Mol Gen Genet (1992) 235:317–324

MGG

© Springer-Verlag 1992

Identification of the site of translational frameshifting required for production of the transposase encoded by insertion sequence IS *1*

Yasuhiko Sekine[1], **Hiromichi Nagasawa**[2], **and Eiichi Ohtsubo**[1]

[1] Institute of Applied Microbiology, University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113, Japan
[2] Ocean Research Institute, University of Tokyo, Minamidai 1-15-1, Nakano-ku, Tokyo 164, Japan

Received April 21, 1992 / Accepted June 27, 1992

Summary. Previous genetic analyses indicated that translational frameshifting in the −1 direction occurs within the run of six adenines in the sequence 5′-TTAAAAAACTC-3′ at nucleotide positions 305–315 in IS*1*, where the two out-of-phase reading frames *insA* and B′-*insB* overlap, to produce transposase with a polypeptide segment Leu−Lys−Lys−Leu at residues 84–87. IS*1* mutants with a 1 bp insertion, which encode mutant transposases with an amino acid substitution within the polypeptide segment at residues 84–87, did not efficiently mediate cointegration, except for an IS*1* mutant which encodes a mutant transposase with a Leu−Arg−Lys−Leu segment instead of Leu−Lys−Lys−Leu. An IS*1* mutant with the DNA segment 5′-CTTAAAAACTC-3′ at positions 305–315 carrying the termination codon TAA in the B′-*insB* reading frame could still mediate cointegration, indicating that codon AAA for Lys corresponding to second, third and fourth positions in the run of adenines is the site of frameshifting. The β-galactosidase activity specified by several IS*1*−*lacZ* fusion plasmids, in which B′-*insB* is in-frame with *lacZ*, showed that the region 292–377 is sufficient for frameshifting. The protein produced by frameshifting from the IS*1*−*lacZ* plasmid in fact contained the polypeptide segment Leu−Lys−Lys−Leu encoded by the DNA segment 5′-TTAAAAAACTC-3′, indicating that −1 frameshifting does occur within the run of adenines.

Key words: Adenine run – Amino acid sequencing – Cointegration – LacZ fusion protein

Introduction

Insertion sequence IS*1* is the smallest active IS element in bacteria (Ohtsubo and Ohtsubo 1978; Johnsrud 1979) and is involved in various kinds of genomic rearrangements, including the cointegration event between two replicons (Iida and Arber 1980; Ohtsubo et al. 1980, 1981). IS*1* encodes two out-of-phase reading frames, *insA* and B′-*insB*, where B′ is an open reading frame extending in-frame from the initiation codon ATG of the *insB* frame for 126 bp. The B′ reading frame, which overlaps the 3′ end of the *insA* frame, is in the −1 frame with respect to *insA*. Previous genetic analyses indicated that a frame shifting event occurs in the −1 direction within a run of six adenines, which lies in the overlap region between *insA* and B′, to fuse *insA* and B′-*insB* by translation, producing the InsA-B′-InsB fusion protein that has IS*1* transposase activity (Sekine and Ohtsubo 1989). The InsA protein, which is produced unless the frameshifting event occurs, may play a role as a negative regulator of transposition (Machida and Machida 1989; Sekine and Ohtsubo 1989; Zerbib et al. 1990). Since the efficiency of frameshifting determines the ratio between the amount of InsA and that of transposase, frameshifting is thought to be a mechanism which controls transposition of IS*1* (Sekine and Ohtsubo 1989; Escoubas et al. 1991). The production of transposase encoded by other IS elements, such as IS*3* and perhaps those related to IS*3*, has been suggested to depend on −1 frameshifting within a run of adenines in these elements (Sekine and Ohtsubo 1991). Recently, frameshifting in IS*150* which is related to IS*3*, has been demonstrated (Vögele et al. 1991).

We present here further genetic analyses which support the concept of frameshifting in IS*1* and show that the precise site of frameshifting is codon AAA for Lys in the run of adenines in *insA*. We also present here the result of amino acid sequencing analysis showing that −1 frameshifting does occur in the run of adenines to produce transposase with the polypeptide segment Leu−Lys−Lys−Leu.

Materials and methods

Bacterial strains and plasmids. Bacterial strains used were *Escherichia coli* K12 derivatives, MV1184 (*Δ*[*lac*−

Correspondence to: E. Ohtsubo

proAB] *ara strA thi*[ϕ80 *lacZΔ*M15] Δ[*srl–recA*]306::Tn*10*/F′ (*traD36 proAB lacI^q^ZΔ*M15) (Vieira and Messing 1987), BW313 (*HfrKL* 16*PO*/45[*lysA*(61–62)] *dut1 ung1 thi1 relA1*) (Kunkel et al. 1987), MC1000 (F^- *araD139* Δ[*ara leu*]*7697* Δ*lacX74 galU galK strA*) (Casadaban and Cohen 1980), JE5519 (F^- *recA1 man aroD argEH str nalA lac gal xyl mtl*) (Ohtsubo et al. 1981), GC4670 (F^- *lon*::Tn*10 thr leu lacY*) (a gift from S. Casaregola), and AD202 (F^- *ompT*::kan *araD139* Δ[*argF–lac*]*U169 rpsL150 relA1 flbB5301 deoC1 ptsF25 rbsR*) (a gift from Y. Akiyama). The strain YS202, which is a *lon*$^-$ derivative of AD202, was constructed by P1 transduction of AD202 to tetracycline resistance, using a lysate grown on GC4670.

