J. Mol. Biol. (1998) **276**, 913–925

JMB

Molecular Genetic Analysis of Transposase–End DNA Sequence Recognition: Cooperativity of Three Adjacent Base-pairs in Specific Interaction with a Mutant Tn*5* Transposase

Maggie Zhou, Archna Bhasin and William S. Reznikoff*

Department of Biochemistry University of Wisconsin-Madison, 420 Henry Mall Madison, WI 53706, USA

*Corresponding author

Transposition of Tn*5* and IS*50* requires the specific binding of transposase (Tnp) to the end inverted repeats, the outside end (OE) and the inside end (IE). OE and IE have 12 identical base-pairs and seven non-identical base-pairs. Previously we described the isolation of a Tnp mutant, EK54, that shows an altered preference for OE *versus* IE compared to wild-type (wt) Tnp. EK54 enhances OE recognition and decreases IE recognition both in DNA binding and in overall transposition. Here we report that base-pairs 10, 11 and 12 of the OE are critical for the specific recognition by EK54 Tnp. These three adjacent base-pairs act cooperatively; all three must be present in order for EK54 Tnp to interact very favorably with the end DNA. The existence of only one or two of these three base-pairs decreases binding of EK54 Tnp. The combined use of EK54 Tnp and a new OE/IE mosaic end sequence containing the OE base-pairs 10, 11 and 12 gives rise to an extraordinarily high transposition frequency. Just as the Tnp is a multifunctional protein, the nucleotides in the 19 bp Tn*5* ends also affect other functions besides Tnp binding. Furthermore, the fact that we were able to isolate end sequence variants that transpose at a higher frequency than the natural ends (OE and IE) with wt Tnp reveals yet another way in which the wt transposition frequency is depressed, i.e. by keeping the end sequences suboptimal.

Keywords: base-pair cooperativity; protein–DNA interaction; Tn*5* transposon; transposition; transposase

Introduction

Tn*5* is a prokaryotic transposable element. The elegance of the Tn*5* transposition system lies in two respects. First, a single protein, Tn*5* transposase (Tnp), is able to carry out a complex series of sequential steps, nearly completing the whole transposition process (Goryshin & Reznikoff, 1998). Dissecting the various functional domains of this complex protein and comparing them with other proteins with related functions can help in the study of protein domain evolution. Second, and perhaps more interesting, is the fact that Tn*5* transposition is regulated at many levels through many different mechanisms (reviewed by Berg, 1989; Reznikoff, 1993). *tnp* promoter activity and IE availability are both regulated by host *dam* methylation; Tn*5* inhibitor protein (Inh) inhibits transposition by forming inactive dimers with Tnp; Tnp is primarily *cis* active, wherease Inh is active both *in cis* and *in trans*; translation of read-through transcripts of *tnp* initiated from fortuitous upstream promoters is prevented by a secondary structure that exists only in such read-through transcripts; furthermore, the Tnp protein is suboptimal in its specific binding to the Tn*5* end inverted repeat sequences (Zhou & Reznikoff, 1997). Understanding the complexity and mechanisms of these multilevel regulations in a system like this can contribute to our general understanding of complex biological systems.

The initial step of Tn*5* transposition involves specific binding of Tnp with the inverted repeats that define the ends of Tn*5* or its component transposable element, IS*50*. These end inverted repeats are called OE (outside end) and IE (inside end). OE and IE are both 19 base-pairs (bp) long, with 12

Abbreviations used: Tnp, transposase; OE, outside end; IE, inside end; wt, wild-type; Inh, inhibitor protein.

0022–2836/98/100913–13 $25.00/0/mb971579

Table 1. *Trans* papillation level of hybrid end sequences with EK54 Tnp

Mutant	Position:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Papillation level[a]	No. of times isolated[b]
IE:		c	t	g	T	c	t	c	t	t	G	A	T	c	a	G	a	T	C	t	VL	0
OE:					A						A	T	A			C		A	G		M	6
1											A	T	A								H	2
2											A	T	A			C					H	3
3											A	T	A					A			H	5
4											A	T	A			C		A			H	4
5											A	T	A			C			G		H	6
6											A	T	A					A	G		H	6
7											A	T	A			C		A	G		H	4
8											A	T	A						G		M	7
9					A						A	T	A								M	3
10					A						A	T	A			C					M	2
11					A						A	T	A					A			M	1
12					A						A	T	A						G			0
13					A						A	T	A			C		A				0
14					A						A	T	A			C			G		M	4
15					A						A	T	A					A	G		M	4
16											A	T				C		A			L	2
17											A	T						A	G		L	1
18											A	T				C		A	G		L	2
19											A					C		A	G		L	1
20												T				C		A	G		L	1
21																C		A	G		L	1

A list of all hybrid end sequences isolated on pRZ5451 (Figure 1) that papillate more frequently than wt IE, with the other end being wt OE, and the EK54 Tnp being expressed from pFMA187. All hybrid end sequences as well as wt OE are shown only at those positions where they differ from wt IE.

[a] *trans* papillation levels of wt IE, wt OE and hybrid end sequences are classified as follows: VL, very low; L, low; M, medium; and H, high.

[b] Although mutants 12 and 13 were not found in this experiment, they were found in *cis* papillation screening (data not shown).

identical bp and seven non-identical bp between them (Table 1). Several genetic analyses of the OE and IE sequences have been carried out (Dodson & Berg, 1985; Phadnis & Berg, 1987; Tomcsanyi & Berg, 1989; Makris *et al*., 1988). In a *dam* background, where the inhibitory effect of Dam methylation on IE utilization is taken out of consideration, the majority of point mutations of the OE and IE resulted in reduced transposition frequencies, except for OE mutants 12A → T, 14A → G and 17A → T, which resulted, respectively, in 1.5-fold, 1.8-fold and 1.8-fold enhancement of the wt OE transposition frequency (reviewed by Berg, 1989). In other words, no point mutant (or double mutant) of OE or IE that was tested resulted in a significantly increased *in vivo* transposition frequency.

Biochemical analyses of the OE base-pairs were carried out by Jilk *et al*. (1996) in order to identify OE base-pairs important for wt Tnp binding. Through a hydroxyl radical missing nucleoside interference experiment, a dimethyl sulfate interference experiment, and gel retardation assays with single point mutants of OE, they concluded that positions 6 to 9 and 13 to 19 of OE are involved in wt Tnp binding, while positions 1 to 5 and 10 to 12 appear to be involved only in secondary reactions in wt Tnp-mediated transposition.

The specific DNA binding domain of Tnp is believed to be at the N terminus for the following reasons. First, Inh, which is identical to Tnp except that it lacks the N-terminal 55 amino acid residues of Tnp, does not bind OE in the same mobility shift assay where Tnp–OE complexes are observed (de la Cruz *et al*., 1993; Weinreich *et al*., 1994). Second, several point mutations as well as short deletion mutations altering the very N terminus of Tnp resulted in significantly reduced or undetectable OE binding activity (Weinreich *et al*., 1993).

As the result of a study more precisely defining the Tnp DNA binding domain, we recently reported the isolation of hypertransposing Tnp mutants in the N terminus with an enhanced OE binding activity. Of particular interest is EK54 Tnp, which manifests an altered preference for OE *versus* IE compared to wt Tnp (Zhou & Reznikoff, 1997). *In vivo*, EK54 increases transposition eight- to ninefold compared to wt Tnp when two OEs were used, but decreases transposition fivefold compared to wt Tnp when two IEs or one IE and one OE were used. A C-terminal truncated version of Tnp (called TnpΔ369), which is defective in dimerization and binds OE or IE as monomers, was used in an *in vitro* gel retardation assay to study the OE and IE binding activity of wt, EK54 and other mutant Tnp proteins. EK54 TnpΔ369 was found to bind OE better than wt TnpΔ369, but it bound IE less well than wt TnpΔ369. We proposed that Lys54 either makes a more favorable interaction(s) with one or more OE specific base(s) than Glu54, but a less favorable interaction with the corresponding IE specific base(s), and/or interacts more favorably with the OE DNA backbone, and less favorably with the IE DNA backbone, than Glu54.

