB* is important in cointegration. In this connection, it is interesting to note that the five substitutions seen in the B′ frame between *insA* and *insB* of IS*1*F, an iso-IS*1* element, were all silent (Fig. 3).

Next, we examined another mutant, IS*1*R24, which has a mutation in the region where the *insA* and B′ frames overlap. This mutation introduced a termination codon, TAG, at codon 322 (CAG) in the B′ frame but did not alter the amino acid sequence encoded by *insA* (Fig. 3). IS*1*R24 lost its cointegration ability (Fig. 3), suggesting the importance of codon 322 (CAG) and the B′ reading frame located downstream of the codon.

(b) Determination of the essential region in the insA *frame.* To determine whether the N-terminal region of the *insA* frame that overlaps the B′ frame is essential for IS*1*-mediated cointegration, we analyzed two mutants, IS*1*R26 and IS*1*R25, with mutations that introduce a termination codon, TGA, in both positions 314 (TCA) and 305 (TTA) of the *insA* frame, respectively (Fig. 3). The mutation at position 315 in IS*1*R26 did not alter the amino acid sequence encoded by the B′ frame (Fig. 3), whereas the mutation at position 306 in IS*1*R25 changed codon 304 (TTT) in the B′ frame to TTG, having caused a conservative amino acid change, phenylalanine to leucine (Fig. 3). IS*1*R26 was found to mediate cointegration (Fig. 3), suggesting that codon 314 (TCA) and the *insA* frame downstream from the codon are not important. On the other hand, IS*1*R25 mediated cointegration at a greatly reduced frequency (Fig. 3). This result suggests that codon 305 (TTA) and the *insA* frame upstream of the codon are important, assuming that the change from codon 304 (TTT, for phenylalanine) to TTG (for leucine) in the B′ frame did not seriously affect cointegration. In fact, iso-IS*1* elements that have this type of change (see Fig. 3) could still mediate cointegration (15).

Analysis of IS*1*R Mutants with 1-Base Insertions or a 2-Base Deletion: Evidence of Frameshifting in the −1 Direction During Production of the Transposase. It is possible that the two polypeptides encoded by part of the *insA* and B′–*insB* frames are independently synthesized. An alternative possibility is that one polypeptide is made as the result of a frameshifting event from the *insA* frame to the B′ frame, which is located in the −1 frame with respect to *insA* (Fig. 2), to give the fusion protein InsA–B′–InsB acting as IS*1* transposase. If the latter is true, the sequence AAAAAC between codons 305 (TTA) and 314 (TCA) was likely to contain the site of the frame-

