

Hairpin Formation in Tn5 Transposition*

(Received for publication, August 19, 1999)

Archana Bhasin, Igor Y. Goryshin, and William S. Reznikoff‡

From the Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706

The initial chemical steps in Tn5 transposition result in blunt end cleavage of the transposon from the donor DNA. We demonstrate that this cleavage occurs via a hairpin intermediate. The first step is a 3' hydrolytic nick by transposase. The free 3'OH then attacks the phosphodiester bond on the opposite strand, forming a hairpin at the transposon end. In addition to forming precise hairpins, Tn5 transposase can form imprecise hairpins. This is the first example of imprecise hairpin formation on transposon end DNA. To undergo strand transfer, the hairpin must be resolved by a transposase-catalyzed hydrolytic cleavage. We show that both precise and imprecise hairpins are opened by transposase. A transposition mechanism utilizing a hairpin intermediate allows a single transposase active site to cleave both 3' and 5' strands without massive protein/DNA rearrangements.

Genomic DNA is known to undergo a variety of rearrangements, including inversions, deletions, duplications, and translocations. Transposition is one of the ways by which these rearrangements occur. The transposon Tn5 is a 5.8-kilobase pair prokaryotic, composite transposon consisting of two inverted repeats, IS50L and IS50R, that flank genes encoding antibiotic resistance. IS50R codes for the Tn5 transposase (Tnp), the protein catalyzing all the steps in transposition. Each IS50 repeat is bracketed by two different 19-bp¹ end sequences, termed the outside end (OE) and the inside end (IE), that are specifically recognized by Tn5 transposase (Fig. 1) (reviewed in Refs. 1 and 2). Transposition of the full Tn5 element requires two OEs, whereas transposition of an IS50 element requires one OE and one IE. *In vitro*, Tn5 transposition requires only a hyperactive mutant transposase, such as EK54/LP372 Tnp, Mg²⁺, transposon DNA defined by two inverted 19-bp end sequences, and target DNA (3). Transposition frequency can be increased by using the mosaic end sequence (ME), a hyperactive, synthetic end sequence that is a hybrid of the OE and IE (Fig. 1) (4).

Tn5 is a member of the "cut and paste" family of transposons that includes Tn10 and Tn7. Although the transposition reactions vary in details, they follow the same basic mechanism. Transposase binds to the transposon DNA at the end recognition sequences. Then, the end sequences are brought together via transposase oligomerization to form a complex nucleopro-

tein structure,² termed a synaptic complex (5, 6). Once a stable synaptic complex has been formed, the transposon end sequence-donor DNA boundary can be cleaved to release a pair of 3'OHs at the ends, which are then used in the next chemical step, strand transfer (Fig. 2A). Strand transfer occurs via a one step transesterification (7, 8) in which the 3'OHs attack phosphodiester bonds in the target DNA in a staggered fashion. For Tn5, the staggered attack leads to 9-bp gaps flanking the integrated transposon. The gaps are presumed to be filled in and ligated in the cell, and this leads to 9-bp duplications of the target sequence flanking the transposon (9, 10). In replicative transposition, the strand transfer product is a cointegrate, which is processed by DNA replication (11).

The cleavage step differentiates cut and paste transposition from replicative transposition. In replicative transposons, such as phage Mu, only the 3' end of the transposon is nicked to release a 3'OH, whereas the 5' end remains covalently joined to donor DNA (7). On the other hand, both the 3' and 5' ends of cut and paste transposons are cleaved from the flanking donor DNA. The Tn5 transposon is excised from the donor DNA by flush double strand breaks precisely at the end sequences (3), as is Tn10 (12). Tn7 is excised with three-nucleotide 5' overhangs (13). In this study, we examine how these double strand breaks arise. Previous work on Tn10 showed that the 3' and 5' nicks did not occur simultaneously, but rather were sequential with the 3' nick preceding the 5' nick (14). In addition, in Mu, a single transposase active site carries out all the chemical steps (cleavage and strand transfer) at one end sequence (15). Similarly, in Tn10, evidence suggests that a single transposase active site can also carry out all chemical steps at one end (16). Like Tn10 transposase, Tn5 transposase is a single polypeptide chain. It is difficult to imagine how a single active site can cleave both the 3' and 5' ends of a transposon without major conformational changes in the transposase and/or DNA. This is commonly known as the polarity problem. Tn7 has bypassed this problem by having its transposase made up of two polypeptides, TnsA and TnsB. TnsB contains the active site responsible for 3' nicking, whereas TnsA contains the active site for 5' nicking (17). V(D)J recombination, an antigen receptor recombination system mechanistically similar to transposition, bypasses this polarity problem by going through a hairpin intermediate (18).

It has recently been found that the Tn10 reaction also proceeds via a hairpin intermediate (19). The model for cleavage via a hairpin is as follows. The first cleavage step is a 3' hydrolytic nick by transposase at the end sequence. The free 3'OH attacks the 5' end, forming a hairpin at the transposon end and releasing the flanking donor DNA. In order to undergo strand transfer, the hairpin needs to be resolved. The hairpin is hydrolytically cleaved by transposase in a step chemically identical to the first cleavage step, 3' nicking. Cleavage of the hairpin frees a 3'OH. This 3'OH joins to target DNA in a strand

* This research was supported by National Institutes of Health Grant GM50692. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Dept. of Biochemistry, University of Wisconsin-Madison, 433 Babcock Dr., Madison, WI 53706. Tel.: 608-262-3608; Fax: 608-262-3453; E-mail: reznikoff@biochem.wisc.edu.

¹ The abbreviations used are: bp, base pair; OE, outside end; IE, inside end; PAGE, polyacrylamide gel electrophoresis.

² A. Bhasin, I. Y. Goryshin, M. Steiniger-White, D. York, and W. S. Reznikoff, manuscript in preparation.

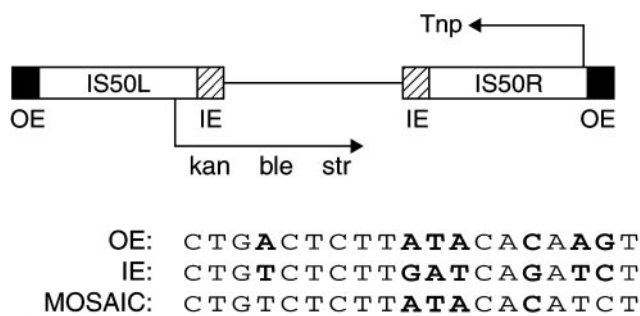


FIG. 1. The structure of the *Tn5* transposon. *Tn5* is composed of two insertion sequences, IS50L and IS50R. IS50R codes for *Tn5* transposase. Each insertion sequence is flanked by 19-bp end sequences termed outside end and inside end. The *black box* represents the OE and the *striped box* represents the IE. The OE and IE differ by 7 bases (shown in *boldface*). A hyperactive, hybrid end sequence, the mosaic end (*MOSAIC*) was used in this study. The *boldface letters* in the mosaic end sequence are outside end sequences, whereas the *lightface letters* are inside end sequences.

transfer reaction chemically similar to the hairpin formation step (Fig. 2B). Note that this mechanism utilizes only two repetitive chemistries, nicking and strand transfer. Furthermore, this model allows for a single transposase active site to perform all the necessary chemical steps without having to undergo a large conformational change, thus relieving the polarity problem.

Using short, radiolabeled transposon end-containing DNAs in the *in vitro* *Tn5* cleavage assay, we studied the cleavage step in *Tn5* transposition. Our results demonstrate that *Tn5* transposases via a hairpin intermediate. Hairpin formation is essential for blunt end cleavage of the transposon. Preformed hairpins can be cleaved to form a blunt-ended transposon. Imprecise hairpins can also be formed and cleaved, although less efficiently. Cleavage of the imprecise hairpin leads to precise 3' end exposure. This finding of imprecise hairpin formation and resolution on the transposon end is novel. Lastly, position 1 of the end sequence is required for efficient hairpin formation.

EXPERIMENTAL PROCEDURES

Preparation of DNA Substrates—DNA substrates were 5' end-labeled using T4 Polynucleotide kinase (Promega) and [γ - 32 P]ATP (DuPont, Redivue). Oligonucleotides were annealed in 20 mM Tris-HCl, pH 7.5, 50 mM NaCl by heating at 65 °C for 10 min and then cooling the reaction at room temperature.

Protein Purification—EK54/LP372 transposase was purified using the IMPACT T7 one-step protein purification system (New England Biolabs). The expression vector pGRTYB35 contains the transposase gene linked to the chitin binding domain through an intein. It was created by digesting pGRTEMP2 (3) with *Bgl*II, filling in with Sequenase, and digesting with *Bsp*HI. The resulting 1.5-kilobase pair transposase gene fragment was agarose gel-purified and ligated into the large *Nco*I-*Sma*I fragment of pTYB4 (NEB). pGRTYB35 was transformed into the strain ER2566 (NEB). These cells were grown to an A_{600} of 0.5, and were then induced with 100 μ M isopropyl-1-thio- β -D-galactopyranoside (final concentration) for 5 h at room temperature (23 °C). Cells were harvested and resuspended in 20 ml (for 1 liter cell culture) of TEGX (20 mM Tris-HCl, pH 7.5, 0.7 M NaCl, 1 mM EDTA, 10% glycerol, 0.1% Triton X-100) containing a protease inhibitor mixture (Roche Molecular Biochemicals). Cells were sonicated, and the lysate was cleared by centrifugation. Cleared lysate was loaded onto a chitin (NEB) column (10 ml of column resin/1 liter of cell culture). The column was washed with 20 column volumes of TEGX, flushed with 3 column volumes of TEGX containing 50 mM dithiothreitol, and then left overnight at 4 °C. Protein was eluted from the column with TEGX, and transposase was collected in the first 10 ml of eluate. The EK54/LP372 transposase was determined to be at least 95% pure by SDS-PAGE analysis and Coomassie staining. Aliquots of 0.4 μ g/ μ l, determined by Bradford assay, were stored at -70 °C.

A 5' end-labeled single-stranded oligonucleotide was used to assay for 3'-5' exonuclease activity in the EK54/LP372 transposase prepara-

tion and in a catalytically inactive mutant transposase preparation.³ A weak single base ladder (up to 9 bases) of DNA degradation products was detected for both protein preparations (data not shown).

In Vitro Cleavage Reactions—The substrate used in many of the *in vitro* cleavage reactions (Figs. 3A and 4) was made by annealing two complementary 53-base mosaic end (the mosaic end is shown in boldface letters) containing oligonucleotides (40-base transposon and 13-base donor DNA) 5'-ACATGCATGCTCACTCACTCAAGATGTGTAT-AAGAGACAGTCGACCTGCAGGG-3' and 5'-CCCTGCAGGTCGACTGTCTCTTATACACATCTTGAGTGAGTGAGCATGCATGT-3' (Research Genetics, PAGE-purified).

In vitro cleavage reactions consisted of radiolabeled mosaic end DNA (see legends to Figs. 3 and 6), 377 nM EK54/LP372 transposase, 0.1 M potassium glutamate, 25 mM Tris acetate, pH 7.5, 10 mM Mg²⁺ acetate, 50 μ g/ml bovine serum albumin, 0.5 mM β -mercaptoethanol, 100 μ g/ml tRNA in a total of 20 μ l. Reactions were incubated at 37 °C for various times (see figure legends). Reactions were stopped by phenol/chloroform extraction and ethanol precipitation in the presence of 20 μ g of glycogen. Reactions were resuspended in water, and an equivalent volume of formamide loading buffer (20) was added. Reactions were then boiled for 5 min and electrophoresed through 15% denaturing polyacrylamide gels. Gels were visualized and quantitated with a phosphorimager (Molecular Dynamics).

Elution of DNA from Polyacrylamide Gels—Putative hairpin DNA generated from *in vitro* cleavage reactions was excised from 15% denaturing polyacrylamide gels. The excised DNA was eluted in TE buffer, (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, pH 7.5) (Fig. 4), or in 20 mM Tris-HCl, pH 7.5, 50 mM NaCl (Fig. 3B) overnight at 4 °C. The eluate was ethanol precipitated in the presence of glycogen and resuspended in water. This DNA was then either electrophoresed through a 15% native gel (Fig. 3B) or subjected to Maxam-Gilbert A+G chemical sequencing (Fig. 4) (21).

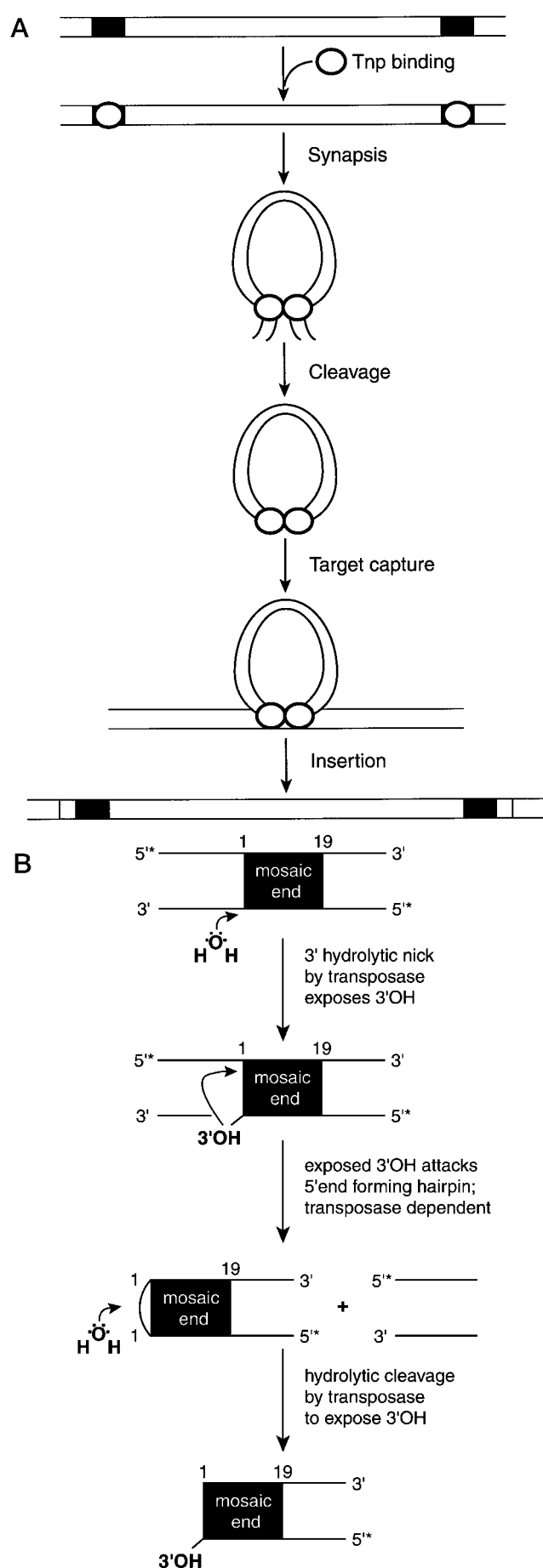
Cloning of the Hairpin—To clone the hairpin, a 53-base oligonucleotide containing the mosaic end (in boldface), the -35 region of the tetracycline promoter (in italics), and a 4-bp *Eco*RI overhang at the 5' end: 5'-AATTTTATCTGTCAAGATGTGTATAAGAGACAGTGTGGGAGCTCGAAGTGGAG-3' was annealed to a 48-base oligonucleotide containing the mosaic end, the -35 region of the tetracycline promoter, and a 4-bp *Aat*II overhang at the 3' end: 5'-CTCAGTTCGAGCTCCCAACACTGTCCTTATACACATCTTGACAACGT-3' (Integrated DNA Technologies, PAGE-purified) to form the substrate. Some of this substrate was 5' end-labeled, whereas some was phosphorylated with T4 polynucleotide kinase and ATP. An *in vitro* cleavage reaction was done with 1.5 nM radiolabeled substrate and 73.5 nM phosphorylated substrate (reaction conditions described above). Aliquots were taken at 10, 20, and 30 min. Reactions were stopped and loaded on a 15% denaturing polyacrylamide gel. The hairpins were isolated from the gel and eluted in TE. Eluates were pooled, ethanol-precipitated in the presence of glycogen, and resuspended in water.

The cloning vector used was pGREC1. pGREC1 was created by digesting pBR322 with *Eco*RI and *Cl*aI, filling in with Sequenase, and ligation with T4 DNA ligase (Promega). pGREC1 is missing the -35 region of the tetracycline promoter, but contains an *Eco*RI site followed by the -10 region of the tetracycline promoter. The plasmid was purified using the QIAfilter Maxi kit. pGREC1 was digested with *Aat*II and *Eco*RI (Promega), and the large fragment (4261-bp) was agarose gel-purified using a QIAquick gel extraction kit. The isolated hairpin (with *Aat*II and *Eco*RI sites) was ligated as a single strand into the pGREC1 *Aat*II/*Eco*RI large fragment using T4 DNA ligase. The ligation was dialyzed and electroporated into DH5 α cells. The transformation was plated onto agar containing tetracycline (15 μ g/ml). Plasmid DNA from colonies was isolated by an alkaline lysis miniprep protocol and dideoxy sequenced according to the protocol of Goryshin and Reznikoff (3).

Time Course—A 40- μ l *in vitro* cleavage reaction consisting of 71.5 nM radiolabeled 53-base mosaic end and 377 nM EK54/LP372 transposase (reaction conditions described above) was incubated at 37 °C. 2.5- μ l aliquots were taken at 0, 5, 10, 15, 25, 35, 45, 60, 75, 90, 120, and 150 min. The reactions were stopped by adding EDTA to a final concentration of 12.2 mM and keeping the aliquots on ice. After all of the time points were taken, all aliquots were subjected to phenol/chloroform extraction and ethanol precipitation in the presence of glycogen. Reactions were resuspended in 50% formamide loading dye, boiled for 5 min, and electrophoresed through 15% denaturing polyacrylamide gels. Gels were visualized with a phosphorimager.

Preformed Hairpin Cleavage and Recovery of Transposition Prod-

³ T. A. Naumann and W. S. Reznikoff, manuscript in preparation.



ucts—Imprecise and precise hairpins were made by annealing self-complementary oligonucleotides containing head-to-head mosaic ends (in boldface), the -35 region of the tetracycline promoter (in italics, overlapping the mosaic end) and a *Pml*I restriction site as a 5' overhang. The hairpin junctions are underlined. The 50-base precise hairpin is 5'-CACGTGTTGACATGTGTATAAGAGACAGCTGTCTC-TTATACACATGTCAA-3', and the 51-base imprecise hairpin is 5'-CACGTGTTGACATGTGTATAAGAGACAGACTGTCTCTTATACACATGTCAA-3' (Integrated DNA Technologies, PAGE-purified). These annealed oligonucleotides were used as substrates in an *in vitro* cleavage reaction.

These oligonucleotides were also used to recover strand transfer products. The precise or imprecise hairpin oligonucleotide was phosphorylated with T4 Polynucleotide kinase (Promega) and ATP. The phosphorylated hairpins were then reacted *in vitro* with EK54/LP372 transposase and the plasmid pGREC1 as target DNA. The reaction conditions were as described above except with 750 nM EK54/LP372 Tnp, 4.8 pmol of hairpin DNA, and 0.04 pmol of pGREC1. The reactions were incubated at 37 °C for 2 h. Insertion by transposition of two cleaved hairpins into pGREC1 leads to linearization of the plasmid. The linearized plasmid was recircularized by ligation of the two *Pml*I restriction sites with T4 DNA ligase (Promega). Upon ligation, the tetracycline resistance gene of pGREC1 was activated because the inserted transposon provided the -35 region of the promoter. Note that in these substrates, position 18 of the mosaic end was changed from a C to a G to accommodate a properly spaced -35 region. This base change had no effect on substrate activity (data not shown).

The ligation mixture was dialyzed, transformed by electroporation into DH5 α cells, and plated onto agar containing tetracycline (15 μ g/ml). Plasmid DNA from colonies was isolated and dideoxy-sequenced.

Dideoxy Experiment—For the dideoxy experiment, we used four oligonucleotides: 20-mer, 5'-TGTTGGGAGCTCGAACTGAG-3'; 39-mer, 5'-ACATGCATGCTCACTCACTCAAGATGTGTATAAGAGACA-3'; 40-mer, 5'-ACATGCATGCTCACTCACTCAAGATGTGTATAAGAGACAG-3'; and 60-mer, 5'-CTCAGTTCGAGCTCCCAACTGTCTCTT-ATACACATCTTGAGTGAGTGAGCATGCATGT-3' (all Integrated DNA Technologies, PAGE-purified). The 39-mer, 40-mer and 60-mer contained the mosaic end (shown in boldface) and were all 5' end-labeled. A ddGTP was added to the 3' end of some of the labeled 39-mer by terminal transferase. The reaction consisted of 0.5 μ M radiolabeled 39-mer, 5 μ M ddGTP (Roche Molecular Biochemicals) and 20 units of terminal transferase (Promega) and was incubated at 37 °C for 45 min. The reaction was heat-inactivated at 65 °C, ethanol-precipitated in the presence of glycogen, and resuspended in water. The 39ddGTP was then tailed with dNTPs and terminal transferase. This reaction consisted of 0.58 μ M 39ddGTP, 15 μ M each dNTP and 20 units of terminal transferase. This reaction product plus the other labeled oligonucleotides were ethanol precipitated, electrophoresed on a 15% denaturing polyacrylamide gel for purification, and visualized by autoradiography. The oligonucleotides were isolated from the gel and eluted in TE. The eluates were ethanol precipitated in the presence of glycogen and DNA concentrations were determined by scintillation counting. The DNA substrates were assembled by annealing oligonucleotides. Annealing of the 60-mer, 20-mer, and 39-mer formed the substrate 39dA. Annealing of the 60-mer, 20-mer, and 40-mer formed the substrate 40dG, whereas annealing of the 60-mer, 20-mer, and 39ddGTP formed the 40ddG substrate. The final concentrations of the annealed substrates were 0.027 pmol/ μ l. These substrates were then used in *in vitro* cleavage reactions consisting of 2.65 nM substrate DNA, 377 nM EK54/LP372 transposase, 5 mg/ml tRNA, 0.1 M potassium glutamate, 25 mM Tris acetate, pH 7.5, 10 mM Mg²⁺ acetate, 50 μ g/ml bovine serum albumin, 0.5 mM β -mercaptoethanol.

Fig. 2. A, a model of *Tn5* transposition. First, transposase (*circle*) binds monomerically and bends the transposon end sequences (*black rectangles*). Then, the end sequences are brought together via transposase oligomerization to form a synaptic complex. Once the synaptic complex is formed, the transposon can be cleaved. The cleaved transposon captures target DNA and inserts the transposon via a one-step transesterification reaction. The 9-bp duplication flanking the insert is shown as a *gray box*. **B**, a model depicting cleavage of the transposon ends via a hairpin intermediate. The first cleavage step is a 3' hydrolytic nick by transposase at the end sequence. The free 3'OH attacks the 5' end, forming a hairpin at the transposon end and releasing the flanking donor DNA. In order to undergo strand transfer, the hairpin needs to be resolved. Hydrolytic cleavage of the hairpin frees a 3'OH for strand transfer.

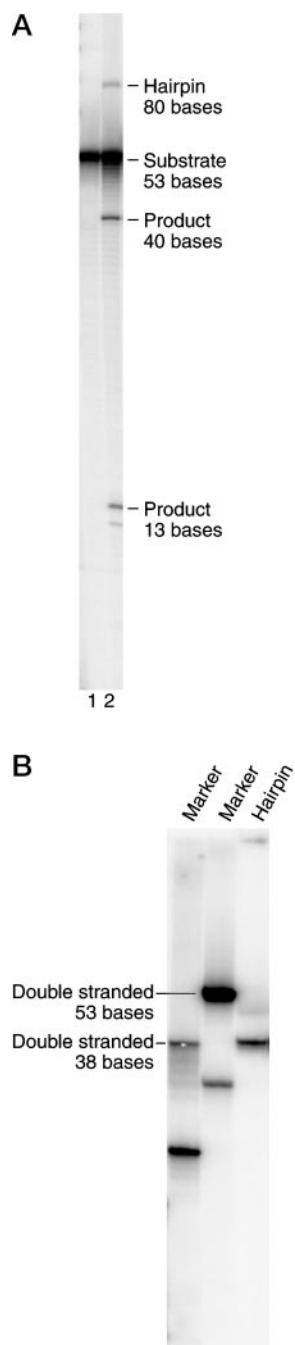


FIG. 3. *A*, *in vitro* cleavage reaction products electrophoresed on a 15% denaturing gel. The reaction contained 7 nM 5' end-labeled 53-bp mosaic end containing DNA and 377 nM EK54/LP372 transposase (see under "Experimental Procedures" for reaction conditions) and was incubated at 37 °C for 1 h. Lane 1 is DNA only, and lane 2 is DNA plus transposase. In the presence of EK54/LP372 transposase, the DNA is cleaved into a 40-base transposon and a 13-base donor DNA. Also, note the low mobility putative hairpin migrating at 80 bases. *B*, the putative hairpin DNA was isolated from a denaturing polyacrylamide gel; eluted in 20 mM Tris-HCl, pH 7.5, 50 mM NaCl; ethanol-precipitated; resuspended in nondenaturing loading dye; and electrophoresed on a 15% native gel. Under native conditions, the low mobility (80 base) putative hairpin DNA migrates at a mobility expected for a 40-bp fragment. This is indicative of a hairpin.

Quