JOURNAL OF BACTERIOLOGY, Sept. 1994, p. 5864–5867
0021-9193/94/$04.00+0

Vol. 176, No. 18

Is the IS*1* Transposase, InsAB′, the Only IS*1*-Encoded Protein Required for Efficient Transposition?

J. M. ESCOUBAS,† D. LANE, AND M. CHANDLER*

Molecular Genetics and Microbiology, UPR9007 du Centre National de la Recherche Scientifique, 31062 Toulouse Cedex, France

Received 2 May 1994/Accepted 2 July 1994

The transposase of the bacterial insertion sequence IS*1* is normally expressed by inefficient translational frameshifting between an upstream reading frame which itself specifies a transposition inhibitor, InsA, and a second consecutive reading frame located immediately downstream. A fused-frame mutant which carries an additional base pair inserted at the point of frameshifting was constructed. This mutant exhibits high transposition activity and should express the transposase, InsAB′, constitutively without frameshifting. Unexpectedly, a second protein species was observed to be expressed from this mutant. We demonstrate here that this protein, InsA*, results from continued frameshifting on the modified frameshift motif. The protein retains the activities of the repressor InsA. Its elimination, by further modification of the frameshift motif, results in a further increase in various transposition activities of IS*1*. These results support the hypothesis that a single IS*1*-encoded protein, InsAB′, is necessary for transposition.

IS*1* is a small, genetically compact, insertion sequence present in the chromosomes of several enterobacterial species (10). It exhibits two partially overlapping open reading frames, *insA* and *insB′*, located respectively in relative reading phases 0 and −1, whose integrity is essential for transposition (5, 9). The product of the first, InsA, binds specifically to both left and right IS*1* terminal inverted repeats, IRL and IRR (16) (Fig. 1A), repressing its own transcription from a promoter located partially in IRL and simultaneously inhibiting transposition (8, 17). A second protein, InsAB′, carries domains contributed by both reading frames and is produced by inefficient translational frameshifting between *insA* and *insB′* (3, 13). Frameshifting has been shown to occur on an A_6C sequence (3, 13) (Fig. 1B), as found also in the synthesis of the Gag-Pol polyprotein in some retroviruses (for a review, see reference 15). The InsAB′ protein can be produced "constitutively" by introduction of a single additional A residue into the A_6C run to form A_7C (2, 13). This insertion mutation physically fuses *insA* and *insB′* and increases the frequency of IS*1* transposition by a factor of between 10^2 and 10^4 (2, 13). The mutation should eliminate expression of the InsA repressor since it places the InsA termination codon out of phase (Fig. 1C). The large stimulation in transposition frequency suggested that the InsAB′ fusion protein is the IS*1* transposase and led to the idea that it is the only IS*1*-encoded protein necessary for transposition.

It has been shown previously (3), by using a phage T7 expression system, that the A_7C frameshift mutant, carried by plasmid pMET11, specifies a protein of the size expected for InsAB′ and that this protein is expressed at a high level compared with that achieved by natural frameshifting which occurs in the wild-type IS*1* coding sequences carried by plasmid pZBT29 (3) (Fig. 2B, lanes 1 and 2). The natural frequency of translational frameshifting, estimated from measurements of the relative amounts of InsAB′ and InsA for the wild-type IS*1* coding sequences, was about 1% under these conditions. Since the InsA termination codon is placed out of phase in the frameshift mutant (Fig. 1C), only one protein species, InsAB′, should be produced. However, inspection of the results (Fig. 2B, lane 2) indicated the presence of a significant amount of a second protein species. This species had an apparent molecular mass of 10.7 kDa, slightly larger than that observed for InsA (as seen with plasmid pZBT29 [Fig. 2B, lane 1]). This raised the possibility that a third protein is normally made in small amounts from wild-type IS*1* and that its presence had been obscured in previous experiments by high levels of InsA. The unexplained presence of this additional protein brings into question the contention that InsAB′ alone is necessary for IS*1* transposition. It was therefore important to determine its origin and its possible effects on transposition activity.