In this study we report our attempt to identify base position(s) in OE that are critical for the OE/IE discrimination manifested by EK54 Tnp. We began by randomizing the seven non-identical positions between OE and IE, and asked which of these positions, when changed from an IE-specific base-pair into an OE-specific base-pair, results in the elevated transposition observed with OE, in the presence of EK54 Tnp. We found three base positions, 10, 11 and 12, of OE to be critical for EK54 recognition. Moreover, these three positions appear to function cooperatively, their individual effects being non-additive. *In vitro* gel retardation assays verified these conclusions, and further suggested that at least some of the nucleotides in the 19 bp Tn*5* ends affect other functions besides Tnp binding, directly or indirectly. However, we performed a missing nucleoside EK54 Tnp-binding experiment, which indicated that OE base-pairs 10, 11 and 12 are not directly involved in the EK54–OE binding reaction. We also discovered two IE variants that are transposed at extraordinarily high frequencies with EK54 Tnp, making them ideal for use in *in vitro* transposition studies (I. Y. Goryshin & W. S. Reznikoff, unpublished results), as well as providing a powerful tool for making transposon-based random insertions in the genome. Those two IE variants also transposed better than the natural ends (IE and OE) with wt Tnp, demonstrating that the wt IE & OE sequences contribute to the low frequency of wt Tn*5* transposition.

Results

Identification of OE-specific nucleotide pairs important for Lys54 recognition

Tn*5* Tnp binds specifically to the OE and the IE (de la Cruz *et al*., 1993; Jilk *et al*., 1996). A mutant of Tnp, EK54, significantly altered the OE *versus* IE preference compared to wt Tnp both *in vivo* (transposition frequency) and *in vitro* (DNA binding) (Zhou & Reznikoff, 1997). EK54 increased OE recognition and decreased IE recognition compared to wt Tnp. In order to identify the nucleotide position(s) in OE that makes it a more favorable target for EK54 Tnp than IE, we screened a population of mosaic end sequences that contained either the OE nucleotide pair or the IE nucleotide pair at each of the seven non-identical positions between OE and IE. As described in Materials and Methods, when EK54 Tnp was expressed either *in cis* or *in trans*, a population of plasmids carrying a wt OE and a mosaic end sequence flanking the *lacZYA* gene was screened for hyper-papillation levels compared to a control construct in which the mosaic end sequence was substituted by a wt IE. The *lacZ* gene is defective in its transcriptional and translational initiation signals, so that only those events of transposition into an actively transcribed and translated region in the correct reading frame will

result in blue papillae on otherwise white colonies. The papillation level of each colony thus provides an estimate of the transposition frequency of the mosaic end sequence in that colony's cells. All screening was carried out in a *dam* strain, JCM101/pOXgen (Zhou & Reznikoff, 1997), to avoid the effect of Dam methylation complicating our comparison, since the IE contains two Dam methylation sites, and methylation of IE is known to inhibit IE-mediated transposition (Yin *et al*., 1988; Zhou, 1997).

A total of 1575 colonies was screened. Forty hyper-papillating mutants were isolated *in cis* (from plasmid pRZ5421) and 65 isolated *in trans* (from pRZ5451; see Figure 1 for the plasmids). After retransformation to confirm that the hyper-papillation phenotype was conferred by the mutant plasmid, the mosaic end in each mutant was sequenced. All mutant ends isolated *in trans* are listed in Table 1. Two additional moderate mutants (4/10/11/12/18 and 4/10/11/12/15/17) and eight weak mutants were isolated *in cis* (10/11/15, 4/10/11/15, 4/10/11/15/17, 4/10/11/15/17/18, 10/12/17, 10/12/18, 10/12/15/17/18, and 10/15/17; Zhou, 1997). Many mutant ends were isolated multiple times. Since there were a total of $2^7 = 128$ different possible mutant ends in the initial screened population, by screening over 1000 colonies there would only be less than a 5% chance that we could have missed examining any of the 128 possible mutants: $1 - [1 - (1 - 1/128)^{1000}]^{128} = 5\%$. Thus, it is likely that we have found all the end mutants that could result in hyper-papillation (compared to IE) in the designed population of mosaic ends.

The most prominent conclusion from Table 1 is that all mosaic ends with OE nucleotide pairs in positions 10, 11 and 12 result in high papillation levels comparable or nearly comparable with OE. Lower levels of papillation (but still higher than IE) are observed when only one or two of these three positions are OE nucleotide pairs. Thus, these three positions appear to be critical in making the end sequence favorable for interaction with the EK54 Tnp.

Another conclusion that can be drawn from the *trans* data in Table 1 is that the A·T base-pair at position 4 of OE seems to inhibit transposition to a certain degree (compare mutant 1 with 9, 2 with 10, 3 with 11, 5 with 14, 6 with 15, and 7 with OE). In other words, OE is not the most optimal end sequence for EK54 Tnp. The degree of increase in transposition frequency brought about by positions 10, 11 and 12 of OE is masked in part by the adverse effect of position 4 of OE. This prompted us to analyze the individual contributions of each of the seven non-identical positions between OE and IE. This is reported in the next section.

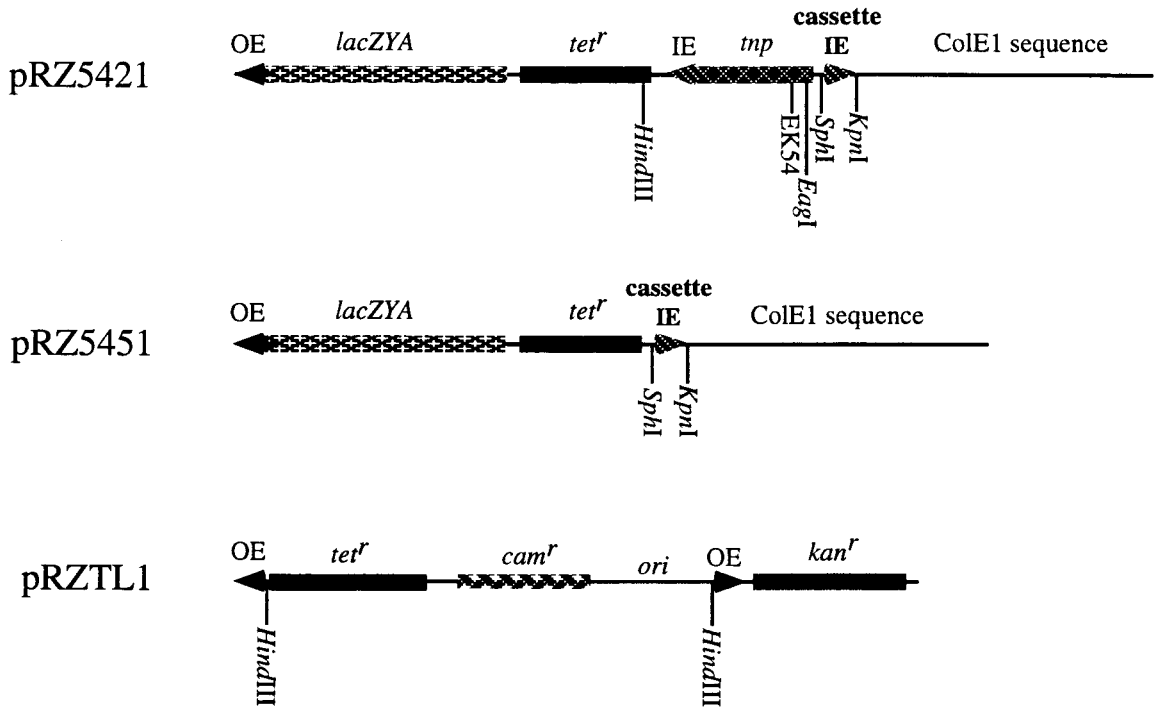

Figure 1. Schematic drawing of three plasmids used in this study. When representing OE or IE, the arrow points towards the outside of the transposable element (from base-pair 19 to base-pair 1). The *Hin*dIII site and the *Eag*I site in pRZ5421 correspond to the same sites in pRZ1496 where a deletion was made to generate pRZ5451 (see Materials and Methods), resulting in the almost complete loss of the *tnp* gene. The cassette IE in pRZ5421 and pRZ5451 is replaced with a population of hybrid cassette ends between the *Sph*I/*Kpn*I sites in papillation screening. The position of the EK54 mutation in the *tnp* gene in pRZ5421 is indicated. In pRZTL1, the two *Hin*dIII sites used in the cloning to construct pRZTL2, 3 and 4, are indicated. *tet*, tetracycline; *cam*, chloramphenicol; *kan*, kanamycin; *tnp*, transposase.

