Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 8049–8053, November 1987
Genetics

Genetic analysis of the interaction of the insertion sequence IS*903* transposase with its terminal inverted repeats

(point mutations/cis-acting protein/DNA–protein interactions/transposition mechanisms)

KEITH M. DERBYSHIRE, LEON HWANG, AND NIGEL D. F. GRINDLEY

Yale University, Department of Molecular Biophysics and Biochemistry, New Haven, CT 06510

Communicated by Joan A. Steitz, August 10, 1987 (received for review June 26, 1987)

ABSTRACT **The insertion sequence IS*903* has perfect, 18-base-pair terminal repeats that are the presumed binding sites of its transposase. We have isolated mutations throughout this inverted repeat and analyzed their effect on transposition. We show that every position in the inverted repeat (with the possible exception of position 4) is important for efficient transposition. Furthermore, various substitutions at a single position can have a wide range of effects. Analysis of these hierarchical effects suggests that transposase contacts the minor groove in the region from position 13 to position 16 but makes major groove (or more complex) interactions with the outer portion of the inverted repeat. Our data indicate that the transposase exhibits relaxed specificity for the "second" end of a transposed segment; the defect in transposition of virtually all mutant inverted repeats can be rescued by a wild-type end. However, this rescue exhibits a pronounced position effect; in most cases, it is efficient only when the wild-type end is close to the 3′ end of the transposase gene. This confirms the cis-acting nature of the transposase protein and suggests the initial transposase–inverted repeat interaction is the rate-limiting step in transposition. From the behavior of transposons with one mutant and one wild-type end, we infer that the inverted repeat contains two functional domains—one for initial complex formation with transposase and the other for effective completion of transpositional recombination. To support this hypothesis we show that an end with a mutation in one domain can significantly rescue an end with a mutation in the other domain.**

One of the striking characteristics of most bacterial transposons is that they have terminal inverted repeats. These repeats define the transposing segment since they are joined precisely to the target site in transposition, and they must be correctly oriented for transposition to occur. It is assumed that these sequences are the recognition sites of the element-encoded transposase, because it has been clearly demonstrated, for many transposons, that these are the only DNA sequences required in cis for efficient transposition. The inverted repeats of different insertion sequences (ISs) vary in length from 8 base pairs (bp) (IS*91*) to 40 bp (IS*2*) and in their degree of homology (1–4). Although the inverted repeats of many insertion elements have been defined by sequence and deletion analysis, a comprehensive analysis of the effects of mutations in an inverted repeat has not been carried out. A single-base transition at the outside end of IS*50* reduces transposition by a factor of 100 (5). Three point mutations in the outer 13 bp of Tn*10* also reduce transposition by a factor of 100 (6). We, therefore, decided to analyze the effects of single substitutions in the inverted repeat of IS*903*.

Tn*903* is a composite transposon consisting of two identical, 1057-bp, IS*903* elements that lie in inverted orientation on either side of a DNA segment encoding resistance to kanamycin (7, 8). Each IS*903* element is capable of transposition and is flanked by perfect 18-bp inverted repeats. These are the only sequence elements required in cis for transposition (9), although it should also be noted that the IS*903* transposase is a cis-acting protein (7). This is a phenomenon that appears to be a general property of IS element transposases; those of IS*1*, IS*10*, and IS*50* also do not efficiently complement mutant transposons in trans (10–12). We have described (13) a saturation mutagenesis technique by which we isolated many mutations throughout the inverted repeat of IS*903*. We have used these to characterize the interaction of the IS*903* transposase with its inverted repeats.

MATERIALS AND METHODS

Standard molecular and genetic techniques were used (14). DNA sequences were determined by the dideoxynucleotide chain-termination procedure (15, 16). All point mutations, with the exceptions described below, were isolated as described (13). Point mutations at the thymidines at positions 11 and 12 (T^{11} and T^{12}) were isolated by the oligonucleotide-directed mutagenesis procedure of Kunkel (17). Oligonucleotides for mutagenesis and sequencing were made on a model 8600 synthesizer from Biosearch (San Rafael, CA).