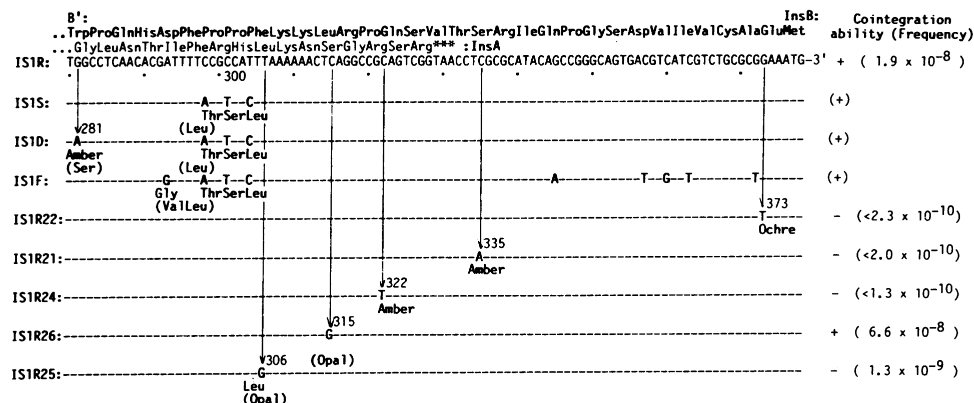

FIG. 3. Cointegration ability of IS*1*R and its mutants (IS*1*R21–IS*1*R26) and of iso-IS*1*s (IS*1*S, IS*1*D, and IS*1*F). Only a segment of each DNA element, spanning the *insA*–B′ overlapping region, is shown. Frequency of the cointegration (per division cycle) mediated by an IS*1* is indicated on the right. The putative amino acid sequences of the two reading frames are shown. The glycine residue at the left end of the sequence of InsA protein is actually residue 76, where the formylmethionine residue of InsA protein is defined as residue 1 (see Fig. 2). IS*1*S, IS*1*D, and IS*1*F are iso-IS*1* elements from chromosomes of *Shigella sonnei*, *Shigella dysenteriae*, and *Shigella flexneri*, respectively (12). Base-substitution mutations as well as amino acids altered due to the mutations in both B′ region and *insA* (above the line and below the line in parentheses, respectively) are indicated.

```
         B':                                                       Cointegration
         ..ProProPheLysLysLeuArgProGlnSerValThrSerArgIleGln        ability (Frequency)
         PheArgHisLeuLysAsnSerGlyArgSerArg*** :InsA
IS1R:    TTCCGCCATTTAAAAAACTCAGGCCGCAGTCGGTAACCTCGCGCATACAG-3'      +   (1.9 x 10^-8)
             ·300     ·          ·          ·          ·
                 InsA-┐┌-B'
         PheArgHisLeuLysLysLeuArgProGlnSerValThrSerArgIleGln
IS1R31:  TTCCGCCATTTAAAAAAACTCAGGCCGCAGTCGGTAACCTCGCGCATACAG-3'     +++ (1.4 x 10^-6)
                   +1A
                    InsA-┐ -B'
         PheArgHisLeuLysAsnLeuArgProGlnSerValThrSerArgIleGln
IS1R32:  TTCCGCCATTTAAAAAACCTCAGGCCGCAGTCGGTAACCTCGCGCATACAG-3'      ++  (9.4 x 10^-8)
                        +1C
            InsA-┐┌-B'
         PheArgHisPheLysLysLeuArgProGlnSerValThrSerArgIleGln
IS1R33:  TTCCGCCATTTTAAAAAACTCAGGCCGCAGTCGGTAACCTCGCGCATACAG-3'      ++  (8.0 x 10^-8)
               +1T
                 InsA-┐   ┌-B'
         PheArgHisLeuLys - LeuArgProGlnSerValThrSerArgIleGln
IS1R34:  TTCCGCCATTTAAAA---CTCAGGCCGCAGTCGGTAACCTCGCGCATACAG-3'      +   (9.2 x 10^-9)
                      -2A
```

FIG. 4. Cointegration ability of mutant IS*1*Rs that have the two frames, *insA* and B′, in-frame. Only the nucleotide sequences of the segments in the overlaps between *insA* and B′ frames are shown.

shifting, because the former codon is important in *insA* frame, but not the latter codon, as described above.

We can test for frameshifting by constructing and examining a mutant IS*1*R with a 1-base insertion or a 2-base deletion that has the two frames in the overlapping region aligned in the same frame and thus should produce an InsA–B′–InsB fusion protein. We expect that such a mutant mediates cointegration at a frequency much higher than that of wild-type IS*1*R, if the mutant encodes a fusion protein identical to that produced by a frameshifting event *in vivo*.

Three mutants (IS*1*R31, IS*1*R32, and IS*1*R33) used for this purpose had a 1-base insertion at different sites in or around the AAAAAC sequence (Fig. 4). IS*1*R31, which had a one-adenine insertion in a stretch of six adenine residues, was able to mediate cointegration with a frequency almost 80 times higher than that of wild-type IS*1*R; IS*1*R32, which had two cytosines instead of one at position 313 just after the run of six adenines, was able to mediate cointegration with a frequency of 5 times higher than that of wild-type IS*1*R; IS*1*R33 had an insertion of a thymine-residue to give four thymines just before the run of adenines (Fig. 4). IS*1*R33 was able to mediate cointegration with a frequency 4 times higher than that of wild-type IS*1*R (Fig. 4). Another mutant IS*1*R34, unlike the three mutants above, had a 2-base deletion reducing the run of adenines from six to four (Fig. 4). This mutant exhibited a cointegration frequency less than that of wild-type IS*1*R (Fig. 4).

