Tipping the balance between replicative and simple transposition

Norma P. Tavakoli¹ and Keith M. Derbyshire^{1,2,3}

¹Division of Infectious Disease, Wadsworth Center, New York State Department of Health and ²Department of Biomedical Sciences, School of Public Health, State University of New York at Albany, Albany, NY 12201-2002, USA

³Corresponding author at: David Axelrod Institute, NYS Department of Health, PO Box 22002, Albany, NY 12201-2002, USA
e-mail: keith.derbyshire@wadsworth.org

The bacterial insertion sequence IS903 has the unusual ability to transpose both replicatively and non-replicatively. The majority of products are simple insertions, while co-integrates, the product of replicative transposition, occur at a low frequency (<0.1% of simple insertions). In order to define the critical steps that determine the outcome of IS903 transposition, we have isolated mutants that specifically increase the rate of replicative transposition. Here we show that the nucleotide immediately flanking the transposon influences both overall transposition frequency and co-integrate formation. In particular, when the 3'-flanking nucleotide is A, co-integrates are increased 500-fold compared with a 3' C. In addition, we have isolated five transposase mutants that increase replicative transposition. These residues are close to the catalytic residues and are thus likely to be part of the active site. These are the first transposase mutations described that affect the product of transposition. Our results are consistent with the hypothesis that a delay in cleavage of the 5'-flanking DNA will increase the effective half-life of the 3'-nicked transposon intermediate and consequently enhance co-integrate formation.

Keywords: co-integrate formation/insertion sequence/replicative transposition/transposase

Introduction

Bacterial transposons move from one genomic location to another by a process that is independent of DNA homology. These transposons move by one of two pathways, non-replicative (also called simple insertion, conservative and cut-and-paste) and replicative (Figure 1a and b), to generate transposition products called simple insertions or co-integrates, respectively (Mizuuchi, 1992; Craig, 1996; Haren *et al.*, 1999). Most transposons move exclusively by one pathway or the other. IS903 and IS1 are unusual as they have been shown to carry out both simple insertion and replicative transposition (Ohtsubo *et al.*, 1981; Weinert *et al.*, 1984; Turlan and Chandler, 1995).

Although the products of non-replicative and replicative transposition look very different, extensive *in vitro* studies

have confirmed models based on *in vivo* experiments to show that the biochemical steps in both pathways are remarkably similar (see below and reviews by Mizuuchi, 1997; Haren *et al.*, 1999). The major distinguishing feature between the two pathways is whether a double-strand break or a single-strand nick occurs at the ends of the transposon before integration (Turlan and Chandler, 2000). A double-strand break effectively removes all connections with the donor DNA and thus precludes co-integrate formation. In contrast, a nick allows maintenance of connections to both donor and target replicons after integration, and, following replication, results in formation of a co-integrate.

The initial step of both transposition pathways involves the generation of a 3' OH at the transposon termini (Figure 2a). In the simple insertion pathway, subsequent cleavage of the 5'-flanking DNA generates an excised transposon (Figure 2b). The 3' OH ends of the excised transposon then act as nucleophiles in a concerted strand transfer reaction that results in integration of the transposon into a target (Figure 2d). The target DNA is cleaved in a staggered manner and the short single-stranded regions that flank the newly inserted transposon are repaired by the host replication apparatus to generate short direct repeats, a signature of the transposition reaction. The mechanism of cleavage of the 5'-flanking strand varies. For IS10 and IS50, the excised transposon is generated via a hairpin intermediate (Kennedy *et al.*, 1998; Bhasin *et al.*, 1999). The hairpin is formed when the 3' OH at each transposon end attacks the phosphodiester backbone at the 5' ends of the transposon. This transesterification not only results in the joining of the 3' OH to the 5' end of the transposon, but also releases flanking donor DNA. The hairpin is resolved by a second, transposase-mediated nick (hydrolysis) to generate the excised transposon intermediate with 3' OH termini. Tn7 also generates double-strand breaks at each transposon end, but by a different mechanism. In contrast to IS10 and IS50, Tn7 utilizes two Tn7-encoded proteins to process each flanking strand (Sarnovsky *et al.*, 1996). TnsB, a transposase-related protein, nicks the 3' ends of the transposon. TnsA, a site-specific endonuclease, cleaves the 5' ends to generate an excised transposon, which is then inserted into a target (May and Craig, 1996; Hickman *et al.*, 2000).

In the replicative pathway (Figure 2e and f), as characterized for bacteriophage Mu, second strand cleavage does not occur. Instead, strand transfer results in fusion of target and donor DNAs to form a strand transfer intermediate, which consists of a copy of the transposon joined to both donor and target DNA (Figure 2e) (Craigie and Mizuuchi, 1985; Naigamwalla and Chaconas, 1997). The exposed 3' OHs of the target in this intermediate are used to prime DNA synthesis, allowing replication of the entire transposon to form a co-integrate structure

Fig. 1. Alternate pathways of transposition for IS903. (a) Simple insertion of IS903 from pKD498 into the conjugative plasmid, pOX38. These transposition events can be detected by transfer of chloramphenicol resistance (Cm^r) into a suitable recipient using a mating-out assay. (b) Replicative transposition results in a co-integrate structure that fuses donor and target replicons via two copies of the transposon. Co-integrates can be identified by co-transfer of both Cm^r and kanamycin resistance (Km^r) by pOX38. Note that co-integrates may also form following simple transposition from a plasmid dimer, but that in all these experiments propagation of plasmid DNAs in *recA*⁻ strains precluded dimer formation. The transposon on pKD498 is indicated by the heavy line and the inverted repeats are indicated by triangles. Flanking donor DNA and target DNA are indicated by thin lines.

(Figure 2f). The co-integrate is composed of the donor and target replicons fused by two copies of the transposable element in direct orientation (Figure 1b).