Plasmids. The vector used for cloning the mutant ends pKD50 (Fig. 1) is a derivative of pBR322 that has been modified in the following ways. (*i*) A 1.5-kilobase (kb) fragment encoding chloramphenicol resistance (9) has been cloned into the *Pvu* II site. (*ii*) An *Xba* I linker has been inserted into the *Sal* I site. (*iii*) A *Sal* I linker has been inserted into the *Eco*RI site (the original *Pvu* II, *Sal* I, and *Eco*RI sites were all destroyed in these steps). (*iv*) A segment of DNA encoding the IS*903* transposase expressed from the *lac*UV5 promoter (18) has been cloned between the new *Sal* I site and the *Bam*HI site of pBR322. IS*903* ends were cloned into pKD50 to generate model transposons as follows. Single-stranded DNA from the M13 clones was made double stranded by primer extension, divided into two aliquots, and then digested with either *Bam*HI and *Pst* I or with *Xba* I and *Hin*dIII. After heat treatment (65°C, 10 min), the two reaction mixtures were mixed with the 6.2-kb *Bam*HI–*Xba* I fragment from pKD50 and a 1-kb *Pst* I–*Hin*dIII fragment containing the kanamycin resistance gene of Tn*903* [the *Hin*dIII site normally present within the kanamycin resistance gene has been destroyed by bisulfite mutagenesis (19)]. Ligation yielded pKD100 (with wild-type transposon ends) and analogous plasmids each with two copies of a single mutant end (Fig. 1). All inverted repeats cloned into pKD50 have been checked by restriction endonuclease and sequence analysis.

Plasmids with a similar structure to pKD100 were made in which the transposon had hybrid ends—one wild-type and one mutant (see Fig. 1). These plasmids contain a 2.5-kb *Pst* I fragment carrying the *Escherichia coli galK* gene inserted

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

Abbreviation: IS, insertion sequence.