Plasmid pSEK15, a pUC18 derivative, carries one copy of IS*1* between the *Kpn*I and *Bam*HI sites within the multiple cloning site segment (Sekine and Ohtsubo 1989). Plasmid pSEK24, a derivative of pSEK15, carries an IS*1* mutant, IS*1*-24, with a base substitution at position 322 (Sekine and Ohtsubo 1989); plasmids pSEK31, pSEK32 and pSEK33, which were derived from pSEK15, have IS*1* mutants each with a 1 bp insertion, IS*1*-31, IS*1*-32 and IS*1*-33, respectively (Sekine and Ohtsubo 1989). Plasmid pUC119 (Vieira and Messing 1987) was used to construct pSEK117, as will be described below. Construction of pSEK17 and its derivatives will also be described below. Plasmid pHS1 is a temperature-sensitive replication mutant of the tetracycline-resistance plasmid pSC101 (Hashimoto-Gotoh and Sekiguchi 1977). Plasmid pR-pMLB (a gift from D. Bastia) is a pBR322 derivative, from which IS*1*–*lacZ* fusion plasmids were constructed as described below.

Media. Culture media used were L broth, L-rich broth, ϕ-medium (Yoshioka et al. 1987) and 2×YT broth (Messing 1983). ϕ-medium was used for transformation of plasmid DNA, and 2×YT broth was used for mutagenesis in constructing mutant plasmids. L-agar plates contained 1.5% (w/v) agar (Eiken Chemical) in L broth. Antibiotics were added in L-agar plates, if necessary, at concentrations of 150 µg ampicillin/ml (Wako Junyaku), 5 or 10 µg tetracycline/ml (Sigma). Peptone dilution buffer (0.1% peptone (Kyokuto Seiyaku) in 0.3% NaCl) was used for dilution of cell cultures.

Enzymes. Restriction endonucleases (*Bam*HI, *Bgl*II,-*Hin*dIII, *Kpn*I, *Pst*I, *Sal*I and *Sph*I), bacterial alkaline phosphatase, T4 DNA polymerase, T4 polynucleotide kinase, and T4 DNA ligase were purchased from Takara Shuzo. Restriction endonuclease *Bst*EII was purchased from New England Biolabs. RNase A was purchased from Sigma. These enzymes were used in the buffers recommended by their suppliers.

DNA preparation. Strain MV1184 or MC1000 harboring a plasmid was grown in L-rich broth. The alkaline lysis method (Maniatis et al. 1982) was used to prepare plasmid DNA for cloning and nucleotide sequencing.

Nucleotide sequencing. Nucleotide sequences were determined by the dideoxynucleotide method (Sanger et al. 1977; Messing 1983) using a 7-DEAZA sequencing kit and M4 primer (Takara Shuzo). Synthetic oligodeoxyribonucleotide D_2 described in Sekine and Ohtsubo (1989) was also used as primer for sequencing derivatives of pSEK117. The DNA chains were labeled with α-[^{32}P]dCTP (15 TBq/mmol, Amersham) and separated in 6 or 8% polyacrylamide gels containing 8 M urea.

Plasmid construction. pSEK17 has only one cleavage site for *Pst*I in the *insA* coding region in IS*1*, and was constructed by self-ligation of plasmid pSEK15 after digestion with *Sph*I and *Sal*I and treatment with DNA polymerase I (Klenow) to remove the *Pst*I site flanked by the *Sph*I and *Sal*I sites present in the pUC18 sequence. Each pSEK17 derivative carrying an IS*1* mutant with a substitution(s) or a 1 bp insertion was then constructed as follows. The *Kpn*I–*Hin*dIII fragment in the cloning site segment in vector plasmid pUC119 was replaced with the *Kpn*I–*Hin*dIII fragment of pSEK17 containing the entire IS*1* sequence, yielding pSEK117. Using pSEK117 as template and oligodeoxyribonucleotides synthesized using a DNA synthesizer 380B (Applied Biosystems) as primers, the IS*1* sequence in pSEK117 was mutagenized by site-directed mutagenesis according to Kunkel et al. (1987). The sequences mutated were confirmed by DNA sequencing. Then the *Pst*I–*Bst*EII fragment of IS*1* in pSEK17 was replaced with the *Pst*I–*Bst*EII fragment of each of the resulting pSEK117 derivatives.

Plasmids pSEK2055, pSEK207, pSEK9000 and pSEK6000 are IS*1*–*lacZ* fusion plasmids having a DNA fragment of wild-type IS*1*. These were constructed as follows. Using pSEK117 as template, two *Bgl*II recognition sites were introduced into appropriate positions (see Fig. 3) flanking the run of adenines within the IS*1* sequence, by site-directed mutagenesis as described above. The sequences mutated were confirmed by DNA sequencing. Each of the *Bgl*II fragments was then inserted into the *Bam*HI site of vector plasmid pR-pMLB. Plasmids pSEK2055-I, pSEK207-I, pSEK9000-I and pSEK6000-I are IS*1*–*lacZ* fusion plasmids having a DNA fragment of IS*1* with a single adenine insertion in the run of adenines. These were constructed in the same way as the IS*1*–*lacZ* fusion plasmids, such as pSEK2055 etc. using, however, plasmid pSEK131, which carries IS*1*-31 with a single adenine insertion in the run of adenines as template; pSEK131 itself was obtained by replacing the *Kpn*I–*Hin*dIII fragment in the cloning site segment of vector plasmid pUC119 with the *Kpn*I–*Hin*dIII fragment of pSEK31 which includes IS*1*-31 (Sekine and Ohtsubo 1989).

Purification of β-galactosidase (LacZ) fusion proteins and amino acid sequencing. Strain YS202, harboring the IS*1*–*lacZ* fusion plasmid pSEK9000 or pSEK9000-I, was grown in L-rich broth (30 l for pSEK9000 and 250 ml for pSEK9000-I), containing 0.2% (w/v) glucose at 30° C until the OD_{600} reached 0.5–0.6, and then the culture was incubated with aeration at 40° C for 60 min to induce the LacZ fusion protein. The protein was purified from these cells according to the procedure of In-

amoto and Ohtsabo (1990), except that a French press was used to disrupt the cells instead of sonication. After lyophilization, the protein was solubilized in water, subjected to SDS-polyacrylamide gel electrophoresis, and subsequently electroblotted onto PVDF-type membrane (ProBlott, Applied Biosystems), according to the method of Aebersold et al. (1986). The protein band of interest was excised and subjected to amino acid sequencing analysis using an Applied Biosystems model 470A sequencer or model 477A sequencer fitted with an on-line Applied Biosystems 120A high performance liquid chromatography analyzer. Here, to avoid possible degradation of the protein we used strain YS202. This strain carries a mutation in the *lon* gene, which encodes a protease involved in the degradation of some unstable proteins (for a review, see Gottesman 1989), and a second mutation in the *ompT* gene, which encodes an outer membrane-associated protease responsible for in vitro cleavage of several proteins, including the SecY protein (Akiyama and Ito 1990), during their purification. Indeed when an *ompT*$^+$ strain was used as a host, cleavage of the protein between two consecutive basic amino acids was observed.