Origin of the additional IS*1*-specified protein species. One possibility is that the additional protein reflects the occurrence of an alternative translational frameshift. Experiments with retroviral frameshift signals have demonstrated that several heptanucleotide sequences are capable of inducing −1 translational frameshifting (15). Among these is an A_7C signature similar to that created in the fused-frame mutant, pMET11. One explanation for the appearance of the 10.7-kDa protein from pMET11 is that it results from frameshifting on the A_7C motif. This explanation is supported by inspection of the nucleotide sequence (Fig. 1C). Following a −1 frameshift on the A_7C sequence, the ribosome would encounter a termination codon (UGA) approximately 16 codons downstream and generate a protein of 100 amino acids with a predicted molecular mass of 10.7 kDa. The resulting protein would carry the NH_2-terminal 85 amino acids of InsA, including the α-helix–turn–α-helix motif thought to be involved in recognizing and binding to the IS*1* terminal repeats (16) (Fig. 1D). To test this possibility, an IS*1* derivative in which the sequence GA_2GA_3C replaces the A_6C motif was generated by site-directed mutagenesis. Like the A_7C mutation, this also fuses the *insA* and *insB′* reading frames, but the GA_2GA_3C mutation disrupts the presumed alternative interactions between tRNAs and mRNA in the −1 reading phase (6, 7) and should reduce

* Corresponding author. Mailing address: Molecular Genetics and Microbiology, UPR9007 du CNRS, 118 Rte de Narbonne, 31062 Toulouse Cedex, France. Phone: (33) 613355858. Fax: (33) 61335886. Electronic mail address: mike@biotoul.ibcg.fr.

† Present address: Oceanographic and Atmospheric Science Division, Department of Applied Science, Brookhaven National Laboratory, Upton, NY 11973.