To understand the unique properties of IS903 that allow it to move by both replicative and non-replicative pathways, we have isolated both transposase and DNA mutants that increase the frequency of replicative transposition. IS903 is an insertion sequence of 1057 bp that has inverted repeats (IRs) of 18 bp and encodes a transposase of 307 amino acids (Grindley and Joyce, 1981). The IS903 transposase contains the highly conserved DDE and YREK motifs common to the IS4 family of transposons, which are thought to play a key role in catalysis (Rezsohazy *et al.*, 1993; Tavakoli *et al.*, 1997). The 18 bp IR has been subdivided into two functional regions based on genetic and biochemical analyses (Derbyshire *et al.*, 1987; Derbyshire and Grindley, 1992). The inner 12 bp (bp 6–18) are essential for binding of transposase, which makes both major and minor groove contacts to this region. The outer domain (bp 1–3) is involved in a step subsequent to binding, probably catalysis, since this is the site of DNA cleavage and is thus likely to be an intimate part of the transposase active site. This is also consistent with the phenotype of mutations in the outer domain, which considerably reduce transposition frequency but do not affect transposase binding (Derbyshire *et al.*, 1987; Derbyshire and Grindley, 1992). Furthermore, mutations at the transposon termini affect the outcome of the transposition reaction (Tavakoli and Derbyshire, 1999). Substitution of the wild-type C at the transposon termini reduces transposition frequency by 2–4 orders of magnitude, but increases co-integrate formation to 30–50%. This suggested that the nucleotides at the transposon termini affect both first and second strand processing, and

Fig. 2. Summary of the biochemical steps that occur in simple and replicative transposition pathways. Nicking at the 3' ends is the initial step in both pathways (a). In the simple insertion pathway, this is followed by cleavage of the 5'-flanking DNA (b), which generates an excised transposon. Interaction with a target (c) allows strand transfer to occur, which results in a simple insertion (d). Replicative transposition (a, e and f) occurs via a strand transfer reaction (e) involving the nicked transposon and a target to generate a strand transfer intermediate. Replication of this intermediate using the exposed 3' OHs in the target DNA as priming sites (f) results in duplication of the transposon and a co-integrate structure. IRs are represented by filled triangles, the transposon DNA is shown as a thick line, flanking and target DNAs as thin lines, and cleavage sites by small vertical arrows.

prompted us to investigate the effect of other DNA and protein mutations on co-integrate formation in IS903 transposition.

Results

Effect of flanking DNA on transposition and co-integrate formation

Point mutations at the termini of the IR decrease overall transposition frequency and increase co-integrate formation, suggesting that the terminal nucleotide plays a role in the transposon excision process (Tavakoli and Derbyshire, 1999). We hypothesized that the nucleotide immediately flanking the IR might also affect the reaction, as transposase mediates cleavage of the phosphodiester backbone at the transposon-flank junction. Four transposon vectors were made that differed only by the nucleotide flanking both transposon ends. Mating-out assays were performed to determine the effect on transposition and co-integrate formation (Table I).

Transposition frequencies were high, as expected for a wild-type transposon but, most remarkably, a 3'-flanking A increased co-integrate formation to 5% (500-fold higher than a C). To determine whether this effect required an A at both flanks, derivatives were made in which the

Fig. 3. Alignment of amino acids around the conserved active site residues of IS903 and IS10 transposases. The DDE (bold) and YREK motifs are shown in the center line and aligned with the conserved residues in the IS903 and IS10 transposases (boxed). The IS903 residues considered candidates for involvement in strand cleavage and transfer (affecting co-integrate formation in IS903) are shown above the alignment. Amino acid substitutions discussed in more detail in the text are indicated. The IS10 mutants discussed in the text, and their phenotypes, are indicated below the alignment. The vertical dots align non-conserved residues that were mutated.

Table I. Effect of flanking DNA on co-integrate formation

IS903 inverted repeat			
+1 -1 -2 -3			
GATTTATTCAACAAAGCC-C T G 3'			
CTAAATAAGTTGTTTCGG-G A C 5'			
Transposase	Flanking DNA (-1 -2 -3)	Transposition frequency	% co-integrates
w/t	CTG	1.0×10^{-3}	0.01
w/t	TTG	1.8×10^{-3}	0.35
w/t	ATG	3.1×10^{-3}	5.4
w/t	GTG	2.0×10^{-4}	<0.02
V119A	CTG	1.9×10^{-2}	0.01
V119A	TTG	8.2×10^{-3}	0.35
V119A	ATG	1.3×10^{-2}	4.1
V119A	GTG	5.5×10^{-3}	0.15
w/t	ACG	3.0×10^{-3}	4.2
w/t	ATG	3.1×10^{-3}	5.4
w/t	AAg	3.0×10^{-3}	1.7
w/t	AGG	2.9×10^{-3}	13.7
w/t	AGC	2.2×10^{-3}	6.6
w/t	AGT	3.0×10^{-3}	6.2
w/t	AGa	3.0×10^{-3}	1.6
w/t	AGG	3.0×10^{-3}	13.7

The sequence of the 18 bp IS903 IR with flanking DNA (-1, -2 and -3) is shown at the top of the table. The flanking nucleotide altered for each assay is indicated in bold and underlined. w/t, wild-type transposase. V119A, transposase with a valine to alanine substitution at amino acid 119. Transposition frequencies are the average of at least six experiments.

transposon was flanked by a 3' C and a 3' A. The co-integrate levels were reduced to levels similar to those of a construct with C at both flanking sequences (data not shown). A 3'-flanking C at one transposon end, therefore, has a dominant effect over an A at the other, suggesting that a double-strand break is made efficiently at the C nucleotide end, which commits the transposon to the simple insertion pathway. We note that a G nucleotide flanking the 3' ends of the transposon (G at -1) consistently lowered the transposition frequency (5-fold compared with other flanking nucleotides). We have observed this subtle reduction in transposition frequency with other transposon constructs flanked by a 3' G (see Table IV and data not shown).