Cointegration assay. Each of the ampicillin-resistance pSEK plasmids carrying wild-type IS*1* or mutant IS*1* was introduced by transformation into the *E. coli* K12 strain JE5519, which already harbored the tetracycline-resistance plasmid pHS1. Cointegration between a pSEK plasmid and pHS1 was assayed according to the method described in Sekine and Ohtsubo (1989).

LacZ assay. Each of the ampicillin resistance plasmids carrying the IS*1*−*lacZ* fusion gene was introduced by transformation into MC1000. Liquid cultures of MC1000 harboring a plasmid were incubated overnight at 30° C in L-rich broth containing 100 μg ampicillin/ml, and diluted 1/100 into L-rich broth. After shaking at 30° C until turbidity at 600 nm had reached 0.15–0.2, the temperature was shifted to 40° C. When the turbidity of the culture was 0.8–0.9, LacZ activity was measured by the method described by Miller (1972).

	Sequence	Mutation	Cointegration frequency
B'-*insB*	(85)K 86K 87L		
insA	84L 85K N		
IS*1*	TTAAAAAACTC (305)		1.9 x 10^{-8} (1.4)
	84L 85K 86K 87L		
IS*1*-31	TTAAAAAA**A**CTC	+1A	1.4 x 10^{-6} (100)
	F		
IS*1*-32	TT**T**---------	+1T	8.0 x 10^{-8} (5.7)
	N		
IS*1*-33	------AA**C**---	+1C	9.4 x 10^{-8} (6.7)
	E		
IS*1*-35	---**G**AA------	+1G	1.8 x 10^{-8} (1.3)
	Q		
IS*1*-36	---**C**AA------	+1C	2.2 x 10^{-7} (16)
	R		
IS*1*-37	---A**G**A------	+1G	2.4 x 10^{-6} (170)
	V		
IS*1*-38	---------**G**TC	+1G	7.5 x 10^{-9} (0.54)

Fig. 1. Frequency of cointegration mediated by wild-type IS*1* or each IS*1* mutant with a 1 bp insertion. The nucleotide sequence of a critical region of IS*1* (305–315) of IS*1* and of each IS*1* mutant is shown together with putative amino acids encoded by the two reading frames, *insA* and B′−*insB*. The leucine residue (L) on the left of the amino acid sequence is actually amino acid residue 84 of the InsA protein, where the first methionine residue is defined as residue 1. The 85th lysine residue (^{85}K) is actually encoded by codon ^{308}AAA in *insA* as described below (see Fig. 2). IS*1*-31 with a single adenine insertion is supposed to produce wild-type IS*1* transposase with a polypeptide segment LKKL without frameshifting. In the other IS*1* mutants, which produce mutant transposases, only the codons altered and the amino acids substituted are indicated. *Boldface* letters indicate the mutated nucleotides. The frequency of cointegration (per division cycle) mediated by wild-type IS*1* or each IS*1* mutant is shown with its relative value in parenthesis by taking the frequency for IS*1*-31 as 100. Cointegration frequencies mediated by IS*1*-31, IS*1*-32 or IS*1*-33 were taken from Sekine and Ohtsubo (1989)

Results

Determination of the site of frameshifting of IS1 *by genetic analyses*

Our previous analyses of the cointegration ability of IS*1* mutants carrying a nonsense mutation in the *insA* or B′-*insB* reading frame have suggested that the run of six adenines at nucleotide positions 307–312, where the two frames overlap (see Fig. 3), is likely to contain the possible frameshift site (Sekine and Ohtsubo 1989). Also, our analyses of several IS*1* mutants with single bp insertions within or close to the run of adenines which results in the placement of *insA* and B′-*insB* in the same reading frame, have shown that one IS*1* mutant with a single adenine insertion in the run of adenines (see mutant IS*1*-31 in Fig. 1) can mediate cointegration at a much higher frequency than does either wild-type IS*1* or any of the other mutants with a 1 bp insertion neighboring the run of adenines (see IS*1*-32 and IS*1*-33 in Fig. 1) (Sekine and Ohtsubo 1989). The mutant IS*1*-31 is considered to produce, without frameshifting, active transposase in which the amino acid sequence at residues 84–87 is Leu−Lys−Lys−Leu (LKKL in Fig. 1), but the other mutants produce inactive transposase with FKKL or LKNL (see Fig. 1). These results have suggested that wild-type IS*1* produces transposase having the polypeptide segment LKKL by −1 frameshifting in the run of six adenines. To support this suggestion, we constructed four other mutants with a 1 bp insertion designed to alter the amino acid sequence in the possible IS*1* transposase. We then examined cointegration between each of these mutant plasmids as donor and pHS1, a temperature-sensitive replication mutant of a tetracycline-resistance plasmid pSC101 as recipient, by selecting for cells harboring cointegrates which can replicate in the presence of tetracycline at 42° C. Three mutants, IS*1*-35, IS*1*-36 and IS*1*-38 (Fig. 1), which generate

