Nucleic Acids Research, Vol. 19, No. 16 4377–4385

High-level ribosomal frameshifting directs the synthesis of IS*150* gene products

Karl Vögele, Edward Schwartz+, Caroline Welz, Emile Schiltz[1] and Bodo Rak*
Institut für Biologie III, Universität Freiburg, Schänzlestraße 1, D-7800 Freiburg and [1]Institut für Organische Chemie und Biochemie, Universität Freiburg, Albertstraße 21, D-7800 Freiburg, FRG

Received June 12, 1991; Revised and Accepted July 19, 1991

ABSTRACT

IS*150* contains two tandem, out-of-phase, overlapping genes, *ins150A* and *ins150B*, which are controlled by the same promoter. These genes encode proteins of 19 and 31kD, respectively. A third protein of 49kD is a transframe gene product consisting of domains encoded by both genes. Specific −1 ribosomal frameshifting is responsible for the synthesis of the large protein. Expression of *ins150B* also involves frameshifting. The IS*150* frameshifting signals operate with a remarkably high efficiency, causing about one third of the ribosomes to switch frame. All of the signals required for this process are encoded in a 83-bp segment of the element. The heptanucleotide A AAA AAG and a potential stem-loop-forming sequence mark the frameshifting site. Similar sequence elements are found in −1 frameshifting regions of bacterial and retroviral genes. A mutation within the stem-loop sequence reduces the rate of frameshifting by about 80%. Artificial transposons carrying this mutation transpose at a normal frequency, but form cointegrates at a ≈100-fold reduced rate.

INTRODUCTION

IS*150* is a 1443 base pair (bp) transposable element which inhabits the genomes of many strains of *Escherichia coli* (1,2). IS*150* belongs to a family of structurally related insertion sequences (1,3) with representatives in such diverse bacterial genera as *Shigella*, *Rhizobium*, *Xanthomonas*, *Neisseria*, *Streptococcus* and *Mycobacterium*, as well as in *Mycoplasma* (4−10). The mobile elements of this group have been implicated in phenomena of economic and medical importance such as tomato and pepper spot disease (6) and the synthesis of capsule polysaccharides in pathogenic bacteria (7). IS*150* can, like other members of the group, act as a mobile promoter, turning on the expression of genetic functions adjacent to insertion sites (11,12). The basis of this property is an outward-reading promoter located near the defined right end of the element. Another notable feature of IS*150* is its high target specificity. The only two known target sites share a common 10-bp core sequence (2).

Characteristic of the elements of the IS*150* family is the arrangement of tandem out-of-phase open reading frames on one strand of the DNA. Analysis of the amino acid sequences deduced from the open reading frames of this group of elements showed that they share a conserved sequence motif found in retroviral integrases (13,14). Recent studies suggest that this motif is essential for integrase binding (14). Interestingly, the elements of the IS*150* family have the same terminal nucleotides (TG.....CA) as retroviruses (15).

IS*150* contains three open reading frames which could code for proteins of 19.7, 33.3, and 14kD (1). Studies with gene fusions indicated that all three ORFs have coding function (E.Schwartz, C.Welz, and B.Rak, unpublished). The corresponding genes were designated *ins150A*, *ins150B*, and *ins150C*. In the course of preliminary expression studies we identified two IS*150*-encoded proteins which corresponded to the sizes predicted for the *ins150A* and *ins150B* gene products. However, we also observed a third protein, which clearly exceeded the coding capacity of any single IS*150* gene.

We now report the results of expression studies which permit the unambiguous identification of the gene products of both genes. Sequencing of a derivative of the large protein and of the corresponding IS*150* transcript shows that it is a transframe gene product produced by translational frameshifting. The frameshifting process operates at a rate exceeding 30%. We show that the signals mediating this process are encoded in a segment of IS*150* between bp 521−603. Moreover, we demonstrate that an inverted repeat located in this region plays an essential role in frameshifting. IS*150* gene products are not required for frameshifting. Furthermore, we present evidence indicating that the production of the *ins150B* gene product also requires translational frameshifting. Our data suggest that alteration of the rate of frameshifting affects the transpositional activity of IS*150*. Thus it seems that the balance of IS*150* gene products and not their absolute levels controls transpositional activity.

MATERIALS AND METHODS

Bacterial strains

Escherichia coli K-12 minicell producing strain 312A (=R140) has been described (16). *E.coli* K-12 strain CSH26 (=R1016) has

* To whom correspondence should be addressed

+ Present address: Institut für Pflanzenphysiologie und Mikrobiologie, Freie Universität Berlin, Königin-Luisen-Straße 12−16, D-1000 Berlin 33, FRG

the genotype *ara* Δ(lac-proAB) *thi* (17). *E.coli* K-12 strain WH207 (=R1306) is a *galK* ΔlacX74 *recA E.coli* K-12 derivative obtained from A.Wissmann. *Salmonella typhimurium* strain R954 is the prototrophic type strain LT2. R1115 is a *recA-1* derivative of R954. SL5235 (=R1403) is a restriction-negative modification-positive derivative of *S.typhimurium* used as a shuttle host to transfer plasmid DNA from *E.coli* to strains of *S.typhimurium* (18).

Media for bacterial growth

Bacteria were grown in LB or M9 medium (17). For plates 15g agar/l was added. Where necessary, media were supplemented with ampicillin (amp; 50μg/ml), kanamycin (kan; 50μg/ml) or chloramphenicol (cm; 30μg/ml).

Plasmid construction and site-specific mutagenesis

Plasmids were constructed using standard recombinant techniques (19). Where necessary, staggered ends were made blunt by treatment with DNA polymerase (Klenow fragment) before ligation. Plasmid structures were verified by restriction analysis or sequencing. Oligonucleotide-directed site-specific mutagenesis was performed using a combination of standard techniques (20,21). Mutagenized fragments were verified by sequencing prior to subcloning. The relevant structures of all plasmids used in this study are given in the Figures. A detailed description of their construction is available upon request. The following plasmids provided the basic structures: pBR322 (22), pBR325 (23), pUC13 (24), pFD24 (16), pFR200 (25), pFR200::IS*150-1* (1), pFDX500 (26), pT7-7 (27), pKK177-3 (28), pHSG415 (29), and pFDX600 (F.Brombacher and B.Rak, unpublished results). Plasmid pFDX600 is essentially a deletion derivative of the thermosensitive plasmid pHSG415; details of its construction will be published elsewhere.

Protein labelling and separation

Minicells were prepared from transformants of strain 312A, incubated in [^{35}S] methionine and the labelled proteins separated on polyacrylamide-SDS gels as described previously (16). For in vivo labelling transformants of strain CSH26 were grown in M9 supplemented with Difco 1×methionine assay (2%), proline (20μg/ml), vitamin B1 (1μg/ml), glucose (0.5%) and the appropriate antibiotics. Cells were grown to an OD_{600} of 0.2 and harvested by centrifugation (500-μl aliquots). The pellet was resuspended in 200μl of the same medium and labelled for 1 min. with 20μCi [^{35}S]-methionine (1300Ci/mMol). Where appropriate a 2.5 min. preinduction period with 1 mM IPTG preceded addition of label. Labelling was terminated by the addition of unlabelled methionine (200μg/ml). In pulse-chase experiments a 2.5 min. labelling was followed by a 25 min. incubation in an excess of nonradioactive methionine. Cells were lysed and the labelled proteins separated beside prestained molecular weight standards (BRL) on polyacrylamide-SDS gels (30).