Table II. Transposition frequency and level of co-integrate formation of IS903 mutants

IS903 Tnp residue	Amino acid change	Transposition frequency	% co-integrates	Equivalent IS10 Tnp residue
w/t		1.0×10^{-3}	0.01	
S122	A	2.8×10^{-6}	—	W98R
	R	1.0×10^{-4}	—	(strand transfer defect)
	G	$<1.0 \times 10^{-8}$	—	
L125	A	$<1.0 \times 10^{-8}$	—	I101S
	S	$<1.0 \times 10^{-8}$	—	(strand transfer defect)
G194	S	1.5×10^{-4}	—	A162T
	A	4.0×10^{-6}	2	(hypernicking mutant)
	T	2.0×10^{-6}	—	
	I	$<1.0 \times 10^{-8}$	—	
	R	$<1.0 \times 10^{-8}$	—	
	N	$<1.0 \times 10^{-8}$	—	
R199	A	3.9×10^{-7}	—	P167S
	S	$<1.0 \times 10^{-8}$	—	(hairpin defect)
	P	$<1.0 \times 10^{-8}$	—	
S256	M	8.5×10^{-6}	30	M289I
	A	5.0×10^{-6}	2	(hypernicking mutant)
	V	2.1×10^{-6}	1	
	T	3.8×10^{-7}	10	
	D	$<1.0 \times 10^{-8}$	—	
	E	$<1.0 \times 10^{-8}$	—	
	K	$<1.0 \times 10^{-8}$	—	
	G	$<1.0 \times 10^{-8}$	—	
	R	$<1.0 \times 10^{-8}$	—	
	I	$<1.0 \times 10^{-8}$	—	

Co-integrates are indicated as a percentage of the transposition events and are an average of at least six assays. Equivalent IS10 transposase (Tnp) mutants and their relevant phenotypes are also listed (see text). A dash (—) indicates that no co-integrates were observed or that the transposition frequency was too low to allow detection of co-integrates.

To demonstrate further that these mutations had increased the frequency of replicative transposition, we utilized a hyperactive transposase mutant (V119A) to increase the overall transposition frequency (Tavakoli and Derbyshire, 1999). This mutation is just two amino acids downstream from the first aspartate of the catalytic DDE motif (Figure 3) and increases transposition frequency ~20-fold. Transposons that incorporated the transposase mutant V119A together with each possible flanking nucleotide were constructed, and co-integrate formation was measured (Table I). As expected, the transposition

Table III. Transposition rescue by the hyperactive transposase mutant V119A does not alter co-integrate formation

IS903 Tnp	Transposition frequency	% co-integrates
w/t	1.0×10^{-3}	0.01
V119A	1.9×10^{-2}	0.01
G194A	4.0×10^{-6}	2.0
G194A + V119A	1.5×10^{-2}	1.0
S256M	8.5×10^{-6}	30.0
S256M + V119A	5.2×10^{-2}	16.0
G194A + S256M	2.7×10^{-6}	30.0
R199A	3.9×10^{-7}	–
R199A + V119A	1.3×10^{-3}	0.5
R199S	$<1.0 \times 10^{-8}$	–
R199S + V119A	2.1×10^{-4}	0.3
R199P	$<1.0 \times 10^{-8}$	–
R199P + V119A	$<1.0 \times 10^{-8}$	–
Y252A	8.0×10^{-7}	–
Y252A + V119A	5.5×10^{-4}	0.5
R255A	4.0×10^{-8}	–
R255A + V119A	7.6×10^{-5}	9.0
K266A	$<1.0 \times 10^{-8}$	–
K266A + V119A	5.6×10^{-5}	<0.05
K266G	$<1.0 \times 10^{-8}$	–
K266G + V119A	4.4×10^{-4}	<0.01

A dash (–) indicates that no co-integrates were observed or that the transposition frequency was too low to allow detection of co-integrates.

frequency of each derivative containing the V119A substitution was increased. Most importantly, the high level of co-integrate formation with a flanking 3' A was maintained at a level similar to that mediated by wild-type transposase. These results, combined with the effect we had observed previously for alterations at the transposon termini, show that the dinucleotide at the transposon–donor junction can have a dramatic effect on the outcome of transposition.

Nucleotides at positions –2 and –3 have a minor effect on co-integrate formation

To extend the analysis of the flanking DNA sequence, nucleotides at positions –2 and –3 were varied. Since the possible number of combinations of nucleotides is high, we limited the combinations to nucleotides that showed the highest percentage of co-integrate formation. Accordingly, constructs were made that retained A at –1, which allowed for the detection of both positive and negative effects on replicative transposition. Only small changes were observed in the level of co-integrate formation when varying the nucleotide at –2 (Table I). The highest level observed was with G at nucleotide –2 (13.7%). Four additional constructs were tested that had A and G at positions –1 and –2, but varied the nucleotides at –3 (Table I). Co-integrate formation varied from 1.6 to 13.7%. These results confirm that the nucleotide at position –1 has the most dramatic effect on the outcome of the transposition reaction (500-fold for a 3' A) and that nucleotides further away play only a minor role in modifying the outcome of transposition.