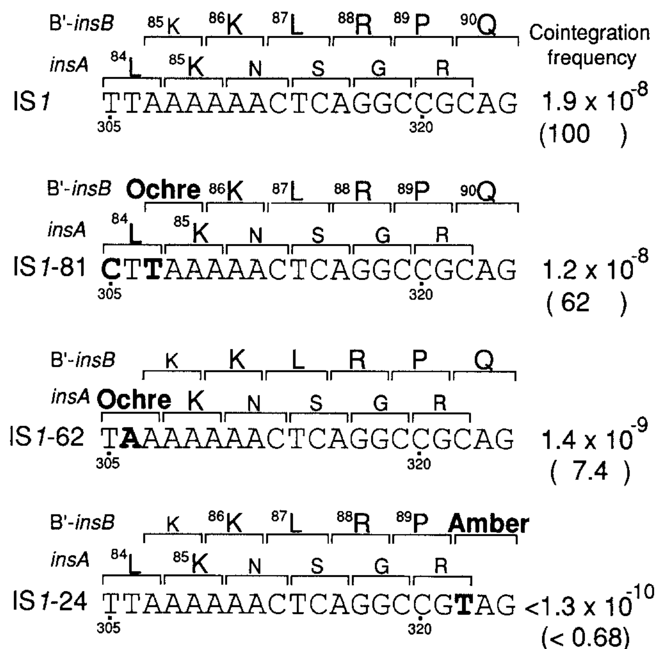

Fig. 2. The cointegration frequency of IS*1* mutants with substitutions. Only the nucleotide sequence of a critical region of IS*1* (305–324) is shown. Substituted nucleotides and codons altered are shown by *boldface* letters. The cointegration frequency (per division cycle) mediated by wild-type IS*1* or each IS*1* mutant is shown with its relative value in parenthesis by taking wild-type IS*1* as 100. The cointegration frequency mediated by IS*1*-24 was taken from Sekine and Ohtsubo (1989). See the legend to Fig. 1 for further details

transposases with LEKL, LQKL and LKKV polypeptide segments, respectively, instead of LKKL, did not efficiently mediate cointegration. Conversely, one mutant (IS*1*-37 in Fig. 1) which generates a transposase with an LRKL segment, with a related amino acid substitution R (Arg) for K (Lys) at the 85th residue, mediated cointegration at an even higher frequency than that of IS*1*-31 (Fig. 1). This supports the suggestion above that IS*1* transposase with the polypeptide segment LKKL at residues 84–87 is the product of frame-shifting and used to mediate cointegration in wild-type IS*1*.

To give rise to the transposase with the polypeptide segment LKKL from the DNA sequence 5′-TTAAAAAACTC-3′ at nucleotide positions 305–315, frameshifting is likely to occur after recognition of the codon TTA at nucleotide positions 305–307 (designated ^{305}TTA) in *insA* as the 84th residue L (^{84}L in Fig. 2) in one of the following two ways (see Fig. 2): (i) codon ^{307}AAA in B′-*insB* is recognized to give ^{85}K; or (ii) codon ^{308}AAA in *insA* is read as ^{85}K and then codon ^{310}AAA in B′-*insB* is recognized as ^{86}K. To determine which is the case, we constructed an IS*1* mutant, IS*1*-81, with two substitutions, C for nucleotide T at 305 and T for nucleotide A at 307, which not only convert codon ^{305}TTA for ^{84}L in *insA* to the synonymous codon CTT, but also convert codon ^{307}AAA for the first K in B′-*insB* to the ochre codon TAA (Fig. 2). This IS*1* mutant retained the ability to mediate cointegration (Fig. 2). This indicates that the mutant produces active transposase upon frameshifting, using the ^{308}AAA codon for ^{85}K as the last codon in *insA* at which −1 frameshifting occurs during translation to give rise to transposase. The other two IS*1* mutants were constructed as negative controls: IS*1*-62 with a substitution of A for T at position 306, which introduces an ochre codon in *insA* at a position upstream of the run of adenines, and IS*1*-24 with a T for C substitution at position 322 which introduces an amber codon in B′-*insB* downstream of the run of adenines. These mediated cointegration at greatly reduced frequencies (Fig. 2) as expected, since both of the termination codons introduced resulted in production of a truncated protein and, therefore, inactive transposase.

Analysis of frameshifting using IS1 − lacZ *fusion plasmids*

To determine the nucleotide sequence required for and the efficiency of, frameshifting in IS*1*, we constructed several IS*1* − *lacZ* fusion plasmids having a DNA fragment of wild-type IS*1* containing the run of adenines. The fragment is flanked by the ATG of *cro* of bacteriophage λ and the *lacZ* gene, such that *insA* is fused with the initiation codon ATG_{cro} in-frame and B′-*insB* is fused with *lacZ* in-frame (Fig. 3). The transcription of the fusion gene from the λ p_R promoter for *cro* is under the control of a thermosensitive repressor, the product of *cI857* which is also carried by the fusion plasmid. The occurrence of −1 frameshifting during translation of *insA* would therefore result in the synthesis of the InsA−B′−InsB−LacZ fusion protein with β-galactosidase (LacZ) activity. We also constructed IS*1*−*lacZ* fusion plasmids having a DNA fragment of IS*1* with a single adenine insertion in the run of adenines, so that *insA* is fused with B′ − *insB* − *lacZ* in-frame. The ratio of LacZ activity specified by the out-of-frame plasmid to that specified by the corresponding in-frame plasmid is considered to reflect the efficiency of frameshifting in the out-of-frame plasmid.