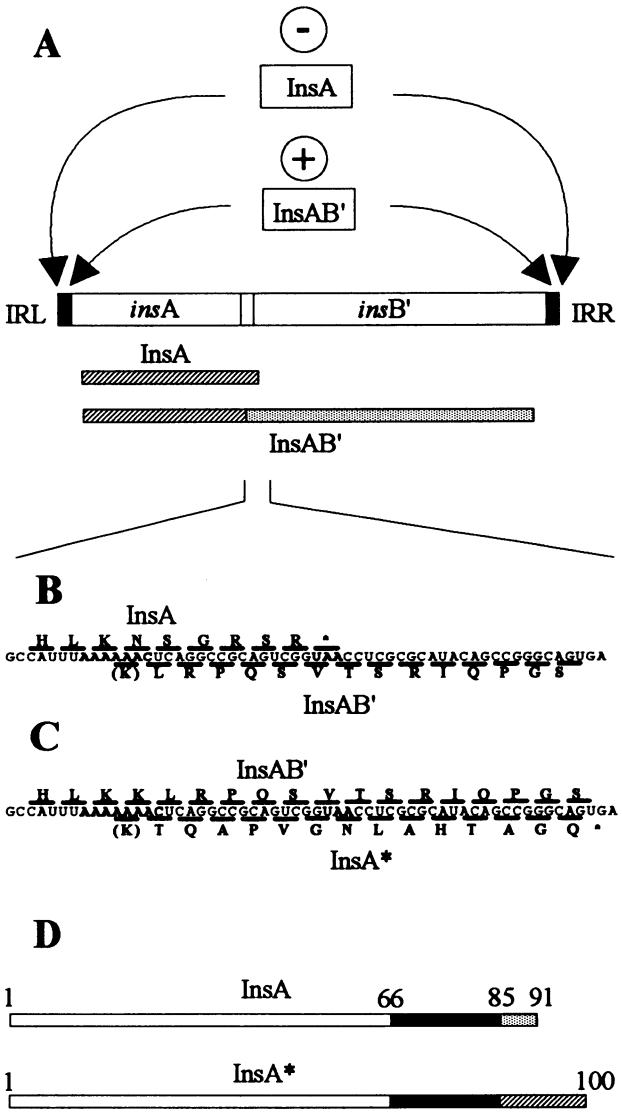

FIG. 1. Organization of IS*1*. (A) Structure of IS*1*. Left (IRL) and right (IRR) inverted terminal repeats are shown as solid boxes. Relative positions of the *insA* and *insB'* reading frames, together with their overlap region, are also shown within the open box representing IS*1*. The InsA protein is represented by a hatched box. The InsA and InsB' components of the InsAB' frameshift product are shown as hatched and stippled boxes, respectively. Arrows indicate the probable region of action of InsA and InsAB' proteins. (B and C) RNA and protein sequence in the crossover region between the two open reading frames. Codons shown above the RNA sequence show the product of direct translational readout. Those below show the product of a −1 translational frameshift. Presumed amino acids at the point of frameshift are shown in parentheses. In panel B, the heptanucleotide A_6C frameshift sequence involved in production of InsAB' from the wild-type IS*1* coding sequence (upper sequence) and the A_7C derivative (lower sequence) are indicated in boldface type as is the UAA termination codon for InsA. In panel C, the A_7C fused-frame sequence is shown in boldface type, as is the UGA termination codon for InsA*. (D) Schematic representation of InsA and InsA*. The solid boxes indicate the position of the potential helix-turn-helix motif in the InsA and InsA* proteins. The difference in the amino acid sequence of the carboxy-terminal end of the protein is distinguished by a stippled box and a hatched box for InsA and InsA*, respectively. The numbers represent amino acid residues.

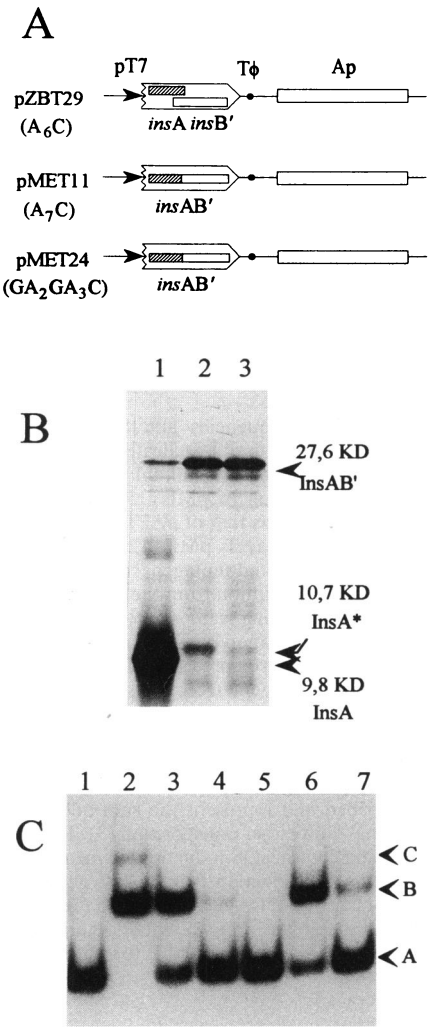

FIG. 2. IS*1* proteins produced in the phage T7 expression system. (A) Structure of the plasmids used. Mutant IS*1* coding sequences are indicated in parentheses below the plasmid name. The construction of pZBT29 and pMET11 has been described previously. Plasmid pMET24 was generated in the same way as pMET11, except that the oligonucleotide used in the site-directed mutagenesis carried the appropriate GA_2GA_3C mutation. Transcription from the phage T7 promoter (pT7) is from left to right. Open reading frames are indicated as hatched and open boxes within IS*1* (larger box). The pointed end of this box represents IRR. Tϕ indicates the presence of a phage T7 transcription terminator. The β-lactamase gene (Ap) is also shown as an open box. (B) Expression of IS*1* proteins in the T7 expression system. The methods used were described previously (3). Lanes: 1, pZBT29; 2, pMET11; 3, pMET24. (C) Gel retardation with crude-cell extracts. The methods used were described previously (18). DNA fragments were prepared by PCR with pMP20 (11) as a matrix. Lanes: 1, no extract; 2 and 3, pZBT29; 4 and 5, pMET24; 6 and 7, pMET11. Lanes 2, 4, and 6 show results obtained with a 1:10 dilution of the crude extracts. Lanes 3, 5, and 7 show results obtained with a dilution of 1:100. Band A, unbound DNA fragment; band B (and C), InsA (InsA*)-specific band.

or eliminate frameshifting over these nucleotides. The amino acid sequence of the InsAB' protein produced by this mutant should be identical to that produced from the A_7C frameshift mutant. The proteins produced from this plasmid, pMET24, in the phage T7 expression system are shown in Fig. 2B, lane 3. It