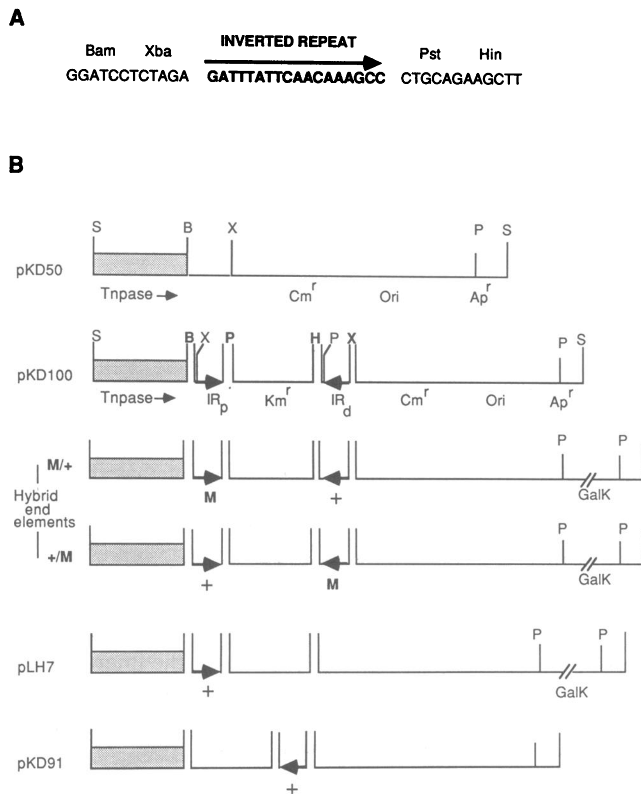

FIG. 1. Construction of IS*903* elements with mutant ends. (*A*) Sequence of the wild-type inverted repeat and its location in the polylinker of M13mp10. Mutant inverted repeats were subcloned from M13 into pKD50. (*B*) The basic structures of pKD50, pKD100, and other constructs are shown, along with relevant restriction sites. pKD100 has two wild-type inverted repeats (IR) that are located proximal (IR_p) and distal (IR_d) to the transposase gene. The hybrid-ended elements show the two locations used for the mutant (M) and wild-type (+) ends. These plasmids also contain the *galK* gene on a 2.5-kb *Pst* I fragment inserted in the *Pst* I site of the ampicillin resistance gene. pLH7 and pKD91 have only a wild-type end and were used as controls. Restriction enzyme sites: S, *Sal* I; B, *Bam*HI; X, *Xba* I; P, *Pst* I; H, *Hin*dIII. Sites shown in bold type for pKD100 are those used in cloning the inverted repeats. Cm^r, chloramphenicol resistance; Ori, origin; Ap^r, ampicillin resistance; Km^r, kanamycin resistance.

into the ampicillin resistance gene of the vector; this was used to simplify the screening of clones during construction of the plasmids and has no significant effect on transposition frequencies. Two other plasmids, pLH7 and pKD91, each of which contain just one IS*903* end, were used as controls and have the structure shown in Fig. 1.

Transposition Assays. A mating-out assay was used. The donor strain RR1023 (*recA56*) was transformed with the various chloramphenicol resistance transposon derivatives to be tested. RR1023 carries a deletion derivative of the F factor, pOX38, that is transfer proficient but contains no ISs (20). The recipient was NG135 (F^-, *recA56 strA*). Single colonies were grown overnight in liquid culture, diluted 1:25, and regrown to midlogarithmic phase. Donors (0.1 ml) and recipients (0.9 ml) were mixed and mated for 90 to 120 min at 37°C without shaking. Appropriate dilutions were then spread on selective plates. Transposition frequencies were then calculated as number of chloramphenicol- and streptomycin-resistant transconjugants per donor cell.

Sequencing Inverted Repeat–Target Junctions. Independent chloramphenicol- and streptomycin-resistant transconjugants from crosses involving various constructs were streaked onto selective plates. Plasmid DNA was then isolated from these clones. To overcome the low DNA yield from these pOX38::Tn derivatives, the majority of the pOX38 DNA was deleted. The DNA was digested to completion with *Cla* I (which cuts only outside the transposon and at least seven times within pOX38) and religated after diluting the DNA. Chloramphenicol-resistant transformants were selected, and these contained smaller, high copy number plasmids containing the entire transposon. This DNA was used directly for sequence analysis (16) using primers that enabled each inverted repeat–target junction to be sequenced.

RESULTS

Point Mutations Throughout the Inverted Repeat Affect IS*903* Transposition. It has been suggested (9) that the outer 20 bp of IS*903* were the only sequences necessary for efficient transposition. To confirm this we have made two transposase-defective derivatives of pKD100. One has a frameshift mutation at the *Mlu* I site early in the gene; the other has a deletion of the entire segment carrying the transposase gene and its promoter (the *Sal* I–*Bam*HI fragment in Fig. 1). When complemented in trans by a copy of the *lac*UV5-promoted transposase gene on pACYC184 both derivatives gave transpositions of chloramphenicol resistance at the same frequency (0.5%–1% of the transposition frequency of pKD100). This confirms the strong preference of the IS*903* transposase for acting in cis and shows that the 18-bp terminal inverted repeats are the only cis-acting sequences required for transposition.

Each of the mutant inverted repeats was subcloned from M13mp10 into the tester plasmid to reconstruct a transposon with two copies of the same mutant end and was then assayed for transposition (Fig. 2). It is clear that every base pair within the inverted repeat (with the possible exception of position 4) is important for efficient transposition. However, different substitutions at a single position may exhibit substantially different effects (the hierarchy of nucleotide substitutions is shown at the relevant positions). One of the more striking effects is seen at the thymidine at position 11 where a change to a guanosine ($T^{11} \rightarrow G$) has only a small effect on transposition whereas a change to adenosine ($T^{11} \rightarrow A$) or cytidine ($T^{11} \rightarrow C$) has very severe effects on transposition. Similarly, the substitutions of adenosines for thymidines at positions 14 to 16 have little or no effect on transposition compared to the other changes.