These results suggest that the frameshifting event occurs *in vivo* from the *insA* frame to the B′ frame to give a fusion protein with IS*1* transposase activity through a 1-base addition, not a 2-base deletion. The fusion protein having the segment between positions 83 and 87 His-Leu-Lys-Lys-Leu encoded by mutant IS*1*R31 with the addition of one adenine is most likely to be the actual IS*1* transposase, since it was much more active in mediating cointegration than those encoded by the other insertion mutants (IS*1*R32 and IS*1*R33) with the corresponding polypeptide segment having asparagine at position 86 instead of lysine or phenylalanine at position 84 instead of leucine, respectively.

DISCUSSION

IS*1* Transposase and InsA Protein. We have shown here that the transposase of IS*1* was produced as the InsA–B′–InsB fusion protein by a frameshifting event from the *insA* frame to the B′ frame. The AAAAAC sequence in the region where the *insA* frame overlaps the B′ frame in IS*1*R is also seen in the other known iso-IS*1* elements and is likely to contain the frameshifting site.

IS*1* is believed to produce much less transposase than InsA protein (18, 19). InsA protein, which can be produced as long as the frameshifting event does not occur, binds to the terminal inverted repeats in IS*1* (19) and in fact contains a consensus amino acid sequence observed in many DNA-binding proteins (see the segment from Cys-66 to Lys-85 in Fig. 2). The InsA–B′–InsB fusion protein also includes the same segment responsible for DNA binding. This may support the previous assumption that IS*1* transposase binds to the terminal inverted repeats (named *insL* and *insR*) that are required for IS*1*-mediated cointegration (14, 31). It is likely that the InsA protein competes with the transposase, which may have a transposition-catalyzing domain in the C-terminal half in addition to the DNA-binding domain that binds to the ends of IS*1*, and thus becomes a transposition inhibitor. The characteristics of the InsA protein as a possible transposition inhibitor as well as the frameshifting event that produces IS*1* transposase in a low amount may be the two major causes for the inefficient transposition of IS*1* that avoids deleterious rearrangement of the host chromosomes containing IS*1*.

Frameshifting Mechanisms. It is possible that the frameshifting in −1 direction occurs during translation of mRNA. We showed that the mutant IS*1*R31 with the insertion of an adenine residue in the possible frameshifting region, AAAAAC, mediated cointegration at a high frequency (Fig. 4). If the transposase produced by a translational frameshifting event *in vivo* is identical to the artificial transframe protein that was produced by IS*1*R31, IS*1* transposase should have had the segment between amino acid residues 83 and 87, His-Leu-Lys-Lys-Leu, in the region encoded by the sequence between bases 302 and 315, CATTTAAAAAACTC, that includes the sequence AAAAAC (Figs. 4 and 5). To give rise to this transposase *in vivo*, frameshifting was likely to occur, such that codon 305 (TTA) in the *insA* frame was recognized as Leu-84 and then codon 307 (AAA) in the B′ frame was recognized to give Lys-85 or that codon 308 (AAA) in the *insA* frame was recognized as Lys-85 and then codon 310 (AAA) in the B′ frame was recognized to insert Lys-86. The frameshifting mechanism, by slippage of the tRNA that is reading the 0 frame codon back one nucleotide to the −1

```
Frame
        IS1R mRNA insA-B'(insB)
        .300                               340.  .349
        GCCAUUUAAAAAACUCAGGCCGCAGUCGGUAACCUCGCGCA//GGCAGUGACGUCAUCGUCUGCG
  0     ..HisLeuLysAsnSerGlyArgSerArgOc
 -1     .ProPheLysLysLeuArgProGlnSerValThrSerArg..

        HTLV-II gag-pro
        CUGAGGAAAAAAACUCCUUAAGGGGGGAGAUCUAAUCUCCCCCCAU
  0     ..GluGluLysAsnSerLeuArgGlyGluIleOc
 -1     .Op GlyLysLysLeuLeuLysGlyGlyAspLeuIleSer..

        BLV gag-pro
        UCAAAUCAAAAAACUAAUAGAGGGGGGACUUAGCGCCCCCCAAACC
  0     ..LysSerLysAsnOc
 -1     .GlnIleLysLysLeuIleGluGlyGlyLeuSerAlaPro..

        MMTV gag-pro
        AAAAUUCAAAAAACUUGUAAAGGGGCAGUCCCCUAGCCCCGC
  0     ..AsnSerLysAsnLeuOc
 -1     .LysPheLysLysLeuValLysGlyAsnSerProSerPro..
```