A third conclusion that can be drawn from the *trans* data in Table 1 is that the G·C base-pair at position 18 of OE seems to inhibit transposition when in the context of the OE 10, 11 and 12 sequence (compare mutant 1 with mutant 8). Below we will show that 18G has a sequence context-dependent effect; i.e. 18G by itself in an IE sequence context enhances the frequency of transposition.

Assessment of the effects of each of the seven non-identical positions on the EK54 and wt Tnp papillation frequencies

In order to better understand the effect of each of the seven non-identical positions between OE and IE on EK54 as well as wt Tnp recognition, we introduced into the IE sequence (on pRZ5451) single point mutations at these seven positions, changing one nucleotide pair at a time into the corresponding OE base-pair. We also introduced a double mutation at positions 10 and 11. The *trans* papillation frequencies of these end mutants were compared with wt IE, wt OE and mutant 1 of Table 1 (henceforth designated IE10/11/12). EK54 or wt Tnp was supplied *in trans* from plasmid pFMA187 (Goryshin *et al.*, 1994). Figures 2 and 3 show the time course of papillation with EK54 and wt Tnp, respectively.

With EK54 Tnp (Figure 2), IE10/11/12 transposes more frequently than any other ends including OE, consistent with Table 1. Strikingly, although the double mutant IE10/11 shows a mild increase in transposition compared to IE (three- to fourfold), none of the three single mutations at positions 10, 11 and 12 caused any increase in transposition compared to IE. In fact, mutation 10A, and especially 12A, decreased transposition. This means that positions 10, 11 and 12 in the triple mutant and in OE function cooperatively in bringing about the highly favorable interaction with EK54 Tnp.

With wt Tnp (Figure 3), IE10/11/12 transposes only slightly better than IE, indicating that the dramatic preference for IE10/11/12 compared to IE seen in Figure 2 was characteristic of EK54 Tnp. Interestingly, transposition with both EK54 and wt Tnp increased two- to threefold over IE with

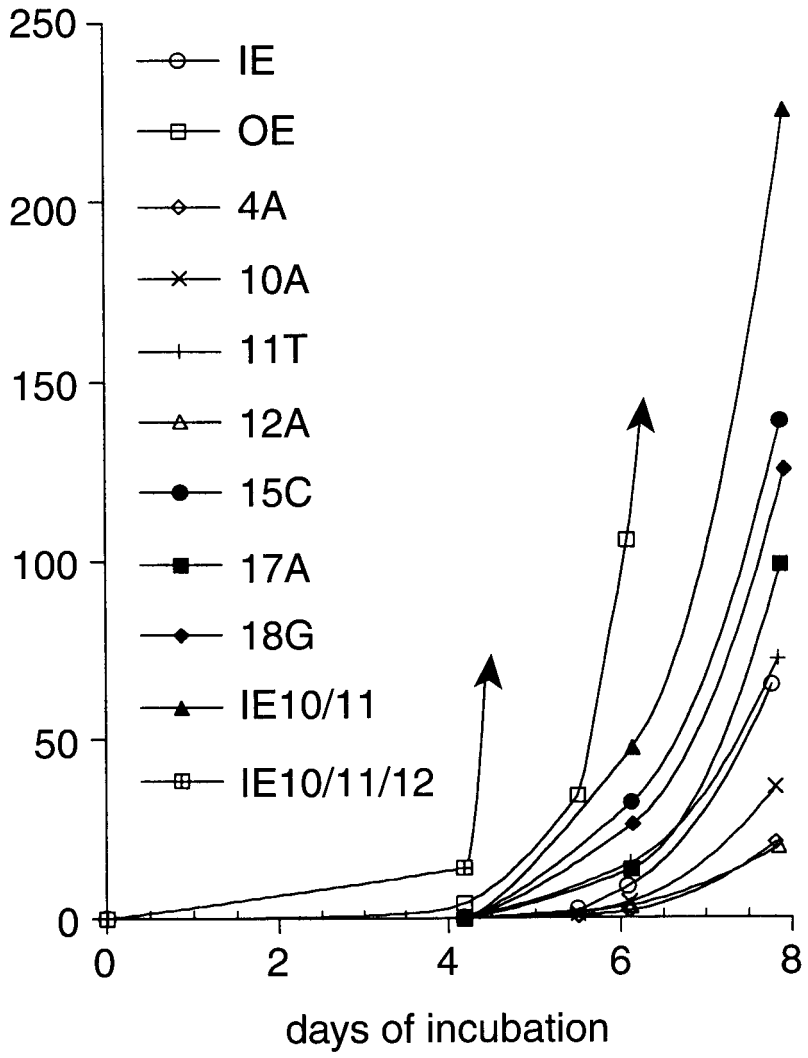

Figure 2. Time-course of *trans* papillation assay (using pRZ5451 derivatives, Figure 1) of IE mutants compared with wt OE and IE, in the presence of EK54 Tnp (expressed from pFMA187). The number of papillae per colony, averaged from five representative colonies, is shown on the ordinate for each end type. Each single point mutant of IE is named by the position of the mutation in IE followed by the OE base now present at that position. Mutant IE10/11 means IE10A/11 T, mutant IE10/11/12 means IE10A/11 T/12A. The mutants are named by the same strand as shown in Table 1. IE10/11/12 generated too many papillae to count within 5.5 days after plating.

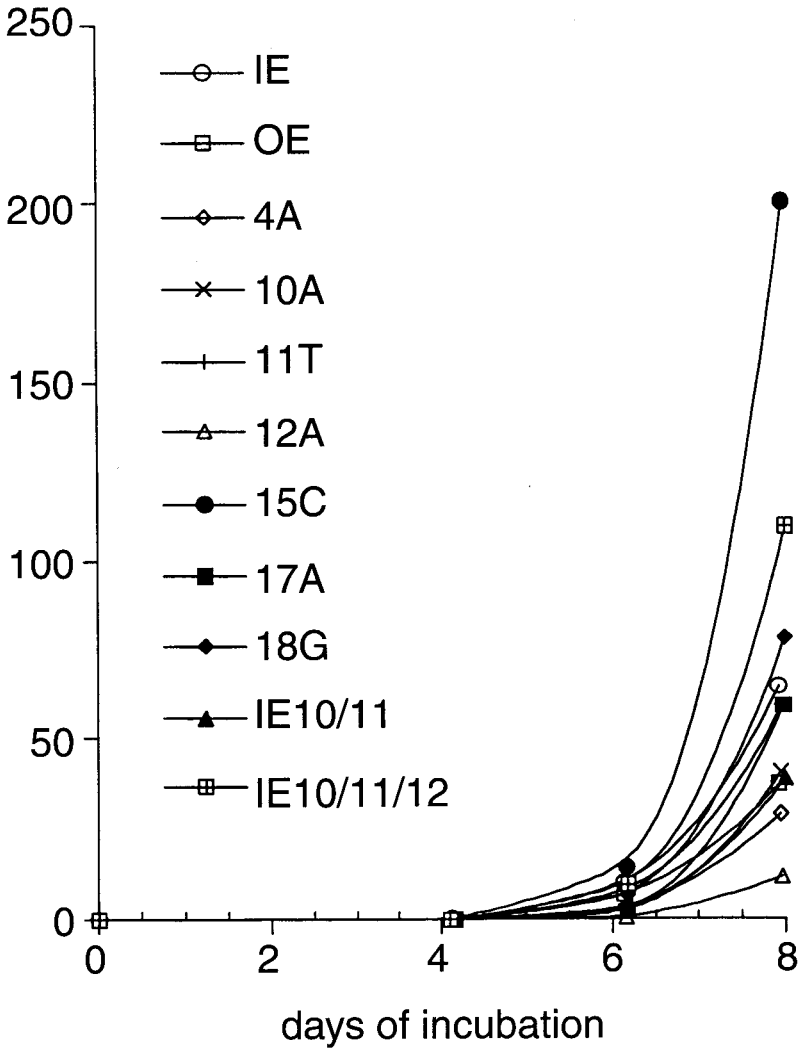

Figure 3. Time course of *trans* papillation assay (using pRZ5451 derivatives, Figure 1) of IE mutants compared with wt OE and IE, in the presence of wt Tnp (expressed from pFMA187). See Figure 2 legend for explanations.

mutation 15C. For wt Tnp, 15C transposes better than both IE and OE, indicating that evolution has resulted in end sequences that are transposed at suboptimal frequencies.