Rescue of Transposition Using a Wild-Type End. In the experiment described above each transposon contained two copies of the same mutation. We have also determined whether a wild-type end could rescue transposition in constructs containing one wild-type and one mutant inverted repeat. A variety of mutant inverted repeats were cloned into two tester plasmids—in one the wild-type end was proximal to the transposase gene and in the other it was in the distal location (Fig. 1). These hybrid-ended elements were then assayed for transposition (Table 1).

A single wild-type end showed significant rescue of all the mutant ends. However, a distinct asymmetry was observed. For many changes this rescue was only seen when the wild-type end was proximal to the transposase gene (consider, for example, mutations in the region from position 6 to 15; see Table 1). By contrast, mutations at the outside end of the inverted repeat (positions 1–3) were substantially rescued regardless of which end was proximal to the transposase gene. This suggested to us that the inverted repeat may consist of two functionally distinct domains (see below).

In these experiments we also used two deletion derivatives as controls to show that the single mutant end in the hybrid transposons was being used in transposition. One derivative, pKD91, contains a single wild-type inverted repeat in the transposase distal location (Fig. 1). Transposition was never detected from this element. The second derivative pLH7

Relative Transposition Frequency

3' ← → 5'

Substitution	C^1	C^2	G^3	A^4	A^5	A^6	C^7	A^8	A^9	C^{10}	T^{11}	T^{12}	A^{13}	T^{14}	T^{15}	T^{16}	A^{17}	G^{18}
T	7	2.9				.003			.05	.009			9.6				7.6	96
C			.07	38	.16	.009					0	6.6	.68	0	.004		4.4	143
G				25	6.4		0	.02			37	0.7	.69			2.5	3.9	
A			.05							0	.43	5.9		13	20	74		8
Hierarchy			C=A	C=G	G>C	T=C				T=A	G> A C	C A >G	T> G C	A>C	A>C	A>G	T= G C	T C >A

FIG. 2. Transposition frequencies of IS*903* elements with mutant ends. Transposons carrying identical copies of the same mutant end were tested for transposition by a mating-out assay. The transposition frequencies are expressed as a percentage of the wild-type construct pKD100, which transposed at a frequency of 1.6×10^{-4} per donor. Each estimate is the average of at least four matings. A value of 0 indicates that less than one transposition event occurred in 10^9 donors assayed. The wild-type sequence is shown above the table.

contains a single wild-type end proximal to the transposase gene (Fig. 1). This element transposed at 0.26% of the wild-type, two-ended element pKD100. These two plasmids indicate the activities of a single wild-type end in each of the two locations. The low level of transposition seen with pLH7 is presumably due to alternative pseudo-sites fulfilling the role of the second end, as has been documented before with IS*903* (18) and with a closely related element IS*102* (21). The transposition frequency of these one-ended elements is lower than that of the hybrid-ended transposons, suggesting that the latter make specific use of the mutant end.

In several cases (see Table 1) we have sequenced the new junctions of the products of transposition of the hybrid-ended transposons. In every case except one, transposition involved use of the combination of wild-type and mutant end, the mutation was retained, the recombination occurred precisely at the end of the mutant inverted repeat (even when position 1 was mutated), and (when not ambiguous) a 9-bp target site was duplicated, which is characteristic of normal IS*903* transposition. The exception was when the mutation $C^{10}\rightarrow T$ was in the distal location, where all the six independent events sequenced had not used the mutant end. We

Table 1. Transposition frequencies of IS*903* elements with one wild-type and one mutant end