The LacZ activity measured after heat induction in lysates of cells harboring pSEK2055, having a DNA insert corresponding to IS*1* coordinates 63–377 and containing the entire *insA* and B′ coding frames, was 1.73 units (Fig. 3). On the other hand, LacZ activity specified by pSEK2055-I, a derivative of pSEK2055 with a single adenine insertion in the run of adenines, was 811 units (Fig. 3). Thus, the efficiency of frameshifting in pSEK2055 was estimated to be 0.21% (Fig. 3). LacZ activity specified by plasmid pSEK9000 having the region 292–377 was 17.9 units, while the LacZ activity specified by the corresponding in-frame plasmid pSEK9000-I was 5220 units (Fig. 3). The efficiency of frameshifting in pSEK9000 was thus estimated to be 0..34% (Fig. 3), which was alsmost the same level as that in pSEK2055. This suggests that the region 63–291, which is upstream of the run of adenines, is not required for frameshifting. The increase of LacZ activity specified by pSEK9000 (and pSEK9000-I), when compared with that specified by pSEK2055 (and pSEK2055-I) (see Fig. 3), may reflect instability of a portion of the InsA

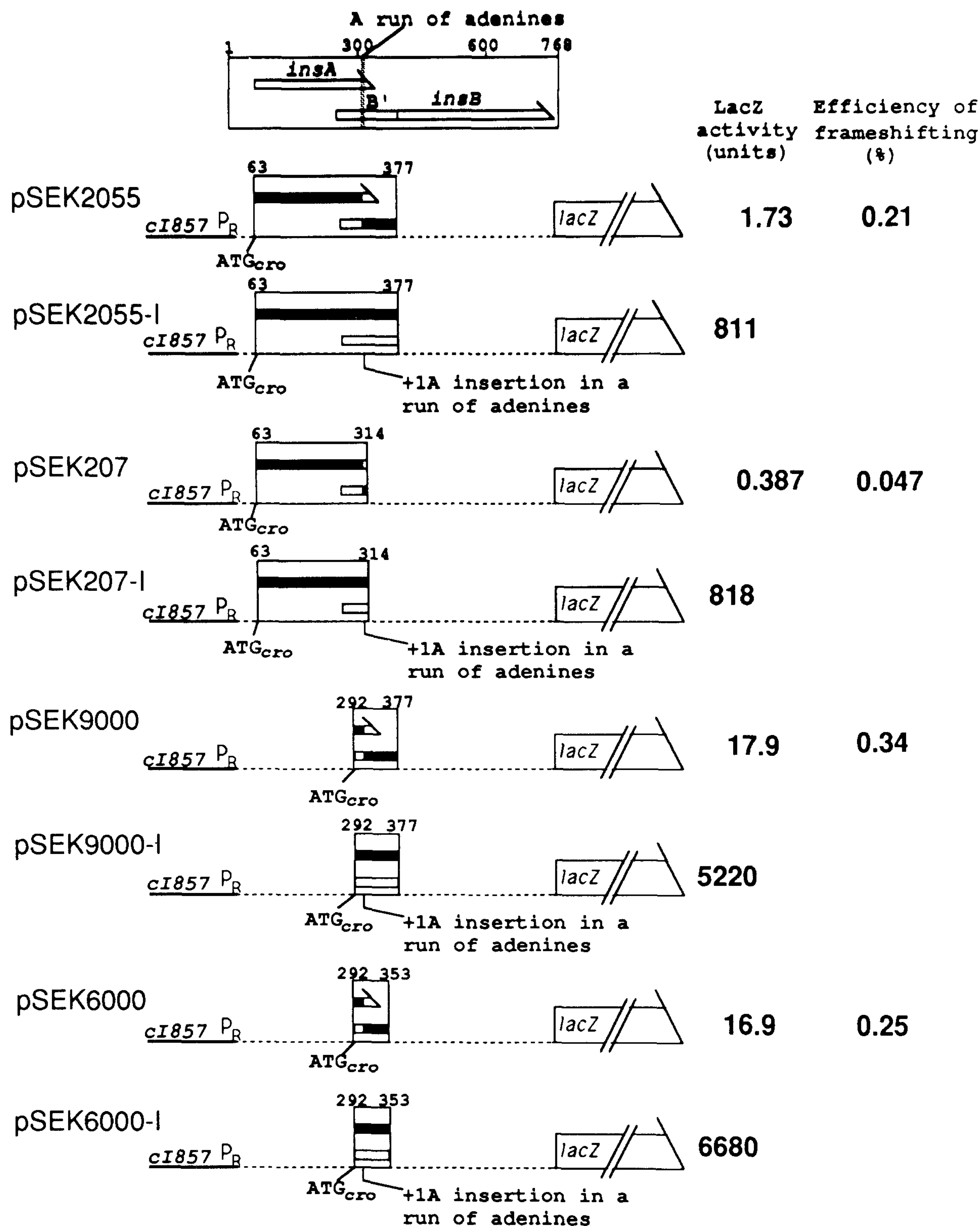

Fig. 3. Structures of IS*1*−*lacZ* fusion plasmids and the activity of LacZ produced in cells harboring each of the plasmids. IS*1* (with coordinates 1–768; Ohtsubo and Ohtsubo 1978) has two overlapping open reading frames, *insA* and B′-*insB*, shown by two *arrows* within IS*1* at the top. Each plasmid carries a DNA fragment of IS*1* containing a part of the *insA* and B′−*insB* frames, each of which is in-frame with ATG_{cro} and *lacZ*, respectively. pSEK2055-I, etc. are the plasmids carrying a DNA fragment with a single adenine insertion in the run of adenines which results in fusion between the two frames to give in-frame expression of *lacZ*. Nucleotide positions of the ends of each IS*1* fragment are indicated. *Filled* portions indicate a frame(s) required to give the InsA−B′−InsB−LacZ fusion protein. Production of the fusion protein is controlled at promoter p_R by a thermosensitive repressor, the product of *cI857*. *Broken lines* show regions of IS*1* deleted in each plasmid. The LacZ activity in the lysate of the cells harboring each plasmid was assayed at least three times; standard errors were less than 15%. Efficiency of frameshifting (%) was obtained by calculating the ratio of LacZ activity specified by the out-of-frame plasmid to that specified by the corresponding in-frame plasmid