Protein purification and sequencing

Strain WH207 harboring plasmids pFDX500 and pFDX1649 was grown in M9 medium supplemented with glucose (0.5%), Difco casamino acids (0.66%), vitamin B1 (1μg/ml), amp (50μg/ml) and kan (50μg/ml) to an OD_{600} of 0.8. IPTG was added to a final concentration of 1mM and the culture was incubated for another 2.5 h. The cells were then harvested, disrupted in a French pressure cell and fractionated by differential centrifugation (31). Subsequently, the extract was enriched by differential ammonium sulfate precipitation (32). Finally the protein was purified on a Mono-Q FPLC anion exchange column (Pharmacia LKB Biotechnology) according to the manufacturer's recommendations. The purified protein (50μg) was sequenced in an Applied Biosystems model 477A pulsed-liquid gas-phase sequencer equipped for on-line identification of amino acid derivatives.

Densitometry

Autoradiograms were digitized with a Cybertech CB-1 video camera. The digitized images were processed using the software package supplied by the same company. The values obtained were corrected for the distribution of methionines present in the various proteins.

Indirect RNA sequencing and comparison with the DNA template sequence

Transformants of strain R954 were grown under the same conditions as described for protein purification. RNA was isolated by a 'hot phenol' method described previously (18). First strand cDNA synthesis was carried out with 100μg total RNA using 1 pMol of a synthetic primer (5'-GCTGTCAAAAGACCGAC-GC-3'; IS*150* bp 458−476) and 100 units of AMV reverse transcriptase (Pharmacia LKB Biotechnology) as recommended by the manufacturer. PCR amplification was driven by 20 pmol of each of the above primer and the M13 standard primer in the presence of 1 unit Taq DNA polymerase (Amersham). After 30 PCR cycles (denaturation: 30 sec. at 94°C, annealing: 30 sec. at 53°C, elongation: 1 min. at 72°C) the reaction mixture was treated with RNAse A and loaded directly onto a 2% agarose gel. The PCR fragment was eluted and sequenced (33) using a T7 sequencing kit (Pharmacia LKB Biotechnology). Control reactions were done omitting reverse transcriptase and with RNAse-treated RNA.

Determination of enzyme activity

β-galactosidase activity was assayed according to the standard procedure (17) from transformants of strain CSH26 grown under the same conditions as for in vivo protein labelling (see above). Cells were lysed by the addition of SDS and chloroform. Where appropriate, IPTG was added to a concentration of 1mM 30 min. prior to lysis.

Transposition assay

Transformants harboring thermosensitive plasmids with the transposons to be tested were inoculated to a density of 10−100 cells/ml in LB supplemented with kan and cm and grown under aeration at 30°C for 24 hrs. Titers were determined by plating dilutions of the stationary-phase cultures on LB kan cm agar and incubating at 30°C. Dilutions were spread on prewarmed (42°C) LB kan and LB cm agar and incubated at 42°C. Colonies were counted and then replica-plated onto LB cm and LB kan agar, respectively, to determine the fraction of double resistant survivors at the non-permissive temperature. The rates of transposition and cointegration are estimated as the ratios of single and double resistant survivors at 42°C, respectively, to double resistant colony formers at 30°C.

RESULTS

Genes *ins150A* and *ins150B* encode three proteins

IS*150* contains three genes: *ins150A* and *ins150B* are coded in tandem, overlapping, out-of-phase ORFs on one strand of the

element; *ins150C* is located on the opposite strand (1). In order to identify IS*150* gene products we introduced an IS*150*-carrying plasmid (pFDX450; Fig. 1) into a minicell-producing strain of *E.coli* and monitored plasmid-directed protein synthesis by labelling with radioactive methionine. Three faint proteins with apparent molecular masses of approx. 19kd, 31kd and 49kd were observed (Fig. 2A; compare lanes 1 and 2). The 19kD and 31kD proteins correlate in size with the hypothetical products of genes *ins150A* (19.7kD) and *ins150B* (33kD), respectively. The third protein with an apparent molecular mass of approximately 49kD is much larger than any hypothetical IS*150* gene product. The synthesis of all three proteins was independent of the orientation of the IS*150*-containing insert in the vector, excluding their formation from adventitious reading frames (data not shown).

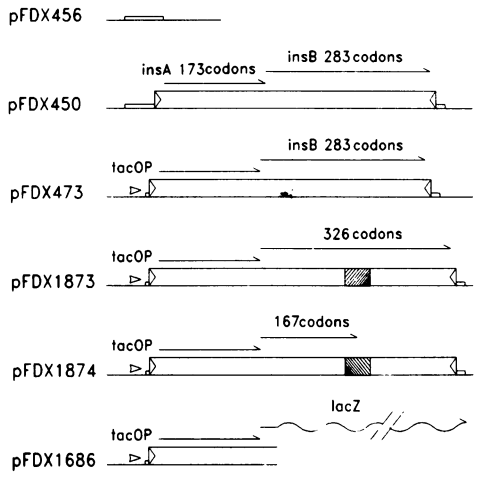


Figure 1. Relevant structures of plasmids used for in vitro and in vivo studies of IS*150* directed gene expression. The thin open bar represents the truncated copy of IS1 into which IS*150* is inserted (11). The inverted repeats of IS*150* are denoted by open arrowheads within the thicker open bar. Genes *ins150A* and *ins150B* are marked by arrows above the bar. A shaded box indicates a 129-bp insertion into gene *ins150B* which in one orientation enlarges the open reading frame by 43 codons (pFDX1873) and in the other, leads to premature termination (pFDX1874). The wavy line denotes gene *lacZ*. All structures are drawn to scale.

To facilitate further investigations we positioned the strong synthetic *tac* promoter 5′ of the left terminus of IS*150* and thus upstream of genes *ins150A* and *ins150B*. The pattern of proteins synthesized by the resulting plasmid (pFDX473; Fig. 1) was, in the absence of the inducer IPTG (isopropyl-β-d-thiogalactoside), indistinguishable from that observed for the predecessor plasmid pFDX450 (Fig. 2A, lane 3). Induction of the *tac* promoter by IPTG led to a marked increase in the synthesis of all three proteins (Fig. 2A, lane 4). This suggests that not only the genes for the 19kD and 31kD proteins but also the gene for the 49kD species lie downstream of the *tac* promoter. In order to confirm the assignment of the 19kD and 31kD proteins and to clarify the origin of the 49kD product we constructed a pair of plasmids in which a 129 bp fragment of the *lacZ* gene was inserted in both orientations into the *ins150B* gene. In one orientation (pFDX1873; Fig. 1) the inserted fragment adds 43 codons to *ins150B*, preserving the reading frame and resulting in a 326-codon long *insB* fusion gene. In the other orientation (pFDX1874; Fig.1) a stop codon is introduced, shortening gene *ins150B* to a length of 167 codons. The first modification should give rise to an enlarged product of approx. 38kD, the second, to a truncated protein of approx. 20.5kD. Proteins of the predicted sizes did indeed appear in place of the 31kD protein confirming its assignment as the product of gene *ins150B* (Fig. 2A, lanes 5 and 6). Furthermore, concomitant and equal shifts were observed for the 49kD protein, indicating that this gene product is, in part, encoded by the same ORF. The apparent sizes of the original and tailored proteins argue for the assignment of the 49kD protein to a composite protein incorporating the *ins150A* and *ins150B* ORFs. This is compatible with the coupled expression of the three proteins when under the control of the *tac* promoter. Pulse-chase experiments gave no indication that the three proteins are subject to specific proteolytic processing (data not shown). The ratio of the three proteins (InsA:InsB:InsAB) as determined by densitometric measurement of the autoradiogram was about 2:1:1.