Transposase mutants that affect co-integrate formation

The phenotypes of the DNA mutants were consistent with the hypothesis that they were delaying the cleavage of the flanking DNA, thereby increasing the effective half-life of

Table IV. Combination of transposase mutants with all four possible flanking nucleotides

Transposase	Flanking nucleotide (–1)	Transposition frequency	% co-integrates
w/t	C	1.0×10^{-3}	0.01
w/t	T	1.8×10^{-3}	0.35
w/t	A	3.1×10^{-3}	5.4
w/t	G	2.0×10^{-4}	<0.02
V119A	C	1.9×10^{-2}	0.01
V119A	T	8.2×10^{-3}	0.35
V119A	A	1.3×10^{-2}	4.1
V119A	G	5.5×10^{-3}	0.15
G194A	C	4.0×10^{-6}	2.0
G194A + V119A	C	2.8×10^{-3}	1.5
G194A + V119A	T	1.7×10^{-3}	2.1
G194A + V119A	A	1.3×10^{-3}	13.0
G194A + V119A	G	3.2×10^{-4}	6.0
S256M	C	8.5×10^{-6}	30.0
S256M + V119A	C	2.7×10^{-3}	15.0
S256M + V119A	T	4.8×10^{-4}	13.0
S256M + V119A	A	2.9×10^{-3}	23.0
S256M + V119A	G	4.8×10^{-4}	14.0
K266G	C	$<1.0 \times 10^{-8}$	–
K266G + V119A	C	4.4×10^{-4}	<0.01
K266G + V119A	T	1.0×10^{-6}	–
K266G + V119A	A	5.2×10^{-4}	7.0
K266G + V119A	G	8.0×10^{-6}	–

A dash (–) indicates that no co-integrates were observed or the transposition frequency was too low to allow detection of co-integrates. Data for wild-type and V119A transposases are taken from Table I.

the nicked transposon intermediate (Figure 2a) and the chance of target capture to form a strand transfer intermediate (Figure 2e). Based on this hypothesis, we predicted that any transposase mutant that prolonged the half-life of the nicked transposon substrate should also elevate replicative transposition. To identify candidate residues, we considered a collection of IS10 transposase mutants that were defective in transposon excision and strand transfer. We focused on five mutants that showed defects in second strand cleavage and strand transfer (W98R, I101S), hairpin formation (P167S) or exhibited a hypernicking phenotype (A162T and M289I) (Haniford *et al.*, 1989; Bolland and Kleckner, 1995; Kennedy and Haniford, 1996; Kennedy *et al.*, 1998). In particular, the phenotype of the latter three mutants would be predicted to increase the effective half-life of the nicked intermediate and thus enhance co-integrate formation, although this had not been tested for IS10. By aligning the IS903 and IS10 transposase sequences, we identified the equivalent residues in the IS903 transposase (Figure 3; S122, L125, G194, R199 and S256). Although these residues are not conserved, they are located in close proximity to the highly conserved active site residues [in the case of IS903, D121, D193 and E259 (Tavakoli *et al.*, 1997); and for IS10, D97, D161 and E292 (Bolland and Kleckner, 1996)] (Figure 3). Site-specific mutagenesis of candidate residues was used to introduce the desired mutations. Mutant transposases were cloned into a transposon donor vector and assayed by *in vivo* mating-out assays. The overall transposition frequencies and levels of co-integrate formation are shown in Table II.

Substitutions at S122 and L125 reduced transposition to undetectable levels or, when transposition was detected, the transconjugants resulted from simple insertions and were not studied further. In contrast, substitutions at

residues G194 and S256 increased co-integrate formation. Five mutant transposases G194A, S256M, S256A, S256V and S256T, showed an increase in co-integrate formation when compared with wild-type transposase. Substitutions of G194 to S and T reduced the transposition frequency but did not affect co-integrate formation, while substitutions to I, R and N reduced transposition to undetectable levels. Other substitutions of S256 to D, E, K, G, R and I also abolished transposition.

Two mutants, G194A and S256M, were characterized further as they generated co-integrates at the highest frequency. Transposition mediated by each mutant transposase was reduced by 2–3 orders of magnitude compared with wild-type, and so we considered the possibility that the elevated formation of co-integrates might simply reflect a reduction in simple transposition events. Double mutants were made with the hyperactive transposase mutant V119A in order to rescue transposition. The V119A mutation rescued the transposition frequency of each mutant but maintained the same level of co-integrate formation (Table III). This demonstrates that the amino acid substitutions are specifically increasing replicative transposition. A double mutant containing both G194A and S256M mutations showed no additional increase in co-integrate formation above the value obtained with S256M (Table III), indicating that the mutations are probably acting in the same way.

The ability of V119A to rescue transposition of the mutant transposases prompted us to re-examine the IS903 R199 substitutions (nominally equivalent to P167 in IS10), which exhibited barely detectable levels of transposition and thus prevented detection of co-integrates (Table II). *In vitro* studies have shown that P167S is defective in hairpin formation, but nicking occurs efficiently at the 3' ends of the transposon (Kennedy *et al.*, 1998). Double mutants were therefore constructed with the V119A mutation in order to suppress the transposition defect of the R199 substitutions. Transposition frequency was rescued for two of the double mutants but not for the R199P transposase. More importantly, a small but reproducible increase (30- to 50-fold compared with wild-type transposase) in co-integrate formation was observed for the R199A and R199S double mutants (Table III).

In an effort to identify other transposase mutations that increased replicative transposition, we screened substitutions of the highly conserved YREK motif (Figure 3). In agreement with previous results (Haniford *et al.*, 1989; Bolland and Kleckner, 1996; Tavakoli *et al.*, 1997; Davies *et al.*, 2000), substitution of these residues had a very deleterious affect on transposition (Table III). Transposition of these mutant transposases could be rescued by introducing the hyperactive V119A mutation, and for two of these derivatives, Y252A and R255A, significant levels of co-integrates were detected (Table III). Several substitutions of K266 were examined (R, G and A), but none increased the level of co-integrate formation above background levels (Table III and data not shown).