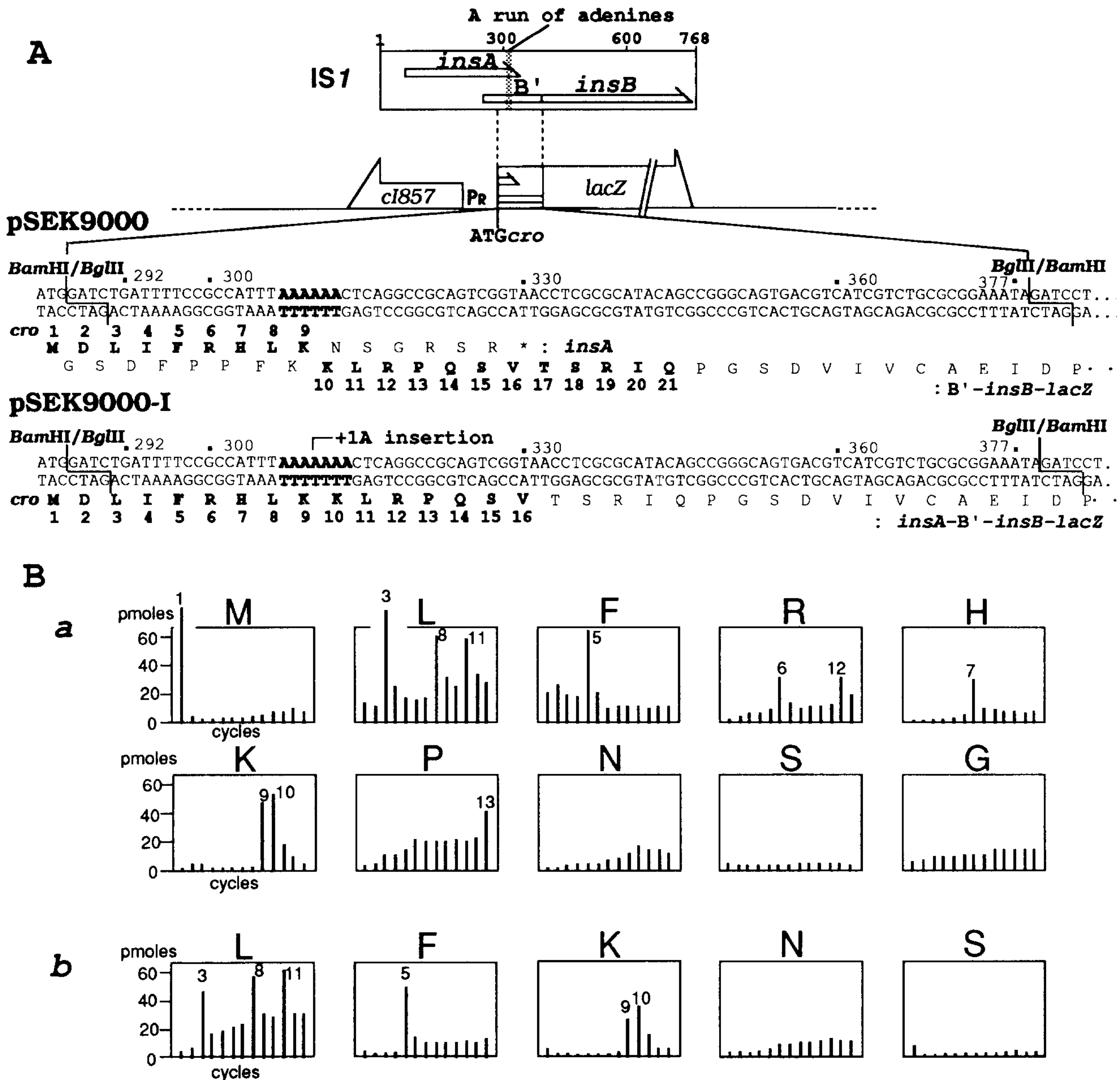

Fig. 4. **A** Nucleotide sequences of a critical region in the IS*1*−*lacZ* fusion plasmids and amino acid sequences of the proteins encoded by the region. Plasmid pSEK9000 has an IS*1* fragment (coordinates 292–377) containing a part of *insA* and of B′−*insB* fused with ATG_{cro} and *lacZ*, respectively. Plasmid pSEK9000-I is a derivative of pSEK9000 with a single adenine insertion in the run of adenines (shown in *boldface*) which results in in-frame fusion between *insA* and B′-*insB-lacZ*. Numbers above the nucleotide sequence represent IS*1* coordinates. Recognition sites of restriction endonucleases used to construct the plasmids are shown. The amino acids encoded by the two reading frames are shown. The amino acids indicated in *boldface* are those determined by sequencing the purified LacZ fusion protein produced; numbers above or below the amino acids indicate cycle numbers of Edman degradation of the LacZ fusion protein. The 9th lysine residue (K) of the protein specified by pSEK9000 is encoded by codon ^{308}AAA in *insA* as described above (see Fig. 2). **B** Critical phenylthiohydantoin (PTH)–amino acids (in pmol) detected during each sequencing cycle of Edman degradation of the purified LacZ fusion protein produced from cells harboring pSEK9000 (*a*) or pSEK9000-I (*b*). The data are not corrected for injection, base line, and tailover. A number over a vertical line represents the major PTH-amino acid recovered from that cycle

protein or a portion of the transcript encoded by pSEK2055 (and pSEK2055-I).

The efficiency of frameshifting in pSEK6000 carrying the region 292–353 was 0.25% (Fig. 3) which was reduced relative to that in pSEK9000; a similar reduction has been reported by Escoubas et al. (1991). The degree of the reduction, however, seemed not to be so great, and the efficiency in pSEK6000 is the same as that in pSEK2055. Moreover, the efficiency in pSEK6100 having the region 292–332 was the same as that in pSEK2055 and pSEK6000 (data not shown). In the two plasmids, pSEK6000 and pSEK6100, several possible secondary structures seen in the region downstream of the run of adenines (Sekine and Ohtsubo 1992) are deleted. These show that the region 333–377 is not essential for frameshifting, suggesting that the contribution of the secondary structures downstream of the run of adenines to the efficiency of frameshifting is small (Sekine and

Ohtsubo 1992). Note here that the efficiency of frameshifting in pSEK207 having the region 63–314 was, however, 0.047% (Fig. 3), a 4.5-fold decrease compared with that in pSEK2055. This suggests that the region 315–377, which is located downstream of the run of adenines, is required to stimulate frameshifting. The reduction of efficiency of frameshifting seen in pSEK207 was considered to be caused mainly by the elimination of the termination codon of *insA*, but not by the elimination of secondary structures, as described above and by Sekine and Ohtsubo (1992).