In another series of experiments we studied the expression of IS*150*-encoded proteins *in vivo* (Fig. 2B). Transformants harboring a wild type copy of IS*150* downstream of the *tac* promoter (pFDX473) produced proteins of the apparent molecular

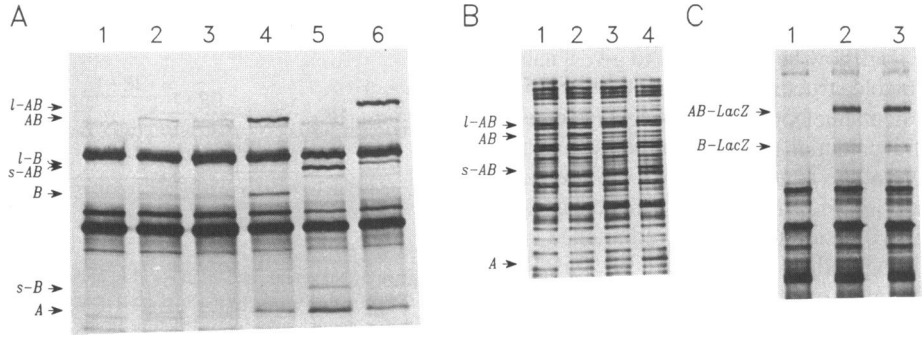


Figure 2. IS*150*-encoded proteins expressed in minicells and *in vivo*. Autoradiograms of [^{35}S]-labelled proteins separated in SDS-polyacrylamide gels. Panel A: labelling in minicells (11% gels). Lane 1: Plasmid pFDX456; Lane 2: plasmid pFDX450; Lane 3: plasmid pFDX473 without induction (−IPTG); Lane 4: plasmid pFDX473 with induction (+IPTG); Lane 5: plasmid 1874 (+IPTG); Lane 6: plasmid pFDX1873 (+IPTG). IS*150* encoded proteins are marked by arrows and labelled as follows: *A*: Ins150A; *B*: Ins150B; *AB*: Ins150AB; *s-B* and *s-AB*: the shortened Ins150B and Ins150AB proteins encoded by plasmid pFDX1874; *l-B* and *l-AB*: the enlarged hybrid Ins150B and Ins150AB proteins encoded by plasmid pFDX1873. Panel B: Pulse labelling in vivo (10% gels). Lane 1: plasmid pFDX456; Lane 2: plasmid pFDX473; Lane 3: plasmid pFDX1873; Lane 4 plasmid pFDX1874. Arrows are labelled as in panel A. Proteins Ins150AB (*AB*) and the enlarged hybrid Ins150AB protein (*l-AB*) are partially hidden by host proteins but are recognized by the more intensive staining. Panel C: Expression of β-galactosidase fusion proteins (6.5% gels). Lane 1: plasmid pFDX456; Lanes 2 and 3: plasmid pFDX1686; Lane 2: pulse; Lane 3: pulse-chase. *B-lacZ*: Ins150B-LacZ fusion protein; *AB-LacZ*: Ins150AB-LacZ fusion protein.

masses of Ins150A and Ins150AB when induced with IPTG (Fig. 2B; compare lanes 1 and 2). A protein corresponding to the *ins150B* gene product was, however, not detectable. Likewise, transformants harboring gene *ins150B* enlarged by the in-frame insert (pFDX1873; Fig. 1) made Ins150A and an enlarged Ins150AB but no enlarged Ins150B was visible (Fig. 2B, lane 3). Transformants carrying the *ins150B* gene abbreviated by the introduction of a premature stop codon (pFDX1874; Fig. 1) again produced the normal Ins150A protein and, in addition, a protein corresponding in size to the artificially shortened *ins150AB* reading frame (Fig. 2B, lane 4). A truncated Ins150B was, however, not found. Densitometric scanning of the autoradiograms showed that the ratio of Ins150A to Ins150AB protein and its modified species was approx. 2:1. This demonstrates that the mechanism responsible for frame coupling operates in vivo as efficiently as in minicells.

In none of the above in vivo experiments was an Ins150B protein detectable. In another attempt to identify an *ins150B* gene product in vivo we constructed a plasmid on which the *ins150B* gene was fused in frame to *lacZ* (pFDX1686; Fig. 1). Transformants expressing this *ins150B-lacZ* gene fusion produced two high-molecular-weight proteins (Fig. 2C, compare lanes 1 and 2). The smaller protein was similar in size to the authentic LacZ protein (not shown) and thus had an apparent molecular mass expected for an Ins150B-LacZ protein. The apparent molecular mass of the larger protein correlated to a gene product incorporating *ins150A* as well as *ins150B* and *lacZ* moieties. Pulse-chase experiments showed that both LacZ fusion proteins are stable under the conditions tested (Fig. 2C, lanes 2 and 3). On the basis of densitometric scanning the ratio of the two proteins (Ins150AB-LacZ:Ins150B-lacZ) was about 5:1. This indicates that the expression of Ins150B is weaker in normal cells than in minicells, where the ratio of Ins150AB:Ins150B expression was 1:1 (see above). The low-level expression of Ins150B explains the failure to detect its synthesis in the preceding experiment.

Translational frameshifting is responsible for synthesis of the InsAB fusion protein

The expression studies reported above demonstrated that three proteins of 19kD, 31kD and 49kD are expressed from overlapping out-of-phase genes *ins150A* and *ins150B*. The expression of the large protein from these genes indicates that their reading frames are somehow being coupled to give a natural fusion protein. Some notable structural features of the *ins150A-ins150B* junction—the heptanucleotide AAAAAAG located 6 nucleotide upstream of a potential hairpin-forming sequence—are strikingly reminiscent of the contexts of retroviral frameshift sites (41). Although the observed structural features clearly link expression of the Ins150AB fusion protein to well-studied cases of ribosomal frameshifting, other mechanisms such as RNA splicing, RNA editing and ribosomal slippage cannot be excluded a priori. Moreover, in none of the cases of alleged frameshifting reported so far has the removal of a single base from the mRNA by an editing mechanism been experimentally excluded. In order to obtain decisive evidence for frameshifting in IS*150* we decided to determine both the amino acid and the RNA sequence in the protein and transcript regions respectively, corresponding to the *ins150A/ins150B* junction. To this end we engineered a gene for a truncated fusion protein, thus bringing the region in question into easy reach of the amino acid sequencing procedure, and outfitted it with strong transcription and translation start signals.