FIG. 5. Sequences involved in frameshifting. Nucleotide sequences of possible mRNAs from IS*1*R are shown. The common AAAAAACU sequence is indicated by boldface letters. Amino acid sequences encoded by the *insA* and B′ frames, which are out-of-frame in the −1 direction, are shown. The leucine residue in InsA protein of IS*1*R shown immediately below the nucleotide sequence is at position 84 and the lysine residue with the boldface letter in B′ frame is at position 86 in the putative IS*1* transposase protein. The amino acids indicated by boldface letters in the figure show, as if the reading frame shifted from codon 308 (AAA, for Lys-85) to codon 310 (AAA, for Lys-86) to produce IS*1*R transposase. Retroviruses shown are bovine leukemia virus (BLV) (32, 33), human T-cell leukemia virus type II (HTLV-II) (34), and mouse mammary tumor virus (MMTV) (35–37). The amino acid sequence of the transframe polyprotein encoded by the *gag–pro* coding region of mouse mammary tumor virus has been determined (37); its translation product is indicated by boldface letters. The potential of G+C-rich stem–loop structures after the frameshift signal in retroviruses (32, 34–36) and IS*1*R are underlined in the nucleotide sequence. IS*1*R has another possible stem–loop structure, which is indicated by lines above its nucleotide sequence.

frame codon, as proposed in a retrovirus system (36, 38), may be applicable to the −1 frameshifting in IS*1* (see below for further discussion). The frameshifting mechanism by a 2-nucleotide translocation in the tRNA reading the 0 frame codon, which was proposed in a bacteriophage system (39), may also be applicable to the frameshifting in IS*1*. It is also likely that the 2-base sequence AA may be recognized as if it were an AAR or AAY codon for lysine or asparagine, respectively, where R is a purine and Y is a pyrimidine. If IS*1* transposase produced *in vivo* contained the segment between residues 83 and 87, His-Leu-Lys-Lys-Leu, the AA sequence might be predominantly recognized as AAR to insert lysine.

An alternative mechanism other than the translational frameshifting would be that transcripts from the promoter located upstream of *insA* may contain mRNA with a 1-base addition at possible frameshifting sites by an error during transcription. Although this possibility cannot be excluded immediately, we believe that the translational frameshifting mechanism is likely to be correct considering the similarity between −1 frameshifting event in IS*1* and those observed in gene expression in retroviruses, as described below.

Frameshifting in Other Systems. There are several other cases of translational frameshifting in the −1 direction in prokaryotic genes, which include the bacteriophage MS2 coat gene and synthetase gene (40) and the bacteriophage T7 gene 10 (41). In these prokaryotic genes, however, no sequence is common to that seen in the possible frameshifting region of IS*1*.