The mutant 18G has an interesting effect on EK54 Tnp-mediated transposition, enhancing the frequency relative to the IE sequence. This is in contrast to the opposite effect of 18G seen in the context of 10A/11 T/12A in Table 1.

Comparison of *in vivo* transposition frequencies with symmetrical end sequences

The above papillation assays were all performed with plasmids in which one transposon end was wt OE, with the other being a mutant end. In order to observe the full effect of some of the mutant ends, we introduced them into the plasmid pRZTL1 (Figure 1), in which the two ends can be replaced by mutant ends symmetrically. We replaced the two OEs in pRZTL1 with two IEs, two IE10/11/12 sequences, or two IE10/11/12/15 sequences, respectively, resulting in plasmids pRZTL2, pRZTL3 and pRZTL4. The mutant IE10/11/12/15 was included because both IE10/11/12 and 15C transposed at a higher frequency than IE in the papillation assay (see above), and we wished to determine the effects of the combined mutants on transposition. Transposition was measured as the frequency of Tet^r colonies, as described in Materials and Methods.

Figure 4 shows the transposition results measured by such a tetracycline resistance assay. EK54 decreases IE transposition compared to wt Tnp, but dramatically increases IE10/11/12 transposition compared to wt Tnp. This is consistent with the result from the papillation assays demonstrating that there is a specific preference of EK54 Tnp for end sequences containing the OE base-pairs at positions 10, 11 and 12.

With EK54 Tnp, the IE10/11/12 construct transposes much more frequently than the OE construct (32-fold), also consistent with the papillation assays (Figure 2). IE10/11/12/15 transposes still more frequently than IE10/11/12, yielding two mutant ends that transpose far more efficiently than the natural ends (OE and IE) when used in combination with EK54 Tnp.

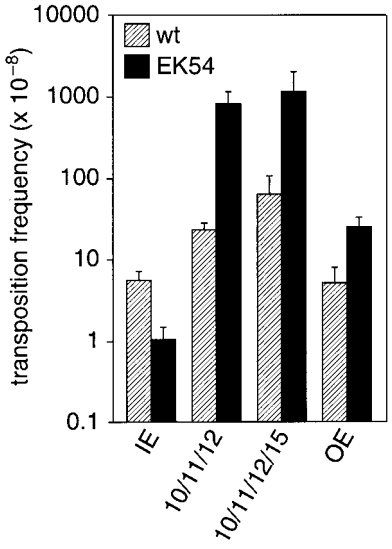

Figure 4. Comparison of *in vivo* transposition frequencies of symmetrical end mutations measured by the tetracycline resistance assay, using pRZTL1, 2, 3 or 4 (Figure 1), shown in a logarithmic scale. The frequencies shown are the average of five samples originated from five separate single colonies after transformation. The mutant ends are identified as the OE-like changes superimposed on the IE sequence.

The apparent *in vivo* transposition frequencies reported in Figure 4 for constructs carrying two OEs or two IEs are all 300 to 500-fold lower than that reported by Zhou & Reznikoff (1997). The reason for this is that the tetracycline resistance assay of transposition frequency employed in this study examines only those transposition events in which the transposon is inserted in the correct orientation into a dispensable actively transcribed region, so that the promoter-defective *tet*^{r} gene is expressed sufficiently to confer a Tet^r phenotype. This kind of transposition event is only a small fraction of total transposition events. Thus, the transposition frequencies measured by this assay are lower than the actual transposition frequencies.

Comparison of *in vitro* DNA binding activity of IE mutants

To answer the question of whether the changes in transposition efficiency caused by the IE mutants are due at least in part to changes in their specific binding affinity to EK54 and wt Tnp, we carried out an *in vitro* gel retardation assay. Since the C-terminal region of Tnp is involved in Tnp dimerization (Weinreich *et al.*, 1993; York & Reznikoff, 1996), which appears to influence Tnp binding to OE (de la Cruz *et al.*, 1993; York & Reznikoff, 1996), we simplified our *in vitro* DNA binding study by using a C-terminal truncated version (Δ369) of wt or EK54 Tnp. TnpΔ369 binds to OE as a monomer (York & Reznikoff, 1996), so the abundance of Tnp–OE complexes is a simple reflection of the Tnp–OE affinity and is not complicated by the dimerization reaction.

Increasing amounts of purified EK54 or wt TnpΔ369 were incubated with a 55 bp labeled DNA fragment (0.4 nM) containing one of the end sequences, before being loaded onto a native polyacrylamide gel. The percentage of the total labeled DNA of each lane that appeared in the TnpΔ369–DNA complex was quantified (Figure 5A and B). The following conclusions can be drawn.

(1) IE10/11/12 is bound to EK54 TnpΔ369 more efficiently than IE but it is bound to wt TnpΔ369 less efficiently than IE. Thus, the increased *in vivo* transposition frequency of IE10/11/12 compared to IE, observed in the presence of EK54 Tnp, was at least in part due to its increased specific binding affinity to the EK54 Tnp. The fact that the single mutants 10A, 11 T and 12A and the double mutant 10/11 all resulted in lower binding than IE suggests that the cooperativity among base-pairs 10, 11 and 12 observed in the presence of EK54 Tnp in *in vivo* transposition assays is also due to their cooperative effects on the specific binding of EK54 Tnp.

(2) IE10/11/12 is bound to EK54 TnpΔ369 less efficiently than OE, yet it transposes with EK54 Tnp much more efficiently than OE in *in vivo* transposition assays (Figures 2 and 4). This implies that the DNA binding affinity as reflected by the gel retardation assay does not always correlate with the overall transposition efficiency, and that at least some of the 19 nucleotide pairs of the ends directly or indirectly affect some other aspect(s) of transposition besides Tnp binding.

(3) With both EK54 and wt TnpΔ369, 15C is bound better than IE, and IE10/11/12/15 is bound better than IE10/11/12. This is consistent with the *in vivo* papillation results of Figures 2 and 3, indicating that mutation 15C increases both wt and EK54 Tnp mediated transposition through enhanced DNA binding.

Analysis of OE bp directly involved in EK54 Tnp binding

We have previously used a hydroxyl-radical interference experiment to define the bases in OE directly involved in wt Tnp binding (Jilk *et al.*, 1996). Positions 10/11/12 were not involved. We repeated these experiments to determine whether any of these positions are important for EK54 Tnp binding. Figure 6 shows that EK54 and wt Tnp binding are affected identically by missing nucleosides in the (−) strand of OE and, in particular, removal of nucleosides at positions 10, 11 and 12 failed to inhibit EK54 Tnp binding. We also observed no difference in EK54 and wt Tnp binding dependencies when the (+) strand was examined (data not shown).