Mutation	Transposition frequency, % of wild type		
	M/M*	M/+*	+/M*
$C^1\rightarrow T$	7†	118†	100†
$C^2\rightarrow T$	2.9	46	164
$G^3\rightarrow A$	0.05	15†	42†
$A^4\rightarrow G$	25	132	55
$A^5\rightarrow C$	0.16	3.9	13
$A^5\rightarrow G$	6.4	36	44
$A^6\rightarrow T$	0.003	0.34	6.3
$A^8\rightarrow G$	0.02	0	2.8†
$C^{10}\rightarrow T$	0.009	0	1.4†
$T^{12}\rightarrow C$	6.6	26†	100†
$T^{14}\rightarrow A$	13	31	138
$T^{14}\rightarrow C$	0	0.1	9
$T^{15}\rightarrow C$	0.004	0.12†	21†
Δ		0	0.26

Hybrid-ended transposons were constructed (Fig. 1) and then assayed for transposition. The transposition frequencies are expressed as a percentage of pKD100, which transposes at 1.6×10^{-4} per donor. A value of 0 indicates less than one transposition event in 5×10^8 donors.

*The end listed first is proximal to the transposase gene. M, mutant end; +, wild-type end; Δ, deletion of the end in the M location.

†Elements whose new junctions were sequenced after transposition.

assume that these transposition events are analogous to those generated by pLH7 (the one-ended element), implying that this mutant end is so defective that it is no longer recognized by the transposase.

The Inverted Repeat Has Two Functional Domains. As suggested above, the asymmetric behavior of the transposons with one mutant and one wild-type end suggested that the inverted repeat might contain two functionally distinct domains. One domain would include positions 1–3, and the other domain would include positions 6–15 (with positions 4 and 5 ambiguous). Knowing that a wild-type end can rescue mutants within each of these regions in a characteristic fashion, then, if each region indeed corresponds to a functionally independent domain, we would expect that an end with a mutation in one domain should be rescued significantly by an end with a mutation in the other domain. Moreover, the rescue should have a predictable asymmetry. We have tested this hypothesis by constructing transposons in which one end carries the $G^3\rightarrow A$ mutation and the other end carries the $T^{15}\rightarrow C$ mutation.

It is clear from Fig. 3 that the element with the $G^3\rightarrow A$ mutant end proximal and the $T^{15}\rightarrow C$ end distal to the transposase gene transposes more frequently than the transposon with the $G^3\rightarrow A$ mutation in both ends, even though the $T^{15}\rightarrow C$ end is a much poorer transposition substrate than the

Proximal end \ Distal end	+	$G^3\rightarrow A$	$T^{15}\rightarrow C$
+	100	42	21
$G^3\rightarrow A$	15	0.05	0.6
$T^{15}\rightarrow C$	0.12	0	0.004

FIG. 3. Rescue of transposition using mutants from different domains. Two transposons were constructed that had a combination of the mutant ends $G^3\rightarrow A$ and $T^{15}\rightarrow C$. Their transposition frequencies are shown in the highlighted boxes. The other data are from Table 1 and Fig. 2 and are presented for ease of comparison. Proximal and distal indicate the location of the end with respect to the transposase gene. The transposition frequencies are expressed as a percentage of pKD100 (+/+ value = 100); the frequencies for the transposons in which the proximal end is $G^3\rightarrow A$ and the distal end is either $G^3\rightarrow A$ or $T^{15}\rightarrow C$ are the average of at least 10 matings. A value of 0 indicates that less than one transposition event occurred in $>10^9$ donors assayed.

$G^3 \rightarrow A$ end. Furthermore, as predicted from our model, this rescue is position dependent; there is no rescue of transposition with the element with the $T^{15} \rightarrow C$ end proximal and the $G^3 \rightarrow A$ end distal. The rescue observed is within reasonable limits of the theoretical maximum expected for this construct. We can derive such a value from the data in Table 1, assuming that each end is effectively wild type for the nonmutated domain. The maximum transposition frequency expected is 3.1% [i.e., the product of the frequencies of the transposon with a $G^3 \rightarrow A$ proximal end and a wild-type distal end (15%) and the transposon with a wild-type proximal end and a $T^{15} \rightarrow C$ distal end (21%)].