In *gag*, *pro*, and *pol* genes in retroviruses, one gene downstream of the other is in the −1 reading frame. Translational frameshifting events in the overlaps can be implicated in the production of Gag–Pro, Gag–Pro–Pol, or Gag–Pol transframe polyproteins (Fig. 5). Either of the two types of sequences, AAAAAAC or UUUA conserved in the overlapping reading frames (Fig. 5), was identified as "frameshift signals" (36). Potential stem–loop structures that follow the frameshift signals (Fig. 5) also were believed to be responsible for the frameshifting (36). Such structures in Rous sarcoma virus were actually required for frameshifting *in vitro* (38). Interestingly, IS*1*R has the identical sequences with the two types of the frameshift signal sequences observed in retroviruses at regions 307–314 and 304–307 (Fig. 5), and one of them, AAAAAAC in IS*1*R, probably includes the frameshifting site, as described above. Furthermore, corresponding stem–loop structures are seen in the region after the AAAAAAC sequence in IS*1*R (Fig. 5). It is possible that the sequence AAAAAAC is also utilized as a signal for frameshifting in the production of IS*1* transposase. Our recent studies on IS*1*R mutants with mutations within the sequence AAAAAAC or TTTA suggest such a possibility, because the mutations within the former sequence caused IS*1* to lose the ability to mediate cointegration, whereas those within the latter sequence did not (unpublished results). Moreover, the UUUA sequence is unlikely to be utilized as a frameshift signal for the production of IS*1* transposase, since iso-IS*1* elements, such as IS*1*S, IS*1*D, and IS*1*F, do not contain the sequence (see Fig. 3).

Hizi *et al.* (37) have determined a segment of amino acid sequence in mouse mammary tumor virus Gag–Pro transframe protein (p30) encoded by a region including UCAAAAAACUUGUA with the motif AAAAAAC (see also Fig. 5) to be Ser-Lys-Asn-Leu-Val. They discussed the possibility that the AAC codon in the *gag* frame is recognized as asparagine and then the CUU codon in the *pro* frame directs the insertion of leucine in the product of the p30 reading frame by tRNA slippage. This frameshifting mechanism can explain our results, too, provided that the frameshifting event in the production of IS*1* transposase occurred at the same site in IS*1* as that in mouse mammary tumor virus. Our recent studies, however, suggest that the frameshifting in IS*1* occurs before the AAC codon within the AAAAAC sequence (unpublished results).

We thank Suguru Tsuchimoto for instruction in the chemical synthesis of oligonucleotides and Dr. Michio Oishi for correcting the English. This research was supported by a grant from the Ministry of Education, Science and Culture, Japan.