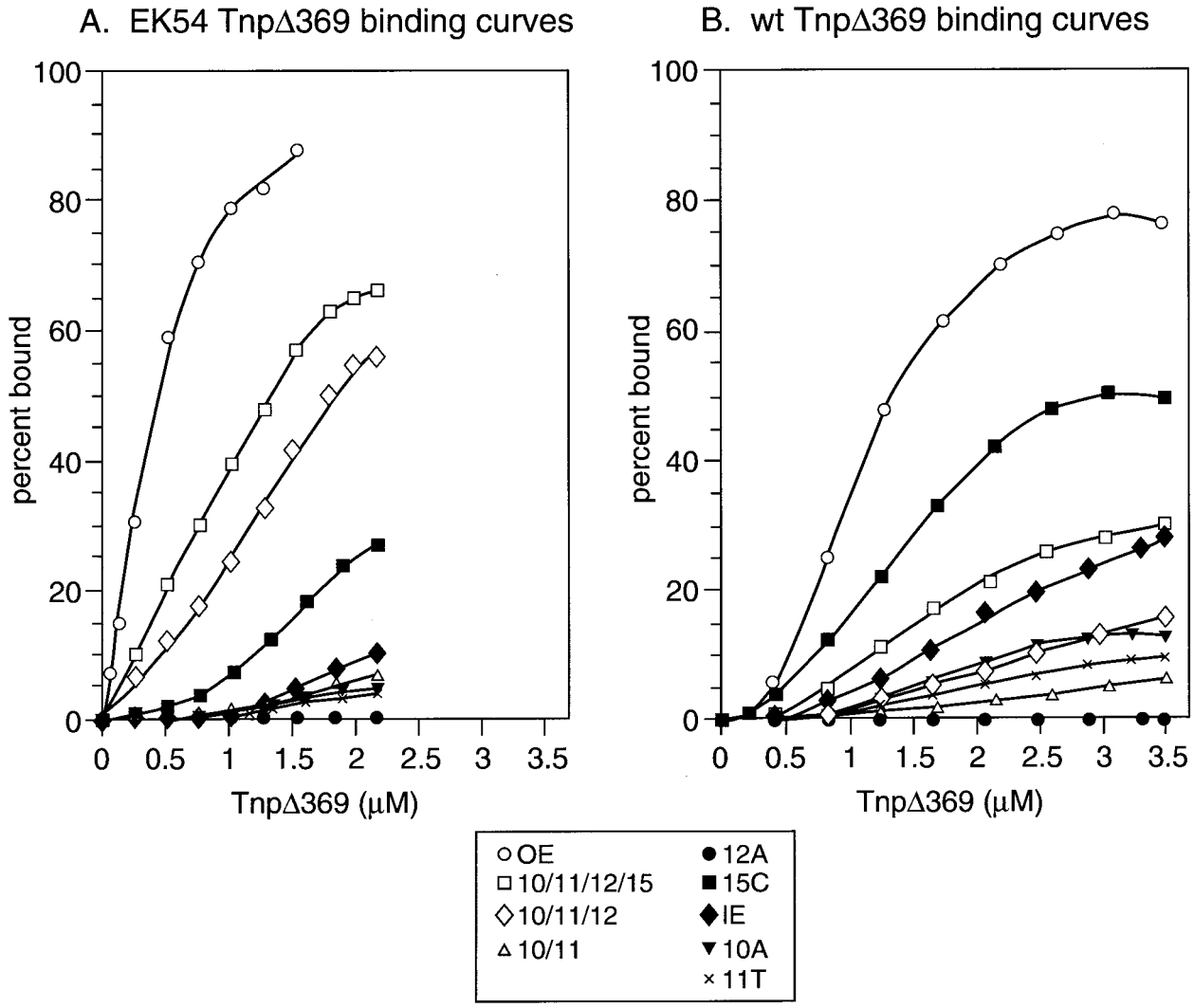

Figure 5. Binding curves of EK54 TnpΔ369 (A) and wt TnpΔ369 (B) with a 55 bp labeled DNA fragment (0.4 nM) containing different end sequences, measured by gel retardation assays. Each curve is the average result of at least two experiments. % bound, the percentage of the total labeled DNA of each lane that is found in the TnpΔ369–DNA complex, measured by a Phosphorimager (Molecular Dynamics). The mutant DNA fragments are identified as the OE-like changes superimposed on the IE sequence.

Discussion

Tn*5* wt Tnp recognizes two forms of end DNA sequences, OE and IE. Binding of Tnp to the ends is believed to be the first committed step in the transposition reaction. In order to study the specific interactions between the Tnp and the ends, we have previously isolated a mutant Tnp, EK54, which increased OE recognition while decreasing IE recognition compared to wt Tnp, in both *in vivo* transposition assays and *in vitro* DNA binding assays (Zhou & Reznikoff, 1997). We reasoned that, since EK54 Tnp has an altered preference for OE *versus* IE, it is possible that residue Lys54 makes a specific favorable interaction(s) with some OE-specific base pair(s), and/or that EK54 Tnp is specifically enhanced in recognizing the OE DNA conformation, but decreased in recognizing the IE DNA conformation.

In this study we tried to dissect the seven non-identical base-pairs between OE and IE, to find out which one(s) of them is (are) critical for the enhanced recognition of OE by EK54 Tnp. By screening a library of mosaic ends with either OE or IE base-pairs at the seven non-identical positions, in the presence of EK54 Tnp, we discovered that all the mosaic ends that could result in significantly higher papillation frequencies than wt IE contain OE base-pairs at positions 10, 11 and 12; and all possible mosaic ends that contain OE base-pairs at positions 10, 11 and 12 resulted in significantly higher papillation frequencies than wt IE (Table 1). Apparently, these three base-pairs together lead to favorable DNA recognition by EK54 Tnp.

The specificity at each base-pair position in the DNA that contributes to the sequence-specific protein–DNA interaction can be explained in three

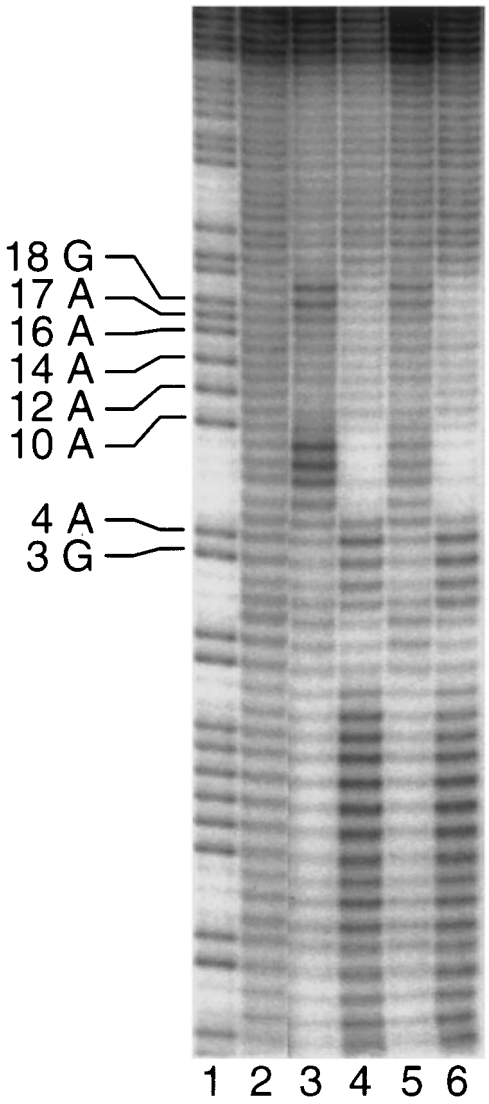

Figure 6. Missing nucleoside analysis of OE bases directly involved in wt and EK54 Tnp binding. Radiolabeled OE-containing DNA was treated in a Fenton reaction to remove random nucleosides, incubated with sufficient purified Tnp to bind approximately 50% of the DNA and then the bound and the free DNAs were separated by polyacrylamide gel electrophoresis as described in Materials and Methods. The DNAs were purified and then analyzed by sequencing gel electrophoresis. The DNA analyzed on this gel was the (−) strand OE DNA as defined by Jilk *et al.* (1996). Lane 1 is the results of the Maxam–Gilbert A + G reaction; positions in the OE sequence are marked. Lane 2 is the hydroxyl-radical cleavage control. Lane 3 is free DNA from the EK54 Tnp binding reaction. Lane 4 is EK54 Tnp-bound DNA. Lane 5 is free DNA from the wt Tnp binding reaction. Lane 6 is wt Tnp-bound DNA. Base positions required for binding are seen as enhancements in the free lanes and diminishments in the bound lanes.

ways. The base-pair can contribute a specific functional group(s) that will interact with a certain functional group(s) in the protein (Seeman *et al.*, 1976). The base-pair can affect the recognition of neighboring base-pair(s) by facilitating or allowing the protein's DNA binding domain to form the appropriate induced fit. It is likely that some positions in the *lac* operator manifest such an effect (Mossing & Record, 1985; Frank *et al.*, 1997). The base-pair can effect the local DNA conformation. The latter case includes the effects of DNA sequence on the ease with which the DNA site can be overtwisted into the proper configuration for complex formation (for example, see Koudelka *et al.* (1988) on phage 434 operator-repressor interaction); the effects of DNA sequence on phosphate positions, which are known to exhibit sequence-dependent variations (Calladine, 1982; Dickerson, 1983); and effects of DNA sequence on the energy of base-pair opening.

In an attempt to determine whether or not OE base-pairs 10, 11 and 12 provide a specific interaction with EK54 Tnp, we conducted a hydroxyl-radical interference experiment (also called a missing nucleoside experiment) using wt and EK54 Tnp and OE DNA (Figure 6). Removal of any one nucleoside in positions 10, 11 and 12 did not appear to significantly affect EK54 Tnp binding. In fact the hydroxyl radical results appeared identical for EK 54 and wt Tnp. This indicates that there is no direct functional group interaction between the EK54 Tnp and base pairs 10, 11 and 12.