Analysis of the amino acid sequence encoded by an IS1 *segment containing the run of adenines*

To obtain direct evidence for frameshifting in IS*1*, we determined the amino acid sequence of the protein produced in the form of the InsA – B′ – InsB – LacZ fusion protein from plasmid pSEK9000, which contained a region of IS*1* sufficient for efficient frameshifting, as described above. In this plasmid, the codon for Leu at the 84th residue in IS*1* transposase becomes the 8th codon in the coding region for the LacZ fusion protein, where the ATG of *cro* is the first codon (see Fig. 4A). The LacZ fusion protein, which was overproduced by heat induction in cells harboring pSEK9000, was readily purified using a LacZ-specific affinity column (see Materials and methods). The purified protein was subjected to 21 cycles of Edman degradation (Fig. 4A), and the relevant phenylthiohydantoin (PTH)-amino acids detected in the first 13 cycles are shown in Fig. 4B (*a*). This shows that translation of the LacZ fusion protein was initiated at the ATG_{cro} and continued in-frame along *insA*, but shifted into B′ – *insB* – *lacZ*. The amino acid sequence detected at cycles 8–11, which was produced from the sequence 5′-TTAAAAAACTC-3′, was LKKL (Fig. 4B, *a*), as expected from the results obtained from genetic analyses described above.

We also overproduced, purified, and sequenced the LacZ fusion protein from cells harboring the corresponding in-frame plasmid pSEK9000-I (Fig. 4A). The pattern of PTH-amino acids (Fig. 4B, *b*) was the same as that from the protein specified by pSEK9000 (Fig. 4B, *a*), as expected.

Discussion

We have demonstrated here that shifting of the reading frame from *insA* to B′-*insB* in IS*1* occurs within the run of six adenines in the −1 direction so as to produce IS*1* transposase with amino acid segment LKKL at residues 84–87, and that the last codon in *insA* recognized during translation of IS*1* transposase is ^{308}AAA in the run of adenines for ^{85}K. These results suggest that a tRNA for ^{85}K, recognizing the codon ^{308}AAA in *insA*, plays a key role in frameshifting in the −1 direction. This suggestion is further supported by mutational analysis of the run of adenines (Sekine and Ohtsubo 1992).

A run of adenines has been implicated in −1 frameshifting in other genes, such as the *dnaX* gene of *E. coli* (Blinkowa and Walker 1990; Flower and McHenry 1990; Tsuchihashi and Kornberg 1990), genes encoded by IS*150* (Vögele et al. 1991) and IS*3* (Sekine and Ohtsubo 1991), and the *gag* and *pro* genes of mouse mammary tumor virus (MMTV) (Hizi et al. 1987; Jacks et al. 1987; Moore et al. 1987). In MMTV, −1 frameshifting occurs at the *gag-pro* overlap, with the sequence AAAAAAC identical to that of IS*1*, to produce the transframe protein, such that the last codon in *gag* (0 frame) is codon AAC (Hizi et al. 1987) and is downstream by one codon when compared with the site in IS*1*. This difference between IS*1* and MMTV and probably other retroviruses which are thought to use the AAAAAAC sequence for frameshifting (Rice et al. 1985; Shimotohno et al. 1985) might be due to the structural or functional differences between prokaryotic and eukaryotic molecules participating in the translational process. The run of six adenines in *dnaX* and IS*150* is followed by a guanine residue instead of a cytosine residue, unlike IS*1* and MMTV. In *dnaX*, the amino acid sequence of the protein produced by frameshifting from the DNA segment 5′-GCAAAAAAGAG-3′ is AKKE (Tsuchihashi and Kornberg 1990), indicating that frameshifting occurs at one of the consecutive codons, AAA and AAG, for lysine (K) in the 0-frame. In IS*150*, the amino acid sequence of the protein produced by frameshifting from the DNA segment 5′-CUAAAAAAGCU-3′ was LKKA (Vögele et al. 1991), indicating that frameshifting occurs at either codon CUA for leucine (L) or one of the consecutive codons, AAA and AAG, for lysine (K) in the 0-frame. Since the exact site for frameshifting is not clear in *dnaX* or IS*150*, it is unknown at present whether the site of frameshifting in these genetic systems is the same as that in IS*1* or not.

Secondary structures downstream of the frameshift site have been demonstrated to stimulate frameshifting in many genetic systems (Jacks et al. 1987, 1988; Brierley et al. 1989; Flower and McHenry 1990; Tsuchihashi and Kornberg 1990; Dinman et al. 1991; Vögele et al. 1991), but such structures seem not to be responsible for efficient frameshifting in IS*1* (for further experiments on, and discussion of the role of the secondary structures in IS*1*, see Sekine and Ohtsubo 1992). We have shown here that the efficiency of frameshifting in IS*1* is 0.2–0.3%. This is very low when compared with other cases, for example, 5–25% in retroviruses (for a review, see Varmus and Brown 1989), 40–50% in *dnaX* (Flower and McHenry 1990; Tsuchihashi and Kornberg 1990), and 30% in IS*150* (Vögele et al. 1991). The lack of the secondary structures which stimulate frameshifting might result in such a low efficiency of frameshifting in IS*1*. It is reasonable to assume that IS*1* adopts a low level of frameshifting, which results in a low level production of transposase, to avoid deleterious rearrangement of the host chromosome containing IS*1*.

Acknowledgements. We thank D. Bastia for kindly sending us plasmid pR-pMLB, S. Casaregola for strain GC4670, and Y. Akiyama

for strain AD202. This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan. Y.S. is the recipient of a JSPS Fellowship for Japanese Junior Scientists.