1. Hirsch, H.-J., Starlinger, P. & Brachet, P. (1972) *Mol. Gen. Genet.* **119**, 191–206.
2. Malamy, M. H., Fiandt, M. & Szybalski, W. (1972) *Mol. Gen. Genet.* **119**, 207–222.
3. Fiandt, M., Szybalski, W. & Malamy, M. H. (1972) *Mol. Gen. Genet.* **119**, 223–231.
4. Ohtsubo, E., Zenilman, M. & Ohtsubo, H. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 750–754.
5. Iida, S. & Arber, W. (1980) *Mol. Gen. Genet.* **177**, 261–270.
6. Saedler, H. & Heiss, B. (1973) *Mol. Gen. Genet.* **122**, 267–277.
7. Ohtsubo, H., Nyman, K., Doroszkiewicz, W. & Ohtsubo, E. (1981) *Nature (London)* **292**, 640–643.
8. Nyman, K., Nakamura, K., Ohtsubo, H. & Ohtsubo, E. (1981) *Nature (London)* **289**, 609–612.
9. Nyman, K., Ohtsubo, H., Davison, D. & Ohtsubo, E. (1983) *Mol. Gen. Genet.* **189**, 516–518.
10. Ohtsubo, H. & Ohtsubo, E. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 615–619.
11. Johnsrud, L. (1979) *Mol. Gen. Genet.* **169**, 213–218.
12. Ohtsubo, E., Ohtsubo, H., Doroszkiewicz, W., Nyman, K., Allen, D. & Davison, D. (1984) *J. Gen. Appl. Microbiol.* **30**, 359–376.
13. Ohtsubo, E., Zenilman, M., Ohtsubo, H., McCormick, M., Machida, C. & Machida, Y. (1981) *Cold Spring Harbor Symp. Quant. Biol.* **45**, 283–295.
14. Machida, Y., Machida, C., Ohtsubo, H. & Ohtsubo, E. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 277–281.
15. Machida, Y., Machida, C. & Ohtsubo, E. (1984) *J. Mol. Biol.* **177**, 229–245.
16. Machida, C., Machida, Y., Wang, H., Ishizaki, K. & Ohtsubo, E. (1983) *Cell* **34**, 135–142.
17. Machida, C., Machida, Y. & Ohtsubo, E. (1984) *J. Mol. Biol.* **177**, 247–267.
18. Armstrong, K., Ohtsubo, H., Bauer, W., Yoshioka, Y., Miyazaki, C., Maeda, Y. & Ohtsubo, E. (1986) *Mol. Gen. Genet.* **205**, 56–65.
19. Zerbib, D., Jakowec, M., Prentki, P., Galas, D. & Chandler, M. (1987) *EMBO J.* **6**, 3163–3169.
20. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33**, 103–119.
21. Matsutani, S., Ohtsubo, H., Maeda, Y. & Ohtsubo, E. (1987) *J. Mol. Biol.* **196**, 445–455.
22. Hashimoto-Gotoh, T. & Sekiguchi, M. (1977) *J. Bacteriol.* **131**, 405–412.
23. Yoshioka, Y., Ohtsubo, H. & Ohtsubo, E. (1987) *J. Bacteriol.* **169**, 619–623.
24. Machida, Y., Machida, C. & Ohtsubo, E. (1982) *Cell* **30**, 29–36.
25. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 90–91.
26. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78.
27. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
28. Galas, D., Calos, M. & Miller, J. (1980) *J. Mol. Biol.* **144**, 19–41.
29. Pabo, C. & Sauer, R. (1984) *Annu. Rev. Biochem.* **53**, 293–321.
30. Luria, S. & Delbrück, M. (1943) *Genetics* **28**, 491–511.
31. Gamas, P., Galas, D. & Chandler, M. (1985) *Nature (London)* **317**, 458–460.
32. Rice, N. R., Stephans, R. M., Burny, A. & Gilden, R. V. (1985) *Virology* **142**, 357–377.
33. Yoshinaka, Y., Katoh, I., Copeland, T. D., Smythers, G. W. & Oroszlan, S. (1986) *J. Virol.* **57**, 826–832.
34. Shimotohno, K., Takahashi, Y., Shimizu, N., Gojobori, T., Golde, D., Chen, I., Miwa, M. & Sugimura, T. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3101–3105.
35. Moore, R., Dixon, M., Smith, R., Peters, G. & Dickson, C. (1987) *J. Virol.* **61**, 480–490.
36. Jacks, T., Townsley, K., Varmus, H. E. & Majors, J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4298–4302.
37. Hizi, A., Henderson, L. E., Copeland, T. D., Sowder, R. C., Hixson, C. V. & Oroszlan, S. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7041–7045.
38. Jacks, T., Madhani, H. D., Masiarz, F. R. & Varmus, H. E. (1988) *Cell* **55**, 447–458.
39. Dayhuff, T. J., Atkins, J. F. & Gesteland, R. F. (1986) *J. Biol. Chem.* **261**, 7491–7500.
40. Atkins, J. F., Gesteland, R. F., Reid, B. R. & Anderson, C. W. (1979) *Cell* **18**, 1119–1131.
41. Dunn, J. J. & Studier, F. W. (1983) *J. Mol. Biol.* **166**, 477–535.