Although further investigation is needed to elaborate the biochemical mechanism of the favorable DNA recognition by EK54 Tnp caused by the three OE base-pairs, the fact that all three positions (10, 11 and 12) are necessary to achieve the specific recognition by EK54 Tnp suggests that a specific alteration on local DNA conformation is involved. For instance, such an alteration may change the flexibility of the DNA, resulting in an altered strength of Tnp–DNA interaction. Positions 10, 11 and 12 are located between the two domains in the OE sequence needed for wt Tnp binding (Jilk *et al.*, 1996; see above). This is reminiscent of the phage 434 operator-repressor interaction described by Koudelka *et al.* (1988), where the central non-contacted base-pairs affect repressor binding affinity by altering the flexibility of the operator DNA.

Interestingly, the effects of base-pairs 10, 11 and 12 are non-additive. Single base-pair changes from the IE to the OE base-pair at each of these three positions generally resulted in worse recognition than IE both in transposition (Figure 2) and in Tnp binding (Figure 5). Whatever makes the DNA more favorably recognized by EK54 Tnp exists only when all three positions are OE base-pairs. Non-additivity of the effects of changing individual base-pairs in a protein binding site has been reported. Disproportionately large reductions in protein–DNA complex stability are often observed for the first substitution in a consensus site or half site, while subsequent substitutions are less detrimental than the first. In other words, the sum of the effects of single mutations from a consensus site are often bigger than the effects of multiple mutations. This behavior has been observed by

Mossing & Record (1985) and Frank *et al*. (1997) for *lac* repressor, by Lesser *et al*. (1990) for *Eco*RI endonuclease, and by Ades & Sauer (1995) for engrailed homeodomain. This phenomenon has also been observed for specific protein–RNA interactions (Pütz *et al*., 1993). However, the phenomenon that we observed is unique in that the effects of single mutations 10A, 11 T and 12A all go in the opposite direction than the effect of the triple mutant IE10/11/12, whereas in the other systems the effects of single mutations are in the same direction as the multiple mutant, though the values are non-additive.

The only other incidence in which a double alteration of a nucleic acid sequence resulted in the opposite effect from the component single alterations was observed by Reich & Sweetnam (1994), on the effects of guanosine replacements by inosine upon free energy changes of the melting of double-stranded oligonucleotides. Two single G to I replacements, one on each strand of a double-stranded 14-mer, individually gave $\Delta\Delta G^{\circ}$ values of 1.6 and 0.0 kcal/mole, respectively, compared to the unreplaced 14-mer. The oligonucleotide with the double replacements gave a $\Delta\Delta G^{\circ}$ value of −1.6 kcal/mole. Though Reich & Sweetnam used inosine replacements in this study, we believe that other base substitutions could also have significant structural consequences on the DNA, and it is conceivable that in some cases a combination of base substitutions may result in completely different structural alterations compared to that caused by any of the component single substitutions.

The requirement that 10/11/12 must function in concert to stimulate EK54-Tnp mediated transposition has an interesting implication with regard to evolutionary theory. Our observation is a molecular version of the ''selective surface'' phenomenon (Wright, 1982) in which species evolution can be viewed as a movement from lower selective value peaks to higher peaks through steps that have negative selective values. The key proposal is that sufficient random processes must occur to allow the population to proceed genetically against selective pressure. In our case, we have so greatly enhanced the random process through our high level of synthetic mutagenesis that no negative selective states are experienced. In nature this might be accomplished while the transposon is passively inherited for many generations in the host.

Although we studied the effects of changes of position 18 in less detail, these also appear to be interesting. An 18 C → G change reduces the EK54 Tnp-mediated transposition frequency in the context of 10/11/12 mosaic end sequence (Table 1) but enhances transposition in the context of the IE sequence (Figure 2). This type of context-specific effect on the Tnp–DNA interaction could be explained by Tnp undergoing a form of induced fit upon binding to the end sequence. We posit that the Tnp N-terminal sequence that contains the end-sequence binding domain assumes a stable structure only upon binding to the end sequence. This model is similar to that proposed for the *lac* repressor–*lac* operator interaction (Mossing & Record, 1985; Lewis *et al*., 1996; Frank *et al*., 1997). Consistent with this hypothesis are the observations that Tn*5* Tnp DNA N-terminal residues are quite sensitive to proteolytic attack both *in vivo* (Weinreich *et al*., 1993) and, in the absence of DNA, *in vitro* (L. Braam & W. S. Reznikoff, unpublished observations).

As mentioned in Results, both IE10/11/12 and IE10/11/12/15 are bound to EK54 TnpΔ369 less well than OE, yet they transpose with EK54 Tnp much better than OE in *in vivo* transposition assays (Figures 2 and 4). This indicates that at least some of the 19 base-pairs of the ends affect other aspect(s) of transposition than Tnp binding (such as synapse formation, cleavage and strand transfer reactions, etc.). Thus, just like Tnp, the Tn*5* end sequences are also involved in more than one function, or the regulation of those functions, confirming the conclusions of Jilk *et al*. (1996). A multifunctional role for transposon end sequences has also been found for several other systems (Huisman *et al*., 1989; Ichikawa *et al*., 1990; Arciszewska & Craig, 1991; Kuo *et al*., 1991; Derbyshire & Grindley, 1992).

One hypothesis that could explain why we were able to find end sequence variants, along with Tnp variants, that transpose much more efficiently than the natural ends, is that evolution has directed the natural ends (OE and IE) to be suboptimal for transposition, possibly to enable them to respond to subtle physiological signals through regulations such as *dam* methylation (Yin *et al*., 1988; Makris *et al*., 1988) and DnaA level (Yin & Reznikoff, 1987), instead of being highly efficient in transposition.

Jilk *et al*. (1996) reported *in vitro* analyses of the OE base-pairs involved in wt Tnp binding (as mentioned in the Introduction). They concluded that positions 6 to 9 and 13 to 19 of OE are involved in wt Tnp binding, while positions 1 to 5 and 10 to 12 appear to be involved only in secondary reactions in transposition. Our finding (Figure 5B), that single, double and triple mutations at positions 10, 11 and 12 of IE all decreased wt TnpΔ369 binding (especially the single mutation 12A, which has non-detectable binding activity), suggests that these three positions of IE do affect wt Tnp binding. The discrepancies of our results and those of Jilk *et al*. may be due to the fact that Jilk *et al*. observed the effects of single point mutations and modifications of the OE sequence, while we observed the effects of mutations of the IE sequence. In other words, the context of IE makes the 12 T → A change intolerable for wt Tnp binding, but in the context of OE when all positions including 10 and 11 are already the OE base-pairs, and binding is already much better than IE, the 12A → T change does not really improve binding much more.

Finally, the discovery that the end variants IE10/11/12 and IE10/11/12/15 in combination with EK54 Tnp provides an extraordinarily high transposition frequency may prove very useful in *in vitro* transposition studies where efficient transposition is desired. Preliminary results using an *in vitro* transposition reaction already support this notion (I. Y. Goryshin & W. S. Reznikoff, unpublished data). Another important application of our super-transposing end variant-Tnp variant combination will be the use in transposon-directed random insertions into the genome in transposon tagging.

Materials and Methods

Media and reagents

The media used in these studies have been described by Zhou & Reznikoff (1997). Native *Pfu* DNA polymerase (for PCR amplifications) was purchased from Stratagene. Phage T4 DNA ligase was from Promega. Sequenase 2.0 was from United States Biochemical Corp.

Bacterial strains and plasmids

All bacterial strains used in this study were derivatives of *E. coli* K-12, except for the B-strain BL21(DE3) pLysS (Studier *et al*., 1990). The structures of some of the plasmids are shown in Figure 1. All plasmids used in this study encoding either the wt or the EK54 Tnp or TnpΔ369 contain the MA56 mutation, which eliminates the translation initiation of the *inh* gene, but otherwise does not seem to have a significant effect on the functions of the Tnp protein (Wiegand & Reznikoff, 1992). The results of all cloning and site-directed mutagenesis manipulations were confirmed by DNA sequence analysis.

pRZ1495 (Makris *et al*., 1988) is a papillation plasmid containing a fixed OE and a replaceable OE (also called OE cassette) flanking the *lacZYA* genes, a tetracycline resistance gene, and the *tnp* gene. pRZ1496 is identical to pRZ1495 except that an IE cassette replaced the OE cassette. We constructed pRZ5450 and pRZ5451 (Figure 1) (both used in *trans* papillation assays) by deleting a *Hin*dIII-*Eag*I fragment (nucleotides 9112 to 12083) containing almost the entire *tnp* gene, from pRZ1495 and pRZ1496, respectively, followed by filling the ends with T4 polynucleotide kinase and ligating the ends. An 11 amino acid residue long peptide is encoded by the resulting plasmids, only nine residues of which are derived from the N-terminal portion of the *tnp* gene.

pRZ5418 and pRZ5421 (Figure 1) (both used in *cis* papillation assays) were constructed by introducing the EK54/MA56 double mutation into pRZ1495 and pRZ1496, respectively. First, we amplified a 413 bp PCR fragment encoding a segment of the *tnp* gene that contains the EK54/MA56 double mutation, using pRZ5412 (Zhou & Reznikoff, 1997) as template. This PCR fragment was then used as a primer in a subsequent PCR reaction, using pRZ1496 as template, and an oligonucleotide complementary to the IE cassette of pRZ1496 as a second primer. The resulting final PCR product was digested with *Sph*I and *Not*I, and used to replace the corresponding *Sph*I-*Not*I fragment in pRZ1495 and pRZ1496.