References

Aebersold RH, Teplow DB, Hood LE, Kent SBH (1986) Electroblotting onto activated glass. J Biol Chem 261:4229–4238

Akiyama Y, Ito K (1990) SecY protein, a membrane-embedded secretion factor of *E. coli*, is cleaved by the OmpT protease *in vitro*. Biochem Biophys Res Commun 167:711–715

Blinkowa AL, Walker JR (1990) Programmed ribosomal frameshifting generates the *Escherichia coli* DNA polymerase III γ subunit from within the τ subunit reading frame. Nucleic Acids Res 18:1725–1729

Brierley I, Digard P, Inglis SC (1989) Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. Cell 57:537–547

Casadaban MJ, Cohen SN (1980) Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J Mol Biol 138:179–207

Dinman JD, Icho T, Wickner RB (1991) A −1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gap-pol fusion protein. Proc Natl Acad Sci USA 88:174–178

Escoubas JM, Prère MF, Fayet O, Salvignol I, Galas D, Zerbib D, Chandler M (1991) Translational control of transposition activity of the bacterial insertion sequence IS*1*. EMBO J 10:705–712

Flower AM, McHenry CS (1990) The γ subunit of DNA polymerase III holoenzyme of *Escherichia coli* is produced by ribosomal frameshifting. Proc Natl Acad Sci USA 87:3713–3717

Gottesman S (1989) Genetics of proteolysis in *Escherichia coli*. Annu Rev Genet 23:163–198

Hashimoto-Gotoh T, Sekiguchi M (1977) Mutations to temperature sensitivity in R plasmid pSC101. J Bacteriol 131:405–412

Hizi A, Henderson LE, Copeland TD, Sowder RC, Hixson CV, Oroszlan S (1987) Characterization of mouse mammary tumor virus *gap-pro* gene products and the ribosomal frameshift site by protein sequencing. Proc Natl Acad Sci USA 84:7041–7045

Iida S, Arber W (1980) On the role of IS*1* in the formation of hybrids between the bacteriophage P1 and the R plasmid NR1. Mol Gen Genet 177:261–270

Inamoto S, Ohtsubo E (1990) Specific binding of TraY protein to *oriT* and the promoter region for the *traY* gene of plasmid R100. J Biol Chem 265:6461–6466

Jacks T, Townsley K, Varmus HE, Majors J (1987) Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary virus *gag*-related polyprotein. Proc Natl Acad Sci USA 82:2829–2833

Jacks T, Madhani HD, Masiarz FR, Varmus HE (1988) Signals for ribosomal frameshifting in the Rous sarcoma virus *gag-pol* region. Cell 55:447–458

Johnsrud L (1979) DNA sequence of the transposable element IS*1*. Mol Gen Genet 169:213–218

Kunkel TA, Roberts JD, Zakour RA (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol 154:367–382

Machida C, Machida Y (1989) Regulation of IS*1* transposition by the *insA* gene product. J Mol Biol 208:567–574

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular Cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Messing J (1983) New M13 vectors for cloning. Methods Enzymol 101:20–78

Miller JH (1972) Experiment in Molecular Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Moore R, Dixon M, Smith R, Peters G, Dickson C (1987) Complete nuleotide sequence of a milk-transmitted mouse mammary tumor virus: two frameshift suppression events are required for translation of *gag* and *pol*. J Virol 61:480–490

Ohtsubo E, Zenilman M, Ohtsubo H (1980) Plasmids containing insertion elements are potential transposons. Proc Natl Acad Sci USA 77:750–754

Ohtsubo E, Zenilman M, Ohtsubo H, McCormick M, Machida C, Machida Y (1981) Mechanism of insertion and cointegration mediated by IS*1* and Tn*3*. Cold Spring Harbor Symp Quant Biol 45:283–295

Ohtsubo H, Ohtsubo E (1978) Nucleotide sequence of an insertion element, IS*1*. Proc Natl Acad Sci USA 73:2316–2320

Rice NR, Stephans RM, Burny A, Gilden RV (1985) The *gag* and *pol* genes of Bovine leukemia virus: nucleotide sequence and analysis. Virology 142:357–377

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Sekine Y, Ohtsubo E (1989) Frameshifting is required for production of the transposase encoded by insertion sequence *1*. Proc Natl Acad Sci USA 86:4609–4613

Sekine Y, Ohtsubo E (1991) Translational frameshifting in IS elements and other genetic systems. In: Kimura M, Takahata N (ed) New Aspects of The Genetics of Molecular Evolution. Japan Scientific Societies Press, Tokyo/Springer-Verlag, Berlin, pp 243–261

Sekine Y, Ohtsubo E (1992) DNA sequences required for translational frameshifting in production of the transposase encoded by IS*1*. Mol Gen Genet 235:325–332

Shimotohno K, Takahashi Y, Shimizu N, Gojobori T, Golde D, Chen I, Miwa M, Sugimura T (1985) Complete nucleotide sequence of an infectious clone of human T-cell leukemia virus type II: an open reading frame for the protease gene. Proc Natl Acad Sci USA 82:3101–3105

Tsuchihashi Z, Kornberg A (1990) Translational frameshifting generates the γ subunit of DNA polymerase III holoenzyme. Proc Natl Acad Sci USA 87:2516–2520

Varmus H, Brown P (1989) Retroviruses. In: Berg DE, Howe MM (ed) Mobile DNA. American Society for Microbiology, Washington, DC, pp 53–108

Vrieira J, Messing J (1987) Production of single stranded plasmid DNA. Methods Enzymol 153:3–11

Vögele K, Schwartz E, Welz C, Schiltz E, Rak B (1991) High-level ribosomal frameshifting directs the synthesis of IS*150* gene products. Nucleic Acids Res 19:4377–4385

Yoshioka Y, Ohtsubo H, Ohtsubo E (1987) Repressor gene *finO* in plasmids R100 and F: constitutive transfer of plasmid F is caused by insertion of IS*3* into *finO*. J Bacteriol 169:619–623

Zerbib D, Polard P, Escoubas JM, Galas D, Chandler M (1990) The regulatory role of the IS*1*-encoded InsA protein in transposition. Mol Microbiol 4:471–477

Communicated by M. Sekiguchi