To monitor the success of each stage of the work and to facilitate subsequent protein purification we based the construction on the *ins150B-lacZ* gene fusion (pFDX1686; see above) rather than on the wild type *ins150B*. We made two major alterations: (i) we deleted the first 93 bp of IS*150* thereby eliminating the *ins150A* translational start (pFDX1648; Fig. 4); (ii) we employed site-directed mutagenesis to introduce a Shine−Dalgarno sequence 7bp upstream of a methionine codon, which is the fifteenth-to-last codon of *ins150A* (pFDX1649; Fig. 4; see also Fig. 3). Comparison of the enzyme levels synthesized by the isogenic plasmids with and without the artificial Shine−Dalgarno sequence (pFDX1649 and pFDX1648; Fig. 4) demonstrated that more than 90% of the β-galactosidase activity synthesized by the former plasmid is attributable to the new ribosome binding signal, which directs high-level expression of a truncated Ins150A'B'-LacZ fusion protein (Fig. 4). Using the plasmid described above (pFDX1649) we purified the InsA'B'-LacZ fusion protein by two rounds of precipitation with ammonium sulfate followed by FPLC. The consecutive stages of purification were documented by electrophoresis in SDS-polyacrylamide gels (Fig.5). The N-terminal amino acid sequence of the purified fusion protein was determined by Edman degradation (Fig. 3). The first five residues of this sequence (MetTyrLeuLysLys) are identical to the deduced amino acid sequence of Ins150A'. The following residues, however, (AlaGluSerLeuSerSerHisGluLys-Val) match the sequence predicted for a protein encoded in the −1 frame. To obtain the RNA sequence in the critical region of the *ins150A/ins150B* transcript, we transformed *Salmonella typhimurium* LT2 (a strain devoid of IS*150* sequences; (1)) with

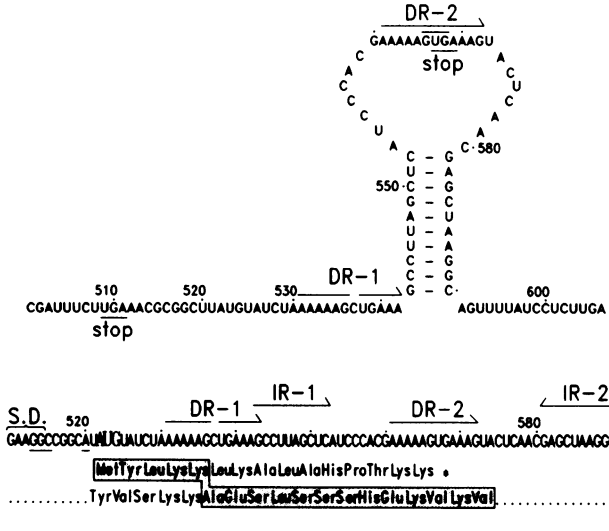


Figure 3. Nucleotide and amino acid sequences of the IS*150* frameshifting region. *IS150* coordinates are given above the sequences. Upper panel: Sequence of mRNA with a potential stem-loop structure in the region where *ins150A* and *ins150B* overlap. A UGA stop codon (bp 509−511), which is in phase with *ins150B* and thus delimits the left border of this ORF is indicated. The first and second units of a direct repeat (DR-1 and DR-2) are indicated by arrows. A hypothetical start codon of *ins150B* (GUG) and the overlapping stop codon of *ins150A* (UGA) centered in the loop are marked. Lower panel: Linear representation of the same region. Base changes introduced by site-specific mutagenesis are underlined and the Shine-Dalgarno (S.D.) sequence thus created is bracketed. Also marked are the direct and inverted repeats (DR-1, DR-2, IR-1 and IR-2, respectively). Below the RNA sequence the amino acid sequences deduced from the open reading frames of *ins150A* (top) and *ins150B* (bottom) are given together with the amino acid sequence determined by Edman degradation (shaded box).

plasmid pFDX1649 (Fig. 4) and isolated total RNA from the transformant. Linear DNA was prepared from the RNA by reverse transcription and subsequent PCR amplification of the region of the *ins150A/ins150B* junction. The sequence determined in this way was identical to that determined for an isolated DNA fragment containing the corresponding region (Fig. 6). Moreover, a significant background sequence indicative of RNA editing was absent. We conclude that specific and efficient ribosomal frameshifting is responsible for the formation of the Ins150A'B'-LacZ fusion protein and hence for the authentic Ins150AB fusion proteins.

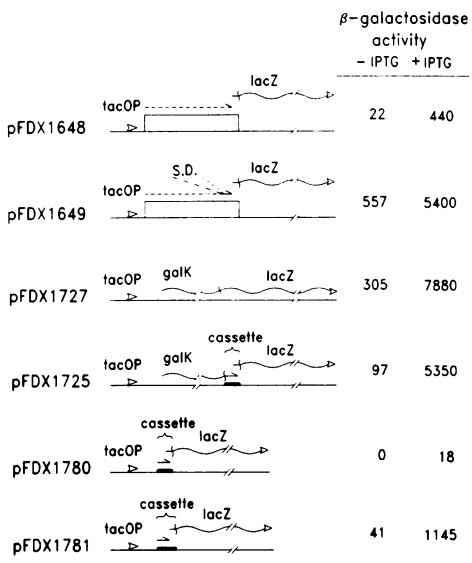


Figure 4. Relevant structures of plasmids constructed for overproduction and purification of an InsA'B-LacZ fusion protein and for investigation of the frameshift signals (left) and the β-galactosidase activity (in units) synthesized by the various plasmids with and without induction of the *tacOP* (+,−IPTG) (right). IS*150* sequences are indicated by an open box. The dashed line marks the 3' part of *ins150A*. S.D.: Shine−Dalgarno sequence (see Fig. 3). The IS*150*-derived cassette in the last three plasmids is given as a solid bar. All other symbols are as in Fig. 1. The structures are drawn to scale.

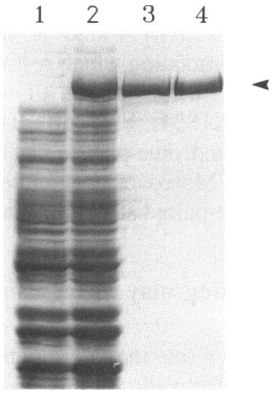


Figure 5. Purification of the Ins150A'B-LacZ protein for amino acid sequencing. Coomassie-stained SDS-polyacrylamide gels. Lane 1: total protein of *E.coli* strain WH207; Lane 2: Total protein of strain WH207/pFDX1649 after induction with IPTG; Lane 3: Extract shown in lane 2 fractionated by ammonium sulfate precipitation; Lane 4: Fraction shown in lane 3 after FPLC chromatography.

Neither remote IS*150* sequences nor IS*150*-encoded gene products play an essential role in frameshifting

Having established that translational frameshifting mediates the formation of the Ins150AB fusion protein we attempted to define the molecular requirements of this process. We designed an experimental system for the convenient quantitative monitoring of frameshifting using *lacZ* as a reporter gene. The basic plasmid (pFDX1725, Fig. 4) carried genes *galK* and *lacZ* coupled by a 83-bp cassette containing the IS*150* frameshift site (IS*150* coordinates 521−603) such that the 3' end of gene *galK* was fused in frame to *ins150A* and the 5' part of *lacZ* was fused in frame to *ins150B*. This *galK-insAB-lacZ* gene fusion was driven by the inducible *tac* promoter. In the presence of inducer this plasmid synthesized high levels of β-galactosidase activity (Fig. 4): The *galK-insAB-lacZ* fusion gave about 68% of the β-galactosidase activity synthesized by a reference plasmid (pFDX1727; Fig. 4) with an in-frame *galK-lacZ* fusion. We can therefore exclude that IS*150* sequences outside the bounds of the frameshift cassette play an important role in frameshifting. Essentially the same result was obtained when the experiment was repeated in *S.typhimurium* LT2, a strain which is devoid of IS*150*, indicating that no IS*150*-encoded functions are involved in the mechanism of translational frameshifting (data not shown).