DISCUSSION

We present a comprehensive genetic analysis of a transposase binding site: the inverted repeat of IS*903*. The data allow us to draw several conclusions about the way the transposase interacts with the inverted repeat. First, the transposase "senses" changes along the full length of the inverted repeat (Fig. 2); every position (with the possible exception of position 4) has at least one change that has a deleterious effect on transposition. This raises an interesting problem in protein–DNA recognition. Most of the protein–DNA site specific interactions that have been well characterized [e.g., the repressors of the λ phages and the *E. coli* cAMP-binding protein (CAP protein) (22, 23)] involve a relatively short segment of DNA (5 or 6 bp) repeated in inverted orientation to allow for symmetrical interaction with a protein dimer. Formally, it is possible that the inverted repeat of IS*903* is recognized as two half sites, since it contains an imperfect dyad symmetry centered on position 10. However, substitutions that improve the dyad symmetry (e.g., $A^{13} \rightarrow G$ and $A^{17} \rightarrow C$) are deleterious, and the effects of symmetric substitutions are highly variable (e.g., the substitution $T^{14} \rightarrow A$ has essentially no effect, whereas the symmetric change $A^6 \rightarrow T$ reduces transposition by a factor of 10^4). It seems, then, that the transposase must recognize the 18-bp inverted repeat as a single, unique element, sensing the individual base pairs either by specific amino acid–base pair contacts or through their effects on the local structure of the DNA double helix (by way of helix bending, changes in twist, etc.).

Consideration of the data in Fig. 2 where there are multiple substitutions at a single base shows that some changes are more deleterious than others. We have examined these hierarchical effects with regard to how the transposase might recognize the inverted repeat. It has been proposed that, if a recognition protein were to use hydrogen bonds for specificity, a minimum of two hydrogen bonds per base pair would be required to specify all bases (24). A single hydrogen bond would result in degeneracies in that two different base pairs could be recognized. Such rules have been applied to explain the basis of *Eco*RI and *Eco*RI* specificity (25) and have been confirmed by analysis of DNA–*Eco*RI cocrystals (26). One of the predictions of this hypothesis is that hydrogen bonding to base pairs in the minor groove cannot distinguish base-pair reversals (e.g., A-T from T-A) because of the symmetrical location of hydrogen bond acceptors. The most striking hierarchies seen in Fig. 2 are those from positions 13 to 16 that are consistent with the above prediction; the most acceptable substitution (in terms of transposition frequencies) is a base-pair reversal. Thus, although we expect the transposase to interact with the inverted repeat in a variety of ways and to be sensitive to local structural changes, we suggest that the transposase recognizes the minor groove in this portion of the inverted repeat. Hierarchies seen elsewhere in the inverted repeat are inconsistent with minor groove interactions (e.g., at position 6 and position 11) and thus probably represent major groove and more complex interactions.

The analysis of transposons with one mutant and one wild-type end provided two interesting observations. First, nearly all mutants were substantially rescued by a wild-type end (Table 1). This was true even for extremely deleterious mutations (e.g., $A^6 \rightarrow T$, $T^{14} \rightarrow C$, and $T^{15} \rightarrow C$) and suggests that the transposase shows substantially relaxed specificity for the second end in the transposition process. This is analogous to the outside-end preference for IS*10* and IS*50*, where the poorer transposition frequency of an element with two inside ends can be rescued by the more active outside end (12, 27, 28). This relaxation of specificity of the IS transposases may have played a significant role in the evolution of these transposons (29).

Second, in many cases rescue by a wild-type end showed a striking asymmetry that was determined by the location of the wild-type end (Table 1). The most efficient (and sometimes the only) rescue of transposition was seen when the wild-type end was proximal to the transposase gene. A clear exception to this position effect is provided by the $G^3 \rightarrow A$ mutant end that is efficiently rescued regardless of location. $C^2 \rightarrow T$ and $C^1 \rightarrow T$ are probably also exceptions. This suggested to us that mutations close to the outside end of the inverted repeat (positions 1–3 with no significant data for position 4) define a distinct domain of the transposon terminus that is not involved in the initial formation of complexes with transposase.