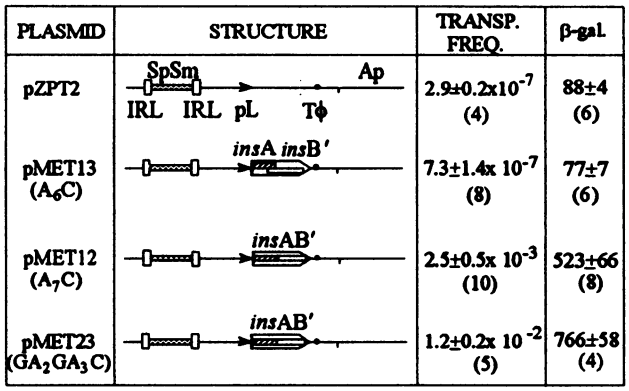

FIG. 3. Transposition frequencies and SOS-inducing capacities. The structure of the plasmids used is indicated on the left of the figure together with the wild-type and two mutant IS*1* coding sequences (in parentheses). The construction of pMET13 and pMET12 has been described previously (3), as has that of pZPT2 (17). Plasmid pMET23 was generated in the same way as pMET12 except that the oligonucleotide used in the site-directed mutagenesis carried the appropriate GA_2GA_3C mutation. Transcription from the phage λ promoter (p_L) is from right to left. Open reading frames are indicated as hatched and stippled boxes within IS*1* (larger box) The pointed end of this box represents IRR. Tφ indicates the presence of a phage T7 transcription terminator. The β-lactamase gene (Ap) is also shown. In addition, the presence of a synthetic IS*1*-based transposon, Ω-on, composed of inverted copies of IRL flanking a gene specifying resistance to streptomycin (Sm) and spectinomycin (Sp) is also shown. Transposition frequencies were determined by using a standard mating-out assay as previously described and represent the sum of the frequencies of cointegrate formation and direct transposition (3). The method used in determining the capacity for SOS induction by measuring the level of β-galactosidase (β-gal.) production from a *lacZ* gene driven from an SOS inducible promoter has also been described (7a). The numbers in parentheses show the number of independent determinations for each value. Standard errors were less than 10% for both transposition and β-galactosidase activities.

is clear that the mutation eliminates the 10.7-kDa species while maintaining constitutive InsAB′ synthesis, supporting the notion that the 10.7-kDa species is produced by frameshifting on the A_7C motif. The protein is called InsA* in the following discussion.

Effect of InsA* on intermolecular transposition activity. In previous experiments, transposition activity was assessed primarily for the wild-type A_6C coding sequences and the A_7C mutant. It was therefore important to determine the influence of InsA* on these results and the effect of its absence. In a first set of experiments, we have compared transposition frequencies obtained by using wild-type A_6C and the A_7C and GA_2GA_3C mutant IS*1* coding sequences in a standard mating-out assay (3). Plasmids pMET13 (A_6C), pMET12 (A_7C), and pMET23 (GA_2GA_3C) carry the IS*1* coding sequences under control of the phage λ p_L promoter. A single IRR is located at its natural position at the 3′ end of *insB′*. In addition, the plasmids carry an artificial IS*1*-based transposon, Ω-on, composed of an Sp^r Sm^r cassette flanked by inverted copies of IRL (11). As discussed elsewhere (3), use of the p_L promoter and elimination of the endogenous IRL promoter and InsA-binding site eliminate any InsA repression of transposase expression. The results of mating-out assays, by using the conjugative plasmid pOX38Km (1) as a recipient replicon, are presented in Fig. 3. As expected, use of the A_7C mutation results in a substantial increase in overall transposition frequency compared with the wild-type A_6C sequence. Furthermore, substitution of GA_2GA_3C for the A_7 mutation results in a further fourfold increase to a transposition frequency of 1%. These results suggest that InsA*, produced by the A_7C mutant, is capable of inhibiting intermolecular transposition.

Effect of InsA* on induction of an SOS signal. Another property associated with the IS*1* transposase, like that of IS*10* (4, 12), is its capacity, in the presence of suitable transposon ends, to induce the host SOS response (7a). This presumably results from the endonucleolytic activity associated with the transposase. To determine whether InsA* influences the ability of InsAB′ to induce the SOS response, the activity of pMET13, pMET12, and pMET23 was determined in a standard SOS induction assay. Briefly, the assay uses an SOS indicator strain in which a *lacZ* gene is driven by the p_L promoter of a defective lambdoid 434 prophage (2). Induction of the SOS response results in cleavage of the 434 repressor and consequent induction of β-galactosidase synthesis. The results obtained with the three plasmids by using a standard assay (7) are presented in Fig. 3. SOS induction by pMET13 is low. A significant increase is observed with pMET12, while pMET23 gives a further stimulation in SOS activity. Thus, as in the case of intermolecular transposition, InsA* interferes with transposase activity.

Sequence-specific DNA binding. The results presented above are consistent with the idea that the effects of InsA*, like those of InsA, are due to its capacity to bind to the ends of IS*1*. To determine the nucleic acid-binding properties of InsA*, extracts of cultures containing high levels of IS*1*-derived proteins were used in gel retardation assays with labeled DNA fragments of 85 bp carrying 28 bp of IRL (18). The results are presented in Fig. 2C. It can be seen that, while extracts derived from strains carrying the A_7C mutant (pMET11) exhibit a distinct retarded band (lanes 6 and 7), those derived from cells carrying the GA_2GA_3C mutant (pMET24) showed little retardation activity (lanes 4 and 5). This suggests that InsA* binding is responsible for the retarded species. It is interesting that although InsAB′ should also carry the InsA-specific nucleic acid-binding domain, we could not observe specific complexes of higher apparent molecular mass expected to arise from InsAB′ binding. This inability to detect efficient InsAB′ binding to the terminal inverted repeats is under investigation. The electrophoretic mobility of the retarded band obtained in the presence of InsA* is similar to that obtained with an extract enriched for InsA itself (plasmid pZB29; lanes 2 and 3). The InsA* protein may therefore behave in a similar way to InsA itself in influencing IS*1* transposition.

The results presented here provide evidence that a significant level of translational frameshifting occurs on the mutant A_7 heptamer derived from the A_6(C) wild-type IS*1* sequence and designed to physically fuse the *insA* and *insB′* reading frames of IS*1*. This mutation results in high-level synthesis of the fusion protein InsAB′ without requiring a −1 translational frameshift. Frameshifting on the mutant A_7 produces a small protein, InsA*, related to the IS*1* transposition inhibitor/repressor, InsA. The level of frameshifting on the A_7 heptamer cannot be assessed accurately from our present data. Estimates from the intensities of the relevant protein species from autoradiographs such as that shown in Fig. 2B suggest that it might be as high as 55% (compared with less than 1% on the wild-type A_6C motif) (3). However, this value is certainly an overestimate, since it has been demonstrated that, at least in vitro, the T7 RNA polymerase specifically amplifies frameshifting on an A_7 sequence (14). We have not attempted to assess the "normal" in vivo rate by more reliable techniques.

Inclusion of the GA_2GA_3C sequence in place of A_7C

eliminates InsA*, confirming that translational frameshifting occurs on the A_7 heptamer. InsA* has a molecular mass slightly larger than that of InsA itself on sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Calculation of the predicted molecular mass of InsA* gives a value of 10.7 kDa compared with 9.8 kDa for InsA. The InsA* protein carries sufficient information to bind specifically to the left terminal inverted repeat of IS*1*. Since the amino acid sequences of InsA and InsA* presumably diverge at amino acid 86 (Fig. 1B to D), the specific DNA sequence recognition domain must be included within the first 85 amino acids of the proteins. We have also provided evidence which suggests that InsA*, like InsA, inhibits IS*1* transposition. This inhibitory activity of InsA* is also illustrated by its ability to reduce the transposase-mediated induction of the SOS response. The most important aspect of these results is that elimination of InsA* production, by use of the GA_2GA_3C motif, leads to an increase in the various activities of InsAB′. The data thus further support the idea that InsAB′ is the IS*1* transposase and is the only IS*1*-encoded protein required for efficient transposition.

We thank all members of the mobile genetic elements group for discussions. We are also grateful to Pierre Prentki for providing the initial stimulus for this analysis.

This work was supported by CNRS grant UPR9007 and NATO travel grant 0.1669.

REFERENCES

1. **Chandler, M., and D. J. Galas.** 1983. Cointegrate formation by Tn9: activity of IS1 is modulated by external sequences. J. Mol. Biol. **170:**61–91.
2. **Elespuru, R. K., and M. B. Yarmolinsky.** 1979. A colorimetric assay of lysogenic induction designed for screening potential carcinogenic and carcinostatic agents. Environ. Mutagen. **1:**65–78.
3. **Escoubas, J. M., M.-F. Prère, O. Fayet, I. Salvignol, D. J. Galas, D. Zerbib, and M. Chandler.** 1991. Translational control of transposition activity of the bacterial insertion sequence IS*1*. EMBO J. **10:**705–712.
4. **Haniford, D. B., A. R. Chelouche, and N. Kleckner.** 1989. A specific class of IS*10* transposase mutants are blocked for target site interactions and promote formation of an excised transposon fragment. Cell **59:**385–394.
5. **Jackowec, M., P. Prentki, M. Chandler, and D. J. Galas.** 1988. Mutational analysis of the open reading frames in the transposable element IS*1*. Genetics **120:**47–55.
6. **Jacks, T., M. D. Power, F. R. Masiarz, P. A. Luciw, P. J. Barr, and H. E. Varmus.** 1988. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. Nature (London) **331:**280–283.
7. **Jacks, T., K. Townsley, H. E. Varmus, and J. Majors.** 1987. Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus gag-related polyproteins. Proc. Natl. Acad. Sci. USA **84:**4298–4302.

7a. **Lane, D., J. Cavaillé, and M. Chandler.** J. Mol. Biol., in press.

8. **Machida, C., and Y. Machida.** 1989. Regulation of IS*1* transposition by the *ins*A gene product. J. Mol. Biol. **208:**567–574.
9. **Machida, Y., C. Machida, and E. Ohtsubo.** 1984. Insertion element IS1 encodes two structural genes required for its transposition. J. Mol. Biol. **177:**229–245.
10. **Nyman, K., H. Ohtsubo, D. Davison, and E. Ohtsubo.** 1983. Distribution of insertion element IS*1* in natural isolates of *Escherichia coli*. Mol. Gen. Genet. **189:**516–518.
11. **Prentki, P., M.-H. Pham, P. Gamas, M. Chandler, and D. Galas.** 1987. Artificial transposable elements in the study of the ends of IS*1*. Gene **61:**91–101.
12. **Roberts, D., and N. Kleckner.** 1988. Tn*10* transposition promotes RecA-dependent onduction of λ prophage. Proc. Natl. Acad. Sci. USA **85:**6037–6041.
13. **Sekine, Y., and E. Ohtsubo.** 1989. Frameshifting is required for production of the transposase encoded by insertion sequence 1. Proc. Natl. Acad. Sci. USA **86:**4609–4613.
14. **Tsuchihashi, Z., and P. O. Brown.** 1992. Sequence requirements for efficient translational frameshifting in the Escherichia coli *dna*X gene and the role of an unstable interaction between $tRNA^{Lys}$ and an AAG lysine codon. Genes Dev. **6:**511–519.
15. **Weiss, R. B., D. M. Dunn, M. Shuh, J. F. Atkins, and R. F. Gesteland.** 1989. E. coli ribosomes rephase on retroviral frameshift signals at rates ranging from 2 to 50%. New Biol. **1:**159–169.
16. **Zerbib, D., M. Jakowec, P. Prentki, D. Galas, and M. Chandler.** 1987. Expression of proteins essential for IS*1* transposition: specific binding of InsA to the ends of IS*1*. EMBO J. **6:**3163–3169.
17. **Zerbib, D., P. Polard, J.-M. Escoubas, D. Galas, and M. Chandler.** 1990. The regulatory role of the IS*1*-encoded InsA protein in transposition. Mol. Microbiol. **4:**471–477.
18. **Zerbib, D., P. Prentki, P. Gamas, E. Freund, D. J. Galas, and M. Chandler.** 1990. Functional organization of the ends of IS*1*: specific binding site for an IS*1*-encoded protein. Mol. Microbiol. **4:**1477–1486.