Expression of *ins150B* also involves translational frameshifting

The β-galactosidase activity synthesized under the direction of plasmid pFDX1725 (68% compared to the control plasmid pFDX1727) is theoretically attributable to the sum of the activities of the Ins150AB-LacZ and Ins150B-LacZ fusion proteins. The results of the in vivo expression studies (in which a 5:1 ratio of Ins150AB to Ins150B synthesis was found) suggest that a significant fraction of the total activity is due to synthesis of Ins150B-LacZ. In order to measure the synthesis of Ins150B-LacZ directly we assayed β-galactosidase activity of transformants

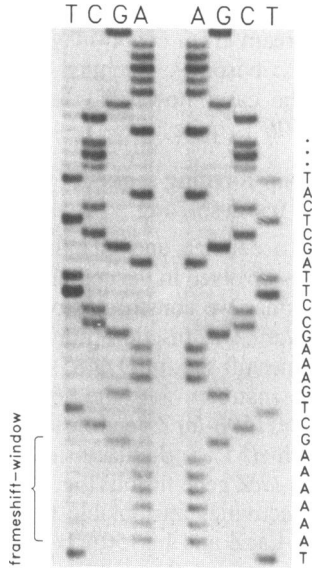


Figure 6. Sequencing ladder of a PCR-amplified cDNA prepared fromT *S.typhimurium* strain R954 transformed with plasmid pFDX1649 (left) and of a fragment prepared from plasmid pFDX1649 DNA (right). The A_6G motif (frameshift window) is labelled. The sequence read from the ladder is given on the right. Compare the corresponding RNA sequence given in Fig. 3.

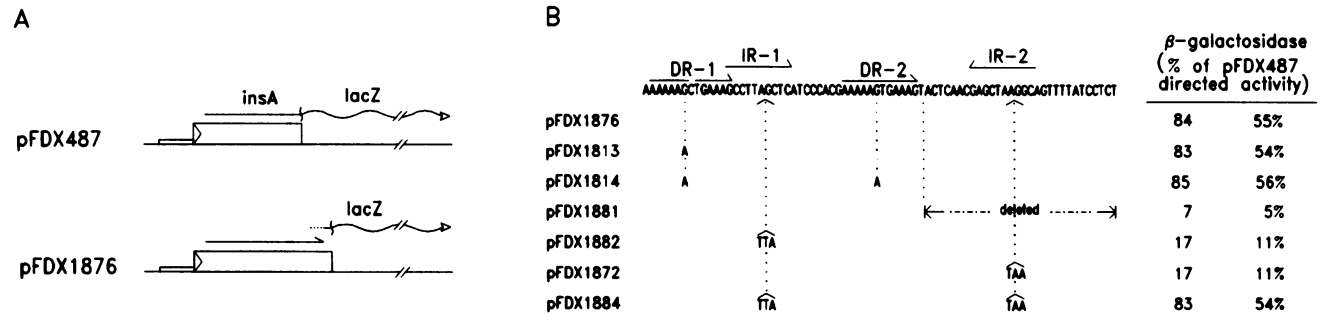


Figure 7. Analysis of sequence elements responsible for high-level translational frameshifting. Panel A: Relevant structures of the basic plasmids used. Plasmid pFDX487 contains an in-frame *ins150A-lacZ* fusion. The β-galactosidase activity synthesized by this plasmid (153 units) is take as 100%. Plasmid pFDX1876 contains a comparable in-frame fusion in *ins150B*. Note that in both plasmids gene expression is driven by the IS*150* promoter. Panel B: Mutations in the various derivatives of plasmid pFDX1876 and the β-galactosidase activities synthesized by these derivatives (in units and as% of pFDX487-directed activity).

harboring the identical 83-bp IS*150* cassette-*lacZ* in-frame fusion lacking *galK* sequences and hence the *galK* initiation signal (pFDX1780; Fig. 4). Surprisingly this plasmid produced only about 0.3% of the activity synthesized by the *galK-insAB-lacZ* fusion (Fig. 4). This suggests that an efficient translational start signal is not present in the cassette. A comparable construction containing an additional 20 bp of IS*150* sequence (bp 501−520) on the 5′ side (pFDX1781; Fig. 4; see also Fig. 3) produced considerable levels of enzyme activity. We conclude that an efficient ribosome binding signal for gene *ins150B* maps to the 20bp segment present in plasmid pFDX1781. The only possible translational start codon immediately downstream of this site is an ATG at bp 522, which is in the *ins150A* phase. A Shine−Dalgarno-like sequence is, however, not present. An in-frame *lacZ* gene fusion a few bp downstream of the ATG codon synthesized high levels of β-galactosidase activity (data not shown). Together these results indicate that the ATG is the the major translational start of gene *ins150B* and that expression of *ins150B* is mediated by frameshifting within the A AAA AAG sequence motif. Thus the major product of *ins150B* is a transframe protein, henceforth called Ins150A′B. We cannot rule out that the downstream GTG codon may be recognized as an initiation codon by ribosomes reaching the overlapping stop codon. If this is the case, however, the contribution to the translation of *ins150B* is quite low.

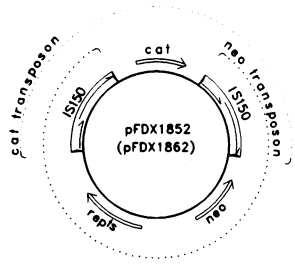
A potential stem-loop-forming sequence is required for high-level translational frameshifting

In order to pursue a detailed and quantitative analysis of the structural parameters involved in the coupling of the *ins150A* and *ins150B* reading frames we constructed two *lacZ* gene fusions (Fig.7): In the first *lacZ* was fused in frame to the *ins150A* gene upstream of the frameshift site (pFDX487), in the second, to the *ins150B* gene downstream of the frameshifting region (pFDX1876). The *ins150B-lacZ* gene fusion synthesized a little over half as much (55%) β-galactosidase activity as the comparable *ins150A-lacZ* gene fusion (Fig. 7). This value should be the sum of the activities attributable to the two frameshift products Ins150AB-LacZ and Ins150A′B-LacZ. Densitometric measurements demonstrated that the Ins150A′B-LacZ fusion protein contributes about one sixth (or 17%) of the total activity (see above). The remaining 38% should represent the contribution of the Ins150AB-LacZ fusion protein.This is somewhat higher than that expected on the basis of the densitometric measurements (33%) but is within the bounds of the accuracy of the experimental system. Thus the ratio of enzyme activities synthesized in this system reflects the overall rate of frameshifting. We used the *ins150B-lacZ* fusion (pFDX1876) to construct a series of mutants as follows (Fig. 7): First we probed the role of the direct sequence repeat (Fig. 3) in frameshifting. We employed site-directed mutagenesis to introduce a G→A exchange in the first repetitive element (DR-1). The resulting mismatch had no significant effect on the production of fusion protein as measured in the β-galactosidase assay (pFDX1813; Fig. 7). Neither did the double mutation consisting of G→A exchanges at corresponding positions of both units have an ostensive effect (pFDX1814; Fig. 7). Frameshifting is apparently insensitive to these alterations in the direct repeat. Furthermore the G in the frameshift motif A AAA AAG is not an important factor in the frameshifting mechanism. We then examined alterations in the potential stem-loop structure for an effect on frameshifting. We created a 30-bp deletion in *ins150B*, which removed the right arm of the palindrome but preserved the reading frame (pFDX1881; Fig. 4). Note that the potential *ins150B* start codon GTG within the loop is also preserved in this construction. This lesion resulted in a drastic reduction in β-galactosidase activity indicating that an essential sequence was affected. We strongly suspected that the essential sequence was the palindrome itself and proceeded to introduce topical mutations into both flanks. Klenow polymerase was used to fill in the symmetric DdeI sites within the palindrome, thus creating 3-bp insertions (pFDX1882 and pFDX1872; Fig. 7); both insertion mutations had the same effect: only a fifth of the β-galactosidase activity found in the wild type was synthesized. The double mutation which restored and lengthened the palindrome restored the synthesis of wild-type levels of enzyme activity (pFDX1884; Fig. 4B). These results show conclusively that the palindrome plays a key role in determining the rate of frameshifting. Moreover, they provide cogent evidence that an alternatively base-paired structure is a prime agent in the coupling mechanism.