We propose the following hypothesis to account for our data that is based on the cis-action of the IS*903* transposase and on previous models (27–29). We suggest that immediately following (or during) its synthesis the transposase searches along the adjacent DNA for an inverted repeat, generally finding the one closest to the 3′ end of the gene. The proximity of this recognition site is a major factor in determining the efficiency of transposition. In support of this statement we find that either inserting a 1-kb segment between the 3′ end of the transposase gene and the proximal inverted repeat or reversing the orientation of the transposase gene in a construct otherwise the same as pKD100 (which effectively moves its 3′ end 1 kb away from the nearest inverted repeat) reduces transposition by a factor of 20–50 (K.M.D., N.D.F.G., P. Rice, and M. Kramer, unpublished data). It seems then, that during its initial search the transposase has a very short functional half-life. Once one end is located and bound by transposase, we suggest that the complex then searches along the DNA for a second end to allow the transposition to take place. Although this second step is also sensitive to distance, as shown by the effect of increasing transposon length (9), it is not nearly as sensitive as the first transposase–inverted repeat interaction, implying that once a complex forms the transposase is functionally much more stable. The transposases of other elements have also been shown to act preferentially in cis. However, they appear to be much less sensitive to distance; the IS*10* transposase shows a reduction of slightly more than a factor of 10 in transposition frequency if an element is placed 50 kb away compared to an element placed 1 kb away (11).

According to this hypothesis, those mutants that exhibit a position effect in the transposons with a single wild-type end, fail to form the initial complex when located adjacent to the transposase gene. The transposase, therefore, has to search an additional 1 kb to find the wild-type end with a consequent, large reduction in transposition efficiency. The corollary of this is that those mutations that do not result in a position effect are recognized by transposase and form the initial stabilizing complex with nearly normal efficiency. Such mutations either define a domain of the inverted repeat that is not involved in formation of the initial complex or do lie in a DNA binding domain but have little effect on binding. We suggest that the former class includes $C^1 \rightarrow T$, $C^2 \rightarrow T$, and $G^3 \rightarrow A$, whereas $T^{14} \rightarrow A$, which has only a mild effect on

transposition even in transposons with two mutant ends, is an example of the latter class (note that $T^{14} \rightarrow C$ shows a pronounced position effect).

If there are two functionally independent domains within the 18-bp inverted repeat, then, since mutations within each domain can be rescued by a wild-type end, one would expect an end with a mutation in one domain (but wild type in the other) to rescue an end with a mutation in the other domain. (This is analogous to genetic complementation between mutations in different genes but responsible for the same phenotype). In addition, since rescue of mutations in each domain exhibits a characteristic position effect, rescue of one mutant end by another should show a predictable asymmetry. We have demonstrated the mutual and position-dependent rescue of a pair of mutant ends, one with the $G^3 \rightarrow A$ mutation, the other with the $T^{15} \rightarrow C$ mutation; the hybrid element transposes at least 10-fold better than transposons with two copies of either $G^3 \rightarrow A$ or $T^{15} \rightarrow C$, but only if the $G^3 \rightarrow A$ mutation is located adjacent to the transposase gene. We suggest that the $G^3 \rightarrow A$ mutation with the $C^1 \rightarrow T$ and $C^2 \rightarrow T$ mutations define a distinct domain of the inverted repeat. Since ends with these mutations can efficiently initiate transposition (demonstrated by the high levels of transposition of hybrid transposons with these ends proximal to the transposase gene), then we conclude that their defect in transposition is in a step subsequent to transposase binding. As this domain is at the end of the repeat that is cut and joined to target DNA, it seems likely that the transposase (or host factor) interacts with this region when mediating these steps. We suggest that the larger domain (from the adenosine at position 6 to the thymidine at position 15) is involved in transposase binding; this would be consistent with the reduced transposition frequencies when ends defective in this region are proximal to the transposase gene.