In papillation screening, the two strands of the hybrid end cassette were first synthesized as degenerate oligonucleotides: 5′-ctg(t/a)ctctt(g/a)(a/t)(t/a)ca(g/c)a(t/a)(c/g)ttgatcccggatccgcatg-3′ and 5′-cggatccgggatcaa(g/c)(a/t)t(c/g)tg(a/t)(t/a)(c/t)aagag(a/t)caggtac-3′, with the two bases at each of those positions indicated by parentheses occurring at a 50% chance for each. The two degenerate oligonucleotides were annealed to each other, forming a 37 bp degenerate cassette flanked by an *Sph*I overhang on one side and a *Kpn*I overhang on the other. The degenerate cassette was used to replace the IE cassette in pRZ5421 and pRZ5451, between the corresponding *Sph*I and *Kpn*I sites, creating plasmids homologous to pRZ5421 and pRZ5451, respectively, but carrying mosaic ends.

The following single point mutants of IE: 12A, 15C and 17A, and the double mutant IE10/11 were isolated in pRZ5451 as follows. The two strands of the hybrid end cassette were again sythesized as degenerate oligonucleotides of the same sequence as given in the above paragraph, except that the two bases in each parenthesis occurred at a different ratio: the IE base occurred 6/7 of the time, and the OE base 1/7 of the time. The rest of the cloning procedures were identical to those above. The IE single point mutants 4A, 10A, 11 T and 18G of pRZ5451 were obtained as follows. Four oligonucleotides, 5′-attggtacctgActcttgatcag-3′, 5′-attggtacctgtctcttAatcagatctt-3′, 5′-attggtacctgtctcttgTtcagatcttg-3′ and 5′-attggtacctgtctcttgatcagatGttgatcccggatccgc-3′, carrying single mutations 4A, 10A, 11 T and 18G, respectively, were synthesized. PCR reactions were performed, using pRZ1496 as template, one of the four mutagenic oligonucleotides as one primer, and oligonucleotide S16 (Goryshin & Reznikoff, unpublished results), which anneals to nucleotides 12054 to 12073 in pRZ1496, as the other primer. PCR products were double-digested with *Sph*I/*Kpn*I, and the resulting 37 bp *Sph*I-*Kpn*I fragments carrying the single mutations were annealed to the *Sph*I-*Kpn*I large fragment of pRZ5451.

pRZTL1 (Figure 1; Goryshin & Reznikoff, 1998) was used in the tetracycline resistance assay of transposition frequencies. In pRZTL1, two OEs flank a tetracycline resistance gene, a chloramphenicol resistance gene and the origin of plasmid replication derived from p15. The vector sequence outside of the two OEs is only 1.3 kb long, and contains a kanamycin resistance gene (*kan*^{r}). In order to replace the two OEs in pRZTL1 symmetrically with two IEs, two IE10/11/12 sequences, or two IE10/11/12/15 sequences, we synthesized one oligonucleotide for each type of substitution. The oligonucleotides were: 5′-acagaagctt(agatctgatcaagagacag)tcgacctgcagggggggg-3′ for IE substitution, 5′-acagaagctt(agatctgtataagagacag)tcgacctgcagggggggg-3′ for IE10/11/12 substitution, and 5′-acagaagctt(agatgtgtataagagacag)tcgacctgcagggggggg-3′ for IE10/11/12/15 substitution, with the parenthesis indicating the 19 bp end sequences. Three PCR reactions were carried out, using pRZTL1 as template and each of the three oligonucleotides as primers. The plasmid was designed so that the same mutagenic primer can anneal to both OE sequences, yielding a product of 1.3 kb that contains the *kan*^{r} gene in the vector sequence between the two OEs. This PCR product was digested with *Hin*dIII, exposing *Hin*dIII overhangs on both sides, and used to replace the corresponding *Hin*dIII fragment in pRZTL1. The resulting plasmids are called pRZTL2 (for IE substitutions), pRZTL3 (for IE10/11/12 substitutions) and pRZTL4 (for IE10/11/12/15 substitutions).

pUC19 derivatives were used in the *in vitro* DNA binding assays. The *Sph*I-*Kpn*I cassettes from pRZ5451 derivatives containing the following mutants of IE, 10A, 11 T, 12A, 15C, IE10/11, IE10/11/12 and IE10/11/12/15, were used to replace the short *Sph*I-*Kpn*I fragment in pUC19. The resulting plasmids are homologous to pRZ7067OE (Makris, 1989) and pRZ7067IE (Jilk *et al*., 1996), and are considered pRZ7067IE derivatives. A 55 bp long *Eco*RI-*Hin*dIII fragment from these plasmids containing the end sequences is used in gel retardation assays with wt or EK54 TnpΔ369.

Papillation assay for *in vivo* screening of mosaic end sequences

The papillation assay (initially described by Krebs & Reznikoff, 1988) was used to screen for mosaic end sequences between OE and IE that result in a higher transposition frequency compared to IE, when transposition is driven by EK54 Tnp. In *cis* papillation screening, a population of plasmids homologous to pRZ5421, carrying a wt OE and a mosaic end sequence (see above), flanking the *lacZYA* genes, a tetracycline resistance gene and the EK54 *tnp* gene, was electroporated into a *lac-dam* strain JCM101/pOXgen (Zhou & Reznikoff, 1997), plated on Trp^--XG-PG agar, and incubated at 37°C. The rate of appearance of these papillae in each colony provides an estimate of the transposition frequency of the mosaic end sequence in that colony's cells. Colonies that resulted in hyperpapillation levels compared to the control colonies carrying pRZ5421 (in which a wt IE cassette is in place of the hybrid end cassette) were isolated. The *trans* papillation screening was carried out exactly as the *cis* papillation screening, except that the EK54 Tnp was expressed from another plasmid pFMA187 (Goryshin *et al*., 1994), and the transposon ends were carried on a population of plasmids derived from pRZ5451 (see above).

In vivo tetracycline resistance assay of transposition frequencies

JCM101/pOXgen was transformed with pRZTL1, 2, 3 or 4 carrying two OEs, two IEs, two IE10/11/12 sequences or two IE10/11/12/15 sequences, respectively, then transformed with pRZ5412 carrying either wt or EK54 *tnp*. Five colonies from each transformation were streaked out onto the same kind of LB plates containing chloramphenicol and ampicillin (LB cam-amp). A single colony from each streak was inoculated into 4 ml LB culture and shaken overnight at 37°C. The overnight cultures were diluted and plated on LB plates containing tetracycline (LB tet) and on LB cam-amp plates. In each of the pRZTL plasmids the two ends flank a tet^r gene that is defective at its −35 promoter region, so that only a transposition event into an actively transcribed region can result in the expression of the tet^r gene. The number of colonies that grow on LB tet plates divided by the number of colonies that grow on LB cam-amp plates gives the transposition frequency of each sample.

In vitro DNA binding assays

Purification of EK54 and wt TnpΔ369 as well as gel retardation assays with DNA fragments containing OE, IE and various IE derivatives were carried out as described by Zhou & Reznikoff (1997). The 55 bp long DNA fragments containing the end sequences were obtained from pRZ7067 derivatives (see above) by a *Eco*RI/*Hin*dIII double digest.