Hypernicking and hairpin defect mutants of IS10 do not allow co-integrate formation

We also examined the effect of the IS10 mutations on IS10 transposition, as these had only been examined under *in vitro* conditions that would not have allowed detection

of replicative transposition intermediates (Bolland and Kleckner, 1995). Transposition of IS10 was monitored using derivatives that allowed direct selection for co-integrates. No co-integrates were detected with either the wild-type IS10 transposase (<1 in 10⁵ events) or any of the mutants tested (A162T, P167S and M289I; data not shown). These results are in contrast to those with the wild-type IS903 and mutants G194A, R199A and S256M, which generate co-integrates, and thus imply that there is a fundamental difference in the mechanism of transposition between the two elements (see Discussion).

Combining transposase mutants and dinucleotide pairs

We examined whether the protein and DNA mutants are additive in their effect on replicative transposition. Both G194A and S256M (combined with V119A) transposases were introduced into transposon vectors that were flanked by different nucleotides. In each case, the G194A and S256M transposases increased the co-integrate frequency (Table IV), but a flanking A did not significantly increase the co-integrate levels above those observed with other flanking nucleotides. This suggests that the effect of both the protein and DNA substitutions is at a similar step in the replicative transposition pathway.

A K266G substitution does not increase replicative transposition, but is sensitive to the flanking nucleotide

A substitution of glycine for lysine at transposase residue 266 exhibited an unusual sensitivity to the nature of the flanking DNA. Transposition mediated by K266G transposase was almost undetectable, but could be rescued significantly by introducing the V119A substitution (Table IV). However, the efficiency with which the V119A substitution rescued transposition mediated by transposase K266G was very dependent on the flanking nucleotide. Transposition was significantly higher when flanked by a 3' A or C compared with a G or T (520-fold comparing a flanking 3' A with a 3' T; Table IV). This is in contrast to the effect of the flanking nucleotide on transposition mediated by the wild-type, V119A and several other mutant transposases that we have screened. In these latter cases, a 5-fold reduction in transposition frequency is observed when the transposon is flanked by a 3' G, but for other flanking nucleotides there was little, or no variation in transposition.

Discussion

IS903 is unusual among bacterial transposable elements in that it has the ability to transpose by two pathways, which generate different products: simple insertions or co-integrates. Here, we have described a series of DNA and protein mutants that specifically increase the frequency of replicative transposition by IS903. These results can be rationalized in light of current models of cut-and-paste versus replicative transposition (reviewed in Craig, 1996; Turlan and Chandler, 2000), and also provide an explanation for the ability of IS903 to move by both pathways. The predominant product of IS903 transposition is a simple insertion. We propose that normally IS903 rapidly processes the flanking DNA to generate an excised

transposon resulting in a simple insertion. The mutants we have described would delay the processing of the 5'-flanking DNA at both ends of the transposon. This would effectively increase the half-life of the nicked product (Figure 2a), and thereby increase the chance of target capture to form a strand transfer intermediate (Figure 2e) that is then replicated to generate a co-integrate. The low level of co-integrates that can be detected with the wild-type transposon is most likely to be due to the effect of the 3'-flanking nucleotide. As observed in Table I, this can have a 500-fold effect on the frequency of co-integrate formation (from 0.01 to 5% depending on whether the transposon is flanked by a C or an A) even with a wild-type transposon. To our knowledge, this is the first time that substitutions in the flanking DNA have been documented to affect the outcome of transposition.

A considerable amount of biochemical evidence, from work with a number of transposons, has suggested that the distinguishing feature between simple insertion and replicative transposition is whether a double-strand break, or a nick, is made at the ends of the transposon before target capture (Craig, 1996; Turlan and Chandler, 2000). This was confirmed by a combination of *in vivo* and *in vitro* experiments with Tn7 (May and Craig, 1996). Tn7 is unusual in that it encodes both a transposase, TnsB, which cleaves and joins the 3' ends of the transposon to a target, and an endonuclease, TnsA, which specifically cleaves the 5' strand at each end of the transposon (Hickman *et al.*, 2000). Tn7 was converted from a cut-and-paste transposon to a replicative transposon by generating an endonuclease-defective TnsA protein. This allowed the normal 3' end cleavage and joining by TnsB, but prevented complete excision from the donor and therefore resulted in co-integrates.

We have described the first mutants of a transposase protein that directly enhance the ability to form co-integrates. The fact that these mutants also reduce overall transposition suggests that they are almost certainly modifying the transposase activity of the protein rather than identifying a second, endonuclease activity specific for cleavage of the 5'-flanking DNA (like Tn7). Furthermore, these residues are in close proximity to the catalytic residues (DDE motif) known to be essential for transposition (Figure 3) (Tavakoli *et al.*, 1997). Structural studies with the related IS50 transposase in complex with its transposon ends show that the equivalent residues in IS50 (R189, H194, Y319, R322 and W323) are all proximal to active site residues and the transposon ends (Davies *et al.*, 2000). Particularly relevant to this analysis, Y319 and R322 of the YREK motif are thought to play an important role in stabilizing the hairpin conformation of the DNA by making critical contacts to the non-transferred strand. Presumably the IS903 substitutions G194A, R199A, Y252A, R255A and S256M are inhibiting the efficiency of the 5'-strand cleavage step in transposition by perturbing the overall conformation or flexibility of the transposase active site. We note that G194, R199 and S256 are not conserved among the transposase family (Figure 3), consistent with the phenotype of these mutants being a result of structural perturbation of the active site, rather than a direct role in catalysis.