We have shown that the transposase of IS*903* has a strong preference for an inverted repeat close to the 3′ terminus of its gene. This means that in the composite transposon Tn*903* the most active ends are the outside ends of the element. This is also true for IS*10* (27) and IS*50* (12, 28), although the preference here is dictated more by binding affinity than cis-action. This outside end preference may help ensure the integrity of the composite transposon.

We thank C. Hughes for excellent technical assistance; V. Derbyshire, G. Hatfull, C. Joyce, and R. Macnab for many useful discussions throughout this work and their comments on this manuscript. This work was supported by Grant GM28470 from the National Institutes of Health.

1. Kleckner, N. (1981) *Annu. Rev. Genet.* **15**, 341–404.
2. Iida, S., Meyer, J. & Arber, W. (1983) in *Mobile Genetic Elements*, ed. Shapiro, J. A. (Academic, New York), pp. 159–221.
3. Heffron, F. (1983) in *Mobile Genetic Elements*, ed. Shapiro, J. A. (Academic, New York), pp. 223–260.
4. Diaz-Aroca, E., Mendiola, M. V., Zabala, J. C. & de la Cruz, F. (1987) *J. Bacteriol.* **169**, 442–443.
5. Sasakawa, C., Phadnis, S. H., Carle, G. F. & Berg, D. E. (1985) *J. Mol. Biol.* **182**, 487–493.
6. Way, J. C. & Kleckner, N. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3452–3456.
7. Grindley, N. D. F. & Joyce, C. M. (1981) *Cold Spring Harbor Symp. Quant. Biol.* **45**, 125–133.
8. Oka, A., Sugisaki, H. & Takanami, M. (1981) *J. Mol. Biol.* **147**, 217–226.
9. Weinert, T. A., Derbyshire, K. M., Hughson, F. M. & Grindley, N. D. F. (1984) *Cold Spring Harbor Symp. Quant. Biol.* **49**, 251–260.
10. Machida, Y., Machida, C., Ohtsubo, H. & Ohtsubo, E. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 277–281.
11. Morisato, D., Way, J. C., Kim, H.-J. & Kleckner, N. (1983) *Cell* **32**, 799–807.
12. Isberg, R. S. & Syvanen, M. (1981) *J. Mol. Biol.* **150**, 15–32.
13. Derbyshire, K. M., Salvo, J. J. & Grindley, N. D. F. (1986) *Gene* **46**, 145–152.
14. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
15. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3963–3965.
16. Chen, E. Y. & Seeburg, P. H. (1985) *DNA* **4**, 165–170.
17. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 488–492.
18. Weinert, T. A., Schaus, N. A. & Grindley, N. D. F. (1983) *Science* **222**, 755–765.
19. Shortle, D. & Nathans, D. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2170–2174.
20. Guyer, M. S., Reed, R. R., Steitz, J. A. & Low, K. B. (1981) *Cold Spring Harbor Symp. Quant. Biol.* **45**, 135–140.
21. Machida, Y., Machida, C. & Ohtsubo, E. (1982) *Cell* **30**, 29–36.
22. Pabo, C. O. & Sauer, R. T. (1984) *Annu. Rev. Biochem.* **53**, 293–321.
23. Anderson, J. E., Ptashne, M. & Harrison, S. C. (1987) *Nature (London)* **326**, 846–852.
24. Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 804–808.
25. Rosenberg, J. M. & Greene, P. J. (1982) *DNA* **1**, 117–124.
26. McClarin, J. A., Frederick, C. A., Wang, B.-C., Greene, P., Boyer, H. W., Grable, J. & Rosenberg, J. M. (1986) *Science* **234**, 1526–1541.
27. Morisato, D. & Kleckner, N. (1984) *Cell* **39**, 181–190.
28. Sasakawa, C. & Berg, D. E. (1982) *J. Mol. Biol.* **159**, 257–271.
29. Sasakawa, C., Carle, G. F. & Berg, D. E. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 7293–7297.