Hydroxyl-radical interference experiment

The hydroxyl-radical interference experiment was performed by procedures similar to that described by Jilk *et al*. (1996), except as described below. The target DNA was a 90 bp OE-containing fragment generated by PCR. The DNA was single end-labeled by using a (5′ ^{32}P)-labeled primer during the PCR reaction. The binding reactions were carried out as described by Zhou & Reznikoff (1997), adding sufficent Tnp to bind approximately 50% of the DNA. Following electrophoresis, the image was analyzed using a Molecular Dynamics Phosphor Imaging system.

Acknowledgments

We thank Igor Y. Goryshin for technical help and the use of plasmid pRZTL1, Thomas Record Jr for interesting discussions, Dona York and Keith Derbyshire for helpful suggestions regarding the manuscript, and Barb Shriver for preparing bacterial medium and many other solutions. This work was supported by NIH grant GM50692.

References

Ades, S. E. & Sauer, R. T. (1995). Specificity of minor-groove and major-groove interactions in a homeodomain-DNA complex. *Biochemistry*, **34**, 14601–14608.

Arciszewska, L. K. & Craig, N. L. (1991). Interaction of the Tn7–encoded transposition protein TnsB with the ends of the transposon. *Nucl. Acids Res.* **19**, 5021–5029.

Berg, D. E. (1989). Transposon Tn5. In *Mobile DNA* (Berg, D. E. & Howe, M. M., eds), pp. 185–210, American Society for Microbiology, Washington, DC.

Calladine, C. R. (1982). Mechanics of sequence-dependent stacking of bases in *B*-DNA. *J. Mol. Biol*, **161**, 343–352.

de la Cruz, N. B., Weinreich, M. D., Wiegand, T. W., Krebs, M. P. & Reznikoff, W. S. (1993). Characterization of the Tn5 transposase and inhibitor proteins: a model for the inhibition of transposition. *J. Bacteriol.* **175**, 6932–6938.

Derbyshire, K. M. & Grindley, N. D. F. (1992). Binding of the IS903 transposase to its inverted repeat *in vitro*. *EMBO J*, **11**, 3449–3455.

Dickerson, R. E. (1983). Base sequence and helix structure variation in *B* and *A* DNA. *J. Mol. Biol*, **166**, 419–441.

Dodson, K. W. & Berg, D. E. (1989). Saturation mutagenesis of the inside end of insertion sequence IS50. *Gene*, **85**, 75–81.

Frank, D. E., Saecker, R. M., Bond, J. P., Capp, M. W., Tsodikov, O. V., Melcher, S. E., Levandoski, M. M. & Record, M. T., Jr (1997). Thermodynamics of the interactions of *lac* repressor with variants of the symmetric *lac* operator: effects of converting a consensus site to a non-specific site. *J. Mol. Biol.* **267**, 1186–1206.

Goryshin, I. Y., Kil, Y. V. & Reznikoff, W. S. (1994). DNA length, bending, and twisting constraints on IS50 transposition. *Proc. Natl Acad. Sci. USA,* **91**, 10834–10838.

Goryshin, I. Y. & Reznikoff, W. S. (1998). Tn*5 in vitro* transposition. *J. Biol. Chem.* In the press.

Huisman, O., Errada, P. R., Signon, L. & Kleckner, N. (1989). Mutational analysis of IS*10*'s outside end. *EMBO J.* **8**, 2101–2109.

Ichikawa, H., Ikeda, K., Amemura, J. & Ohtsubo, E. (1990). Two domains in the terminal inverted-repeat sequence of transposon Tn*3*. *Gene,* **86**, 11–17.

Jilk, R. A., York, D. & Reznikoff, W. S. (1996). The organization of the outside end of transposon Tn*5*. *J. Bacteriol.* **178**, 1671–1679.

Koudelka, G. B., Harbury, P., Harrison, S. C. & Ptashne, M. (1988). DNA twisting and the affinity of bacteriophage 434 operator for bacteriophage 434 repressor. *Proc. Natl Acad. Sci. USA,* **85**, 4633–4637.

Krebs, M. P. & Reznikoff, W. S. (1988). Use of a Tn*5* derivative that creates *lacZ* translational fusions to obtain a transposition mutant. *Gene,* **63**, 277–285.

Kuo, C.-F., Zou, A., Jayaram, M., Getzoff, E. & Harshey, R. (1991). DNA-protein complexes during attachment-site synapsis in Mu DNA transposition. *EMBO J.* **10**, 1585–1591.

Lesser, D. R., Kurpiewski, M. R. & Jen-Jacobson, L. (1990). The energetic basis of specificity in the EcoRI endonuclease-DNA interaction. *Science,* **250**, 776–786.

Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G. & Lu, P. (1996). Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science,* **271**, 1247–1254.

Makris, J. C. (1989). Modulation of IS*50* and Tn*5* transposition. PhD thesis, University of Wisconsin-Madison.

Makris, J. C., Nordmann, P. L. & Reznikoff, W. S. (1988). Mutational analysis of insertion sequence 50 (IS*50*) and transposon 5 (Tn*5*) ends. *Proc. Natl Acad. Sci. USA,* **85**, 2224–2228.

Mossing, M. C. & Record, M. T., Jr (1985). Thermodynamic origins of specificity in the *lac* repressor-operator interactions. *J. Mol. Biol.* **186**, 295–305.

Phadnis, S. H. & Berg, D. E. (1987). Identification of base pairs in the IS*50* O end needed for IS*50* and Tn*5* transposition. *Proc. Natl Acad. Sci. USA,* **84**, 9118–9122.

Pütz, J., Puglisi, J. D., Florentz, C. & Giegé, R. (1993). Additive, cooperative and anti-cooperative effects between identity nucleotides of a tRNA. *EMBO J.* **12**, 2949–2957.

Reich, N. O. & Sweetnam, K. R. (1994). Sequence-dependent effects on DNA stabilty resulting from guanosine replacements by inosine. *Nucl. Acids Res.* **22**, 2089–2093.

Reznikoff, W. S. (1993). The Tn*5* transposon. *Annu. Rev. Microbiol.* **47**, 945–963.

Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA,* **73**, 804–808.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60–89.

Tomcsanyi, T. & Berg, D. E. (1989). Effects of adenine (Dam) methylation on transposition activity of O end mutants of IS*50*. *J. Mol. Biol.* **209**, 191–193.

Weinreich, M. D., Mahnke-Braam, L. & Reznikoff, W. S. (1993). A functional analysis of the Tn*5* transposase, identification of domains required for DNA binding and mutlimerization. *J. Mol. Biol.* **241**, 166–177.

Weinreich, M. D., Gasch, A. & Reznikoff, W. S. (1994). Evidence that the *cis* preference of the Tn*5* transposase is caused by nonproductive multimerization. *Genes Dev.* **8**, 2363–2374.

Wiegand, T. W. & Reznikoff, W. S. (1992). Characterization of two hypertransposing Tn*5* mutants. *J. Bacteriol.* **174**, 1229–1239.

Wright, S. (1982). The shifting balance theory and macroevolution. *Annu. Rev. Genet.* **16**, 1–19.

Yin, J. C. & Reznikoff, W. S. (1987). *dnaA*, an essential host gene, and Tn*5* transposition. *J. Bacteriol.* **169**, 1812–1817.

Yin, J. C., Krebs, M. P. & Reznikoff, W. S. (1988). Effect of Dam methylation on Tn*5* transposition. *J. Mol. Biol.* **199**, 35–45.

York, D. & Reznikoff, W. S. (1996). Purification and biochemical analyses of a monomeric form of Tn*5* transposase. *Nucl. Acid. Res.* **24**, 3790–3796.

Zhou, M. (1997). Interactions between Tn*5* transposase and Tn*5* end sequences: study of the DNA binding step. PhD thesis, University of Wisconsin-Madison.

Zhou, M. & Reznikoff, W. S. (1997). Tn*5* transposase mutants that alter DNA binding specificity. *J. Mol. Biol.* **271**, 362–373.

Edited by G. Smith

(Received 27 July 1997; received in revised form 8 December 1997; accepted 8 December 1997)

Note added in proof: The proposed induced fit of the Tn*5* transposase–DNA complex is consistent with the recently reported sequence-dependent deformation of the DNA bound by the DNA-binding domain of Tc*3* transposase (van Puorderoyan, G., Ketting, R. F., Perrakis, A., Plasterk, R. H. A. & Sixma, T. K. (1997). Crystal structure of the specific DNA-binding domain of Tc*3* transposase of *C. elegans* in complex with transposon DNA. *EMBO J.* **16**, 6044–6054).