The rate of frameshifting may affect transposition activity of IS*150*

Does frameshifting play a role in the regulation of transposition frequency by altering the ratio of the IS*150* gene products Ins150A, Ins150A′B and Ins150AB? To assess the latter possibility we studied the effects of experimental perturbation of frameshifting on the rates of transposition and cointegration of IS*150*. For these studies we constructed a temperature-sensitive test plasmid carrying an artificial IS*150* transposon (pFDX1852;

A



B

Plasmid	cm^r, kan^s	cm^s, kan^r	cm^r, kan^r
pFDX1852	$3.2 \pm 1.7 \times 10^{-6}$	$6.2 \pm 3.5 \times 10^{-6}$	$9.0 \pm 4.5 \times 10^{-6}$
pFDX1862	$5.6 \pm 1.7 \times 3.6^{-6}$	$1.5 \pm 0.9 \times 10^{-6}$	$6.1 \pm 2.3 \times 10^{-8}$

Figure 8. Transposition of IS150-flanked artificial transposons. Panel A: Structure of the thermosensitive plasmids pFDX1852 (wild type IS*150*) and pFDX1862 (IS*150* allele carrying the 3-bp insertion in the palindromic sequence as in plasmid pFDX1882, see Fig. 7). All structures are drawn to scale. Panel B: Fraction of single (transposition events) and double resistant colonies (cointegration events) formed at non-permissive temperature.

Fig. 8A). This transposon consisted of two tandemly arranged wild-type copies of IS*150* flanking the *cat* gene for chloramphenicol resistance. In this configuration the vector moiety with the *neo* gene for kanamycin resistance constitutes a second potential transposon (marked in Fig. 8A). A second plasmid (pFDX1862) was identical with the exception that in both flanking IS*150*s the palindromic sequence was disrupted by a 3-bp insertion in the left arm (mutant allele on pFDX1882; Fig. 7). The latter mutation significantly lowers the rate of frameshifting but leaves the reading frames intact (see above). In our experimental system strain R1115, a recA derivative of *S.typhimurium* LT2, was transformed with the test plasmids and the transformants were cultivated at 30°C. *S.typhimurium* LT2 is devoid of IS*150* sequences which could interfere with the transposition assay. Following a temperature shift to 42°C the single and double resistant survivors were counted. These numbers reflect the transfer of antibiotic resistance determinants to the chromosome as the result of transposition and cointegration, respectively. Fig. 8B summarizes the results: The *cat* transposon formed by wild-type copies of IS*150* (pFDX1852) transposed at a rate of 3.2×10^{-6}. The rate of inverse transposition (i.e. transposon of the IS*150*-flanked *neo* determinant) was somewhat but not significantly higher (6.2×10^{-6}). This configuration also mediated cointegration at a rate of 9×10^{-6}. These data indicate that the artificial transposon is transpositionally active and able to catalyze transposition as well as cointegration. The transposon formed by the mutant IS*150*s (pFDX1862) behaved differently: The transposition rates did not differ significantly from that of the 'wild type' transposon (transposition: 5.6×10^{-6} vs. 3.3 X 10^{-6}; inverse transposition: 1.5×10^{-6} vs. $6{,}2 \times 10^{-6}$). The effect on the rate of cointegrate formation, however, was drastic: a more than 100fold reduction in cointegrate formation was observed (6.1×10^{-8} vs. 9×10^{-6}). Thus, in the mutant tested it seems that not the transpositional activity *per se*, but rather the activity to catalyze cointegration has been affected. In this mutant the rate of frameshifting and hence the ratio of IS*150* gene products has been altered. However, the insertion of 3 bp which disrupts the palindromic sequence, also results in the insertion of an additional amino acid in the resultant gene products. We cannot at present exclude an effect of this modification on cointegration. However, in light of the fact that transposition and cointegration are mechanistically linked, the near-wild type rate of transposition of the mutant argues against a critical change in the structure of a transposition enzyme.

DISCUSSION

We have shown that genes *ins150A* and *ins150B* of the insertion element IS*150* encode three proteins with apparent molecular masses of 19, 31 and 49kD. The 19 and 31kD proteins are the products of *ins150A* and *ins150B*, respectively, while the 49kD protein is a transframe gene product determined by the coding regions of both genes (Fig. 2). The amino acid sequence of the part of the protein corresponding to the region where *ins150A* and *ins150B* reading frames overlap shows that Ins150AB is a fusion protein consisting of domains derived from both genes (Fig. 3). The synthesis of fusion proteins from out-of-frame genes has been reported for various systems including the *E.coli dnaX* gene (34−36), the mobile elements IS1 (38−40) and IS*911* (37) and for retroviruses (reviewed in (41)). A mechanism directing specific rephasing of ribosomes at special sites on the mRNA has been proposed as the basis of this phenomenon. However, in the various cases cited above the production of transframe gene products could be the result of specific editing of the mRNA (reviewed by (42)) or of polymerase slippage. While RNA editing has so far not been reported for bacterial systems, recent studies showed that *E.coli* RNA polymerase can slip backwards when it encounters monotonous A stretches. This results in the incorporation of additional A residues into the transcript (43). To clarify this point in the case of IS*150* we determined the sequence of the IS*150* transcript in the critical region and found it to be colinear to the template (Fig. 6). This rules out mechanisms involving RNA editing or polymerase slippage and, together with the amino acid sequence, provides conclusive evidence for ribosomal rephasing.

Experiments with a cassette containing the IS*150* frameshift site showed that all sequences and signals required for efficient frameshifting lie in the region between IS*150* bp 521−603 (Fig. 4). This frameshift window contains structural features found in other well-studied −1 frameshift regions. The amino acid sequence of the IS*150* fusion protein establishes that the site of frameshifting coincides with the heptanucleotide A AAA AA-G (Fig. 3). This motif is associated with −1 frameshifting in the synthesis of the gamma and tau subunits of the *E.coli* DNA polymerase (34−36) and in the production of the transposase of IS*911*, a relative of IS*150* (37). A similar sequence, A AAA AAC, is found at frameshift sites in IS*1* and retroviral systems. Recent studies on a synthetic HIV frameshifting region showed that alteration of the heptanucleotide frameshift sequence from A AAA AAG to A AAA AAA brought about a significant reduction in frameshifting frequency in *E.coli* (44). A similar G→A replacement in the hepta-nucleotide sequence of the IS*150* frameshifting region had no significant effect on frameshifting efficiency (Fig. 7). Downstream of the frameshift site is a palindromic sequence which could form a stable secondary structure in the transcript ($\Delta G = -20.8$ kcal/mol). Potential Stem-loop forming sequences are typical of −1 frameshift regions (41,45). Studies on the coronavirus IBV indicate that the distance between the potential secondary structure and the heptanucleotide is important (46). The spacing of the heptanucleotide A AAA

AAG and the potential stem-loop in the IS*150* frameshift region (6 bp) is identical to that found in the *E.coli dnaX* gene and in insertion sequence IS*911*. Our results suggest that the stem-loop structure plays an essential role in frameshifting. Deletion of one of the inverted repeats lowers the rate of frameshifting to below 5% (Fig. 7). Likewise, the introduction of a 3-bp mismatch into either inverted repeat results in a significant reduction in the rate of frameshifting (by about 80% in both cases). However, the introduction of complementary insertions into corresponding positions of the inverted repeats restores wild-type levels of frameshifting. This indicates that the downstream stem-loop acts to enhance or amplify frameshifting at the frameshift site. In the absence of this element frameshifting may continue, but at a substantially reduced rate (pFDX1881 in Fig. 7). Moreover these results are strong evidence that the effect of the downstream sequences on frameshifting depends on Watson–Crick basepairing in the transcript. Increasing the stability of the hypothetical secondary structure from $\Delta G = -20.8$ kcal/mol to $\Delta G = -25.6$ kcal/mol does not, however, increase the rate of frameshifting above the wild-type level. The IS*150* frameshifting region also contains a 12-bp direct repeat. One copy of the repetitive elements (DR-1) overlaps the heptanucleotide frameshifting signal. The introduction of a single-base mismatch in one of the copies of the repetitive sequence had no significant effect on frameshifting rates, suggesting that the direct repeat is not involved in frameshifting. Our data also show that IS*150* gene products are neither required nor have a significant effect on frameshifting within the element.

An experiment designed to measure the expression of gene *ins150B* using a simple *ins150B-lacZ* gene fusion gave a surprising result: the major product of *ins150B* is also a transframe protein initiated in the *ins150A* frame just upstream of the frameshift site. The ribosome binding site which directs translation of this protein, which we call Ins150A'B, must lie between bp 501 and 521 (Fig. 3). A typical Shine–Dalgarno sequence is not present. An ATG at bp 522–524 is the only potential start codon in the region (see Fig. 3). The complete translation initiation region of *ins150A'B* directs the synthesis of moderate levels of β-galactosidase activity in a *lacZ* gene fusion.

The rate of frameshifting in IS*150* is high. Densitometric measurements show that Ins150A and Ins150AB are made in minicells as well as in vivo at a ratio of approx. 2:1. This suggests that about one third or 33% of the ribosomes which initiate at the *ins150A* start codon and arrive at the frameshift site shift into the −1 frame and continue elongating. If ribosomal frameshifting is a simple gating mechanism causing a certain fraction of incoming ribosomes to rephase, then the value should be representative for the overall rate of frameshifting i.e. for the total volume of translation ensuing from both the *ins150A* and *ins150A'B* starts. Densitometric measurement shows that the ratio of InsAB-LacZ to InsA'B-LacZ in vivo is about 5:1 (Fig. 2C) i.e., InsA'B-LacZ is about 17% of the total frameshift product. A comparable *ins150B-lacZ* gene fusion (pFDX1876 in Fig. 7) gave 55% of the activity of the control *ins150A-lacZ* gene fusion. Deducting the activity attributable to the Ins150A'B-LacZ frameshift protein, which is approx. 17%, we arrive at a value of 38% for the activity stemming from the Ins150AB-LacZ frameshift protein. Thus, in this system we arrive at a somewhat higher estimate of the frameshifting rate than that obtained by densitometry (33%). The margin of error of densitometric measurement may in part account to this discrepancy. The GTG start codon of *ins150B* (Fig. 3) may contribute to the translation of the *ins150B*; this contribution, however, cannot exceed 5% activity synthesized by an *ins150A-lacZ* gene fusion (Fig. 7).

A cassette encompassing IS*150* coordinates 521–603 and thus containing all of the essential frameshift signals as well as the ATG assumed to be the start codon of the InsA'B protein but lacking the *insA'B* translation initiation signals was inserted between genes *galK* and *lacZ* (plasmid pFDX1725 in Fig. 4). This fusion produced high levels of enzyme activity (68% that of a *galK-lacZ* control fusion). We assume but cannot yet prove that in this case the heterologous sequences fused to the cassette stimulate translational initiation at the ATG codon and thereby direct the synthesis of artificially high levels of enzyme activity. This assumption is supported by another observation (K.Vögele and B.Rak, unpublished): an identical gene fusion containing a larger cassette (i.e. bp 501–603 instead of bp 521–603) synthesized only 52% of the control activity.

IS*150* is a member of a family of structurally related insertion sequences (1,13,37). The pattern of tandem, overlapping, out-of-phase genes and the typical frameshifting motifs which characterize the insertion sequences of this family suggest that they share a common mechanism of gene expression based on ribosomal frameshifting. There is now strong evidence that frameshifting is involved in the expression of IS*911* gene products (37). Interestingly, frameshifting also mediates gene expression in IS*1*, an apparently unrelated insertion sequence (38–40).

The properties of ribosomal frameshifting predispose it as a strategy of gene expression for IS elements; indeed, it may have evolved with them. Frameshifting provides a mechanism whereby two or more distinct but structurally related proteins can be encoded in a minimum of coding sequence—an important asset for mobile elements, which are ultracompact genetic modules. IS elements are vagabounds which visit and reside in various genetic contexts. Nevertheless, they behave as autonomous entities which regulate their transposition activity independent of the integration site. It has been pointed out that frameshifting ensures the autonomous regulation of transposition (40). This has been nicely shown for IS*1*. The transpositional activity of this element depends not on the absolute levels of the major gene products but on the balance or ratio of these products. Thus fluctuations in the transcription of the corresponding genes do not perturb transposition rates. The response of frameshifting systems to modulation may provide a sensitive mechanism for regulating rates of transposition: A decrease in the rate of frameshifting concomitantly reduces the amount of frameshift product and increases the amount of non-frameshift product and vice versa. Thus, slight changes in frameshifting frequency have a marked effect on the ratio of gene products. If frameshifting in IS*150* is subject to modulation by host translation factors, by the availability of tRNAs or by the frequency of translation it could provide a sensitive and rapid mechanism linking transposition frequency to the metabolic state of the host cell. Studies aimed at determining the function of the IS*150* gene products are now underway. An important clue is provided by the helix-turn-helix motif in protein Ins150A. This domain resembles that found in several DNA binding proteins. Previously we suggested that this protein may mediate specific binding to the inverted repeats of the element, whereas the catalytic activity for transposition may reside within Ins150B (1). This hypothesis was based mainly on two observations: (i) The inverted repeats of the elements of this family are, with the exception of the constant terminal nucleotides TG...CA, unrelated. (ii) The amino acid sequences deduced from the *insB* reading frames show a

high degree of conservation whereas the amino acid sequences of the *insA* frames are much less homologous. The data presented here and recent findings on IS*1* (40) lead us to think that the Ins150AB protein is the transposase and that Ins150A attenuates transpositional activity by competitive interaction with the ends of the elements. Transposition assays with artificial IS*150* transposons suggest that proteins Ins150A and Ins150A'B are not by themselves sufficient to catalyze transposition (C.Welz and B.Rak, unpublished). Despite the fact that in both IS*150* and IS*1* frameshifting has a key role in gene expression, the two elements differ in at least one respect: IS*1* apparently does not express an InsA'B-type transframe protein (40). We have shown that a mutation in the first unit of the inverted repeat, which reduces frameshifting (Fig. 7), drastically reduces cointegrate formation, but has no significant influence on the rate of simple transposition (Fig. 8). In IS*150* the expression of the *ins150B* reading frame is initiated within the phase of *ins150A* to give an Ins150A'B transframe product. Thus, not only expression of Ins150AB, the presumed transposase, but also expression of Ins150A'B is reduced in the mutant. Moreover, the different levels of expression of Ins150A'B in minicells and in vivo (Fig. 2) suggest that the synthesis of Ins150A'B may be subject to modulation by the metabolic state of the host cell. We suggest that the rate of cointegrate formation is regulated by the rate of Ins150A'B expression. This hypothesis is currently being tested.

ACKNOWLEDGEMENTS

We thank K.Schnetz for many helpful suggestions and for careful reading of the manuscript and G.Igloi for oligonucleotide synthesis. This work was supported by the Deutsche Forschungsgemeinschaft through grant Ra276/2-7. This article is dedicated to Peter Starlinger on the occasion of his 60th birthday.

REFERENCES

1. Schwartz,E., Kroeger,M. and Rak,B. (1988) *Nucleic Acids Res.*, **16**, 6789–6802.
2. Hall,B.G., Parker,L.L., Betts,P.W., DuBose,R.F., Sawyer,S.A. and Hartl,D.L. (1989) *Genetics*, **121**, 423–431.
3. Prère,M.-F., Chandler,M. and Fayet,O. (1990) J. Bacteriol., **172**, 4090–4099.
4. Matsutani,S., Ohtsubo,H., Maeda,Y. and Ohtsubo,E. (1987) *J. Mol. Biol.*, **196**, 445–455.
5. Priefer,U.B., Kalinowski,J., Rüger,B., Heumann,W. and Pühler,A. (1989) *Plasmid*, **21**, 120–128.
6. Kearney,B. and Staskawicz,B.J. (1990) *J. Bacteriol.*, **172**, 143–148.
7. Rubens,C.E., Heggen,L.M. and Kuypers,J.M. (1989) *J. Bacteriol.*, **171**, 5531–5535.
8. Thierry,D., Cave,M.D., Eisenach,K.D., Crawford,J.T., Bates,J.H., Gicquel,B. and Guesdon,J.L. (1990) *Nucleic Acids Res.*, **18**, 188.
9. McAdam,R.A., Hermans,P.W.M., van Soolingen,D., Zainuddin,Z.F., Catty,D., van Embden,J.D.A. and Dale,J.W. (1990) *Molecular Microbiology*, **4**, 1607–1613.
10. Hu,W.S., Wang,R.Y.-H., Liou,R.-S., Shih,J.W.-K. and Lo,S.-C. (1990) *Gene*, **93**, 67–72.
11. Schwartz,E., Herberger,C. and Rak,B. (1988) *Mol. Gen. Genet.*, **211**, 282–289.
12. Charlier,D., Piette,J. and Glansdorff,N. (1982) *Nucleic Acids Res.*, **10**, 5935–5948.
13. Fayet,O., Ramond,P., Polard,P., Prere,M.F. and Chandler,M. (1990) *Molecular Microbiology*, **4**, 1771–1777.
14. Khan,E., Mack,J.P.G., Katz,R.A., Kulkosky,J. and Skalka,A.M. (1991) *Nucleic Acids Res.*, **19**, 851–860.
15. Varmus,H. and Brown,P. (1989) In Berg,D.E. and Howe,M.M. (eds.), *Mobile DNA*. American Society for Microbiology, Washington, D.C., pp. 53–108.
16. Rak,B. and von Reutern,M. (1984) *EMBO J.*, **3**, 807–811.
17. Miller,J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
18. Schnetz,K. and Rak,B. (1988) *EMBO J.*, **7**, 3271–3277.
19. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
20. Kramer,W., Drutsa,V., Jansen,H.-W., Kramer,B., Pflugfelder,M. and Fritz,H.-J. (1984) *Nucleic Acids Res.*, **12**, 9441–9456.
21. Kunkel,T.A., Roberts,J.D. and Zakour,R.A. (1987) *Methods Enzymol.*, **154**, 367–382.
22. Bolivar,F., Rodriguez,R.L., Greene,P.J., Betlach,M.C., Heyneker,H.L. and Boyer,H.W. (1977) *Gene*, **2**, 95–113.
23. Bolivar,F. (1978) *Gene*, **4**, 121–136.
24. Vieira,J. and Messing,J. (1982) *Gene*, **19**, 259–268.
25. Rak,B. (1983) *Mol. Gen. Genet.*, **192**, 61–65.
26. Schnetz,K., Sutrina,S.L., Saier,M.H.,Jr. and Rak,B. (1990) *J. Biol. Chem.*, **265**, 13464–13471.
27. Tabor,S. and Richardson,C.C. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1074–1078.
28. Amann,E. and Brosius,J. (1985) *Gene*, **40**, 183–190.
29. Hashimotoh-Gotoh,T., Franklin,F.C.H., Nordheim,A. and Timmis,K.N. (1981) *Gene*, **16**, 227–235.
30. Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
31. Bölker,M. and Kahmann,R. (1989) *EMBO J.*, **8**, 2403–2410.
32. Fowler,A.V. (1972) *J. Bacteriol.*, **112**, 856–860.
33. Bachmann,B., Lüke,W. and Hunsmann,G. (1990) *Nucleic Acids Res.*, **18**, 1309.
34. Blinkowa,A.L. and Walker,J.R. (1990) *Nucleic Acids Res.*, **18**, 1725–1729.
35. Flower,A.M. and McHenry,C.S. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 3713–3717.
36. Tsuchihashi,Z. and Kornberg,A. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 2516–2520.
37. Polard,P., Prere,M.F., Chandler,M. and Fayet,O. (1991) *J. Mol. Biol.*, **in press**.
38. Sekine,Y. and Ohtsubo,E. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 4609–4613.
39. Lüthi,K., Moser,M., Ryser,J. and Weber,H. (1990) *Gene*, **88**, 15–20.
40. Escoubas,J.M., Prère,M.-F., Fayet,O., Salvignol,I., Galas,D., Zerbib,D. and Chandler,M. (1991) *EMBO J.*, **10**, 705–712.
41. Hatfield,D. and Oroszlan,D. (1990) *TIBS*, **15**, 186–190.
42. Weiner,A.M. and Maizels,N. (1990) *Cell*, **61**, 917–920.
43. Harley,C.B., Lawrie,J., Boyer,H.W. and Hedgpeth,J. (1990) *Nucleic Acids Res.*, **18**, 547–552.
44. Weiss,R.B., Dunn,D.M., Shuh,M., Atkins,J.F. and Gesteland,R.F. (1989) *The New Biologist*, **1**, 159–169.
45. Craigen,W.J. and Caskey,C.T. (1987) Cell, **50**, 1–2.
46. Brierley,I., Digard,P. and Inglis,S.C. (1989) *Cell*, **57**, 537–547.