IS10 and IS50 transposition proceeds through a hairpin intermediate whose resolution results in complete excision

of the element from the flanking DNA (Figure 2b) (Kennedy *et al.*, 1998; Bhasin *et al.*, 1999). This allows a single active site of transposase to cleave two DNA strands of opposite polarity. In the case of IS903, no hairpins have been detected as yet. However, the effect of the substitutions provides genetic evidence suggesting that IS903 excises itself from flanking donor DNA via a hairpin intermediate. The P167S substitution of IS10 is defective in hairpin formation but generates a 3' nick with normal efficiency *in vitro*. Based on this phenotype, we predicted that substitution of the IS903 residue R199 might increase co-integrate formation in an *in vivo* assay. R199A and R199S transposases both generated co-integrates 30- to 50-fold greater than wild-type transposase (Table III). Transposition via a hairpin intermediate also provides a rational explanation for the effect of nucleotide changes at both the 3' flank (Table I) and the transposon termini (Tavakoli and Derbyshire, 1999). These DNA mutants would allow nicking to occur at the 3' end of the transposon but inhibit hairpin formation and thus favor co-integrate formation. Confirmation of these predictions based on our *in vivo* analysis will require development of an *in vitro* assay for IS903 transposition.

Our previous genetic and biochemical analyses indicated that the outer part of the inverted repeat represents a functionally distinct domain (Derbyshire *et al.*, 1987; Derbyshire and Grindley, 1992). As this domain contains the site of transposase cleavage, we predicted that it was likely to be in intimate contact with the active site of the protein, now elegantly confirmed for IS50 with the solution of the structure of the IS50 synaptic complex (Davies *et al.*, 2000). Thus it is perhaps not surprising that the dinucleotide at the transposon-flank junction can influence transposition (Table I). Analysis of our results suggests that this dinucleotide might influence transposition by affecting cleavage of both strands of DNA. Transposition frequency was consistently lower when the 3'-flanking nucleotide was a G (Table IV and data not shown), but this did not affect co-integrate formation. This is most consistent with inhibition of the first nicking step of transposition (Figure 2a). The sensitivity of transposition to the flanking nucleotide was enhanced dramatically by a K266G substitution (Table IV). Other substitutions of this residue (K266R or K266A) that still allowed transposition behaved similarly to wild-type transposase in that transposition was reduced 5- to 10-fold by a flanking 3' G (data not shown). This implies that the glycine residue in some way impairs the ability of the transposase to initiate transposition when flanked by a 3' G or T and suggests that K266 must be at the active site in close contact with DNA. Consistent with this, K266 is part of the highly conserved YREK motif found in the IS4 family of transposases (Rezsohazy *et al.*, 1993; Haren *et al.*, 1999), and in the IS50-IR complex the equivalent residue (K333) makes contacts with a phosphate on the transferred strand at the transposon terminus (Davies *et al.*, 2000). This residue is also conserved in the retroviral integrases (Kulkosky *et al.*, 1992), and DNA cross-linking studies have shown that in human immunodeficiency virus integrase this lysine (K159) is cross-linked to the terminal deoxyadenosine of retroviral DNA (Jenkins *et al.*, 1997).

We observed a dramatic increase in replicative transposition with a C-A dinucleotide at the transposon-flank

junction. This dinucleotide had no effect on overall transposition frequency, but increased co-integrate formation to 5%, suggesting an inhibition of second strand cleavage. Presumably with dinucleotides C–C and C–T at the transposon ends, cleavage of both strands is tightly coupled and efficient.

How is the dinucleotide affecting transposition? One possibility is that the efficiency of cleavage of the phosphodiester backbone (top or bottom strand) varies for each dinucleotide pair. Thus a C–A dinucleotide might be more resistant to cleavage and consequently prevent excision of the transposon from the flanking DNA. Alternatively, the flexibility of the dinucleotide pairs at the transposon–flank junction might be critical in adopting a structure that is more conducive to transposition. This latter possibility is particularly attractive in light of recent footprinting studies on an IS50 pre-cleavage synaptic complex in which significant enhancements were observed at the transposon termini, indicative of a strong local distortion (Bhasin *et al.*, 2000). Similar distortions were also detected at the Mu transposon ends in both the type 0 (Wang *et al.*, 1996) and type I synaptic complexes (Lavoie *et al.*, 1991).

Interestingly, we have detected a similar sensitivity to dinucleotide content on insertion of IS903 into a preferred target site (Hu *et al.*, 2001). In this case, we observed a significant effect on use of a defined target when altering the dinucleotide that is cleaved on transposon insertion. Thus, in two different assays at different stages of transposition (excision and integration), the chemistry of transposition is sensitive to the nature of the dinucleotide that is to be cleaved. This is entirely consistent with the detailed biochemical analyses that have demonstrated that transposition involves reiterated steps of hydrolysis and transesterification, and is thus likely to have similar substrate requirements for both excision and integration (Mizuuchi, 1997; Bhasin *et al.*, 1999; Kennedy *et al.*, 2000).

For Mu and Tn7, target capture occurs at an early stage in the transposition reaction before cleavage takes place (Bainton *et al.*, 1991; Naigamwalla and Chaconas, 1997). This ensures complete assembly of the transposition machinery before commitment to transposition. For IS10 and IS50, the element is completely excised from the donor DNA prior to target capture (Haniford and Kleckner, 1994; Sakai and Kleckner, 1997; Goryshin and Reznikoff, 1998), which precludes co-integrate formation (Figure 2c). It has been proposed that there might be a single binding site in the IS10 transposase for both target and flanking donor DNA and thus their binding is mutually exclusive (Sakai and Kleckner, 1997). If this were true, it would explain why we did not detect co-integrates with the hypernicking and hairpin-defective IS10 mutants. In contrast, the fact that derivatives of IS903 can generate co-integrates very efficiently suggests that target capture occurs early in the IS903 transposition reaction and is not excluded by flanking DNA.

IS903 has been shown preferentially to insert adjacent to G–C or C–G base pairs (Hu and Derbyshire, 1998; Hu *et al.*, 2001). Based on our results in Table I, this suggests that in the majority of cases, the next transposition event will generate a simple insertion and not a co-integrate. However, the ability to generate different transposition

end products can have important evolutionary consequences for both the element and its host. Probably the biggest disadvantage of replicative transposition is that intramolecular events can result in both inversions and, more seriously, deletions (Berg and Howe, 1989; Craig, 1996). In contrast, the ability to generate intermolecular co-integrates can dramatically enhance horizontal gene transfer, which can benefit the host population and facilitate dissemination of the transposon. In a donor strain that carries a conjugationally proficient plasmid, co-integrate formation would facilitate mobilization of non-mobile plasmids and, perhaps more importantly, enhance Hfr formation and thus chromosomal transfer into diverse hosts. Co-integrate formation in a recipient of DNA is also likely to be important for acquisition of new phenotypes. Replicative transposition from a newly introduced plasmid or phage into the recipient chromosome would result in permanent acquisition of all the genetic information from the plasmid or phage via replicon fusion. This is particularly important if transfer or transduction occurs into hosts that cannot support replication of that incoming DNA.

Materials and methods

General procedures

Restriction enzymes and DNA-modifying enzymes were purchased from New England Biolabs and Boehringer Mannheim and were used according to the supplier's instructions. Oligonucleotides were synthesized, and DNA sequencing was performed in the Molecular Genetics Core at the Wadsworth Center. Site-directed mutagenesis was carried out as described by Kunkel *et al.* (1991). All mutants generated were sequenced before subcloning to avoid secondary mutations. The substrate for site-directed mutagenesis was pJD2, a pUC118 derivative containing the transposase gene cloned into the polylinker (Tavakoli *et al.*, 1997).

Mating-out assay

The transposon vector was pKD498, which encodes kanamycin resistance (Figure 1; Tavakoli and Derbyshire, 1999). The transposon carried by pKD498 encodes both chloramphenicol and ampicillin resistance, and carries the pBR322 origin of replication. Mutant transposase genes were cloned on *SalI*–*Bam*HI fragments from pJD2 derivatives to generate pKD498-isotype plasmids. pKD498 is flanked at both ends of the transposon by the nucleotides CTG. pKD498-isotype plasmids containing different nucleotides flanking the IR were created by PCR and cloning. Co-integrates can result from simple insertion of a transposon carried by a plasmid dimer or multimer, if transposition is mediated by ends from different elements (Craig, 1996). To ensure that plasmids were monomeric, they were propagated in a *recA*[–] strain and were screened by gel electrophoresis prior to assaying transposition. Mating-out assays were carried out using RR1023 (*recA56*, *pOX38*) as donor and NG135 (*F*[–], *recA56*, *strA*) as recipient (Derbyshire *et al.*, 1987). The frequencies presented are the average of at least six experiments. Selection for chloramphenicol and streptomycin resistance allowed determination of total transposition events into *pOX38*, whereas selection with kanamycin and streptomycin allowed scoring of replicative events (co-integrates, Figure 1b). In addition, several representative co-integrate molecules were sequenced to show that the characteristic 9 bp target duplication had occurred at transposon–target junctions.

Acknowledgements

We thank Wen-Yuan Hu and Salvador Echeverria for making the initial observation that flanking nucleotides influenced replicative transposition. We thank Drs Marlene Belfort, Joan Curcio, Victoria Derbyshire and members of the Derbyshire laboratory for their insightful comments on the manuscript. We gratefully acknowledge David Haniford for supplying the IS10 mutants, and the Wadsworth Center Molecular Genetics Core facility for DNA sequencing and oligonucleotide synthesis. This work

was supported by the National Institutes of Health, grant GM50699 to K.M.D.

References

- Bainton,R., Gamas,P. and Craig,N.L. (1991) Tn7 transposition *in vitro* proceeds through an excised transposon intermediate generated by staggered breaks in DNA. *Cell*, **65**, 805–816.
- Berg,D.E. and Howe,M.M. (1989) *Mobile DNA*. American Society for Microbiology, Washington, DC.
- Bhasin,A., Goryshin,I.Y. and Reznikoff,W.S. (1999) Hairpin formation in Tn5 transposition. *J. Biol. Chem.*, **274**, 37021–37029.
- Bhasin,A., Goryshin,I.Y., Steiniger-White,M., York,D. and Reznikoff,W.S. (2000) Characterization of a Tn5 pre-cleavage synaptic complex. *J. Mol. Biol.*, **302**, 49–63.
- Bolland,S. and Kleckner,N. (1995) The two single-strand cleavages at each end of Tn10 occur in a specific order during transposition. *Proc. Natl Acad. Sci. USA*, **92**, 7814–7818.
- Bolland,S. and Kleckner,N. (1996) The three chemical steps of Tn10/IS10 transposition involve repeated utilization of a single active site. *Cell*, **84**, 223–233.
- Craig,N.L. (1996) Transposition. In Neidhardt,F.C. (ed.), *Escherichia coli and Salmonella Cellular and Molecular Biology*. Vol. 2. American Society for Microbiology, Washington, DC, pp. 2339–2362.
- Craigie,R. and Mizuuchi,K. (1985) Mechanism of transposition of bacteriophage Mu: structure of a transposition intermediate. *Cell*, **41**, 867–876.
- Davies,D.R., Goryshin,I.Y., Reznikoff,W.S. and Rayment,I. (2000) Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science*, **289**, 77–85.
- Derbyshire,K.M. and Grindley,N.D. (1992) Binding of the IS903 transposase to its inverted repeat *in vitro*. *EMBO J.*, **11**, 3449–3455.
- Derbyshire,K.M., Hwang,L. and Grindley,N.D. (1987) Genetic analysis of the interaction of the insertion sequence IS903 transposase with its terminal inverted repeats. *Proc. Natl Acad. Sci. USA*, **84**, 8049–8053.
- Goryshin,I.Y. and Reznikoff,W.S. (1998) Tn5 *in vitro* transposition. *J. Biol. Chem.*, **273**, 7367–7374.
- Grindley,N.D.F. and Joyce,C.M. (1981) Genetic and DNA sequence analysis of the kanamycin resistance transposon Tn903. *Cold Spring Harb. Symp. Quant. Biol.*, **45**, 125–133.
- Haniford,D. and Kleckner,N. (1994) Tn10 transposition *in vivo*: temporal separation of cleavages at the two transposon ends and roles of terminal basepairs subsequent to interaction of ends. *EMBO J.*, **13**, 3401–3411.
- Haniford,D.B., Chelouche,A.R. and Kleckner,N. (1989) A specific class of IS10 transposase mutants are blocked for target site interactions and promote formation of an excised transposon fragment. *Cell*, **59**, 385–394.
- Haren,L., Ton-Hoang,B. and Chandler,M. (1999) Integrating DNA: transposases and retroviral integrases. *Annu. Rev. Microbiol.*, **53**, 245–281.
- Hickman,A.B., Li,Y., Mathew,S.V., May,E.W., Craig,N.L. and Dyda,F. (2000) Unexpected structural diversity in DNA recombination: the restriction endonuclease connection. *Mol. Cell*, **5**, 1025–1034.
- Hu,W.Y. and Derbyshire,K.M. (1998) Target choice and orientation preference of the insertion sequence IS903. *J. Bacteriol.*, **180**, 3039–3048.
- Hu,W.Y., Thompson,W., Lawrence,C.E. and Derbyshire,K.M. (2001) Anatomy of a preferred target site for the insertion sequence IS903. *J. Mol. Biol.*, **306**, 403–416.
- Jenkins,T.M., Esposito,D., Engelman,A. and Craigie,R. (1997) Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking. *EMBO J.*, **16**, 6849–6859.
- Kennedy,A.K. and Haniford,D.B. (1996) Isolation and characterization of IS10 transposase separation of function mutants: identification of amino acid residues in transposase that are important for active site function and the stability of transposition intermediates. *J. Mol. Biol.*, **256**, 533–547.
- Kennedy,A.K., Guhathakurta,A., Kleckner,N. and Haniford,D.B. (1998) Tn10 transposition via a DNA hairpin intermediate. *Cell*, **95**, 125–134.
- Kennedy,A.K., Haniford,D.B. and Mizuuchi,K. (2000) Single active site catalysis of the successive phosphoryl transfer steps by DNA transposases: insights from phosphorothioate stereoselectivity. *Cell*, **101**, 295–305.
- Kulkosky,J., Jones,K.S., Katz,R.A., Mack,J.P.G. and Skalka,A.M. (1992) Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell. Biol.*, **12**, 2331–2338.
- Kunkel,T.A., Bebenek,K. and McClary,J. (1991) Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol.*, **204**, 125–139.
- Lavoie,B.D., Chan,B.S., Allison,R.G. and Chaconas,G. (1991) Structural aspects of a higher order nucleoprotein complex: induction of an altered DNA structure at the Mu–host junction of the Mu type 1 transpososome. *EMBO J.*, **10**, 3051–3059.
- May,E.W. and Craig,N.L. (1996) Switching from cut-and-paste to replicative Tn7 transposition. *Science*, **272**, 401–404.
- Mizuuchi,K. (1992) Transpositional recombination: mechanistic insights from studies of Mu and other elements. *Annu. Rev. Biochem.*, **61**, 1011–1051.
- Mizuuchi,K. (1997) Polynucleotidyl transfer reactions in site-specific DNA recombination. *Genes Cells*, **2**, 1–12.
- Naigamwalla,D.Z. and Chaconas,G.C. (1997) A new set of Mu DNA transposition intermediates: alternate pathways of target capture preceding strand transfer. *EMBO J.*, **16**, 5227–5234.
- Ohtsubo,E., Zenilman,M., Ohtsubo,H., McCormick,M., Machida,C. and Machida,Y. (1981) Mechanism of insertion and cointegration mediated by IS1 and Tn3. *Cold Spring Harb. Symp. Quant. Biol.*, **45**, 283–295.
- Rezsohazy,R., Hallet,B., Delcour,J. and Mahillon,J. (1993) The IS4 family of insertion sequences: evidence for a conserved transposase motif. *Mol. Microbiol.*, **9**, 1283–1295.
- Sakai,J. and Kleckner,N. (1997) The Tn10 synaptic complex can capture a target DNA only after transposon excision. *Cell*, **89**, 205–214.
- Sarnovsky,R.J., May,E.W. and Craig,N.L. (1996) The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *EMBO J.*, **15**, 6348–6361.
- Tavakoli,N.P. and Derbyshire,K.M. (1999) IS903 transposase mutants that suppress defective inverted repeats. *Mol. Microbiol.*, **31**, 1183–1195.
- Tavakoli,N.P., DeVost,J. and Derbyshire,K.M. (1997) Defining functional regions of the IS903 transposase. *J. Mol. Biol.*, **274**, 491–504.
- Turlan,C. and Chandler,M. (1995) IS1-mediated intramolecular rearrangements: formation of excised transposon circles and replicative deletions. *EMBO J.*, **14**, 5410–5421.
- Turlan,C. and Chandler,M. (2000) Playing second fiddle: second-strand processing and liberation of transposable elements from donor DNA. *Trends Microbiol.*, **8**, 268–274.
- Wang,Z., Namgoong,S.Y., Zhang,X. and Harshey,R.M. (1996) Kinetic and structural probing of the precleavage synaptic complex (type 0) formed during phage Mu transposition. Action of metal ions and reagents specific to single-stranded DNA. *J. Biol. Chem.*, **271**, 9619–9626.
- Weinert,T.A., Derbyshire,K.M., Hughson,F.M. and Grindley,N.D.F. (1984) Replicative and conservative transpositional recombination of insertion sequences. *Cold Spring Harb. Symp. Quant. Biol.*, **49**, 251–260.

Received February 7, 2001; accepted April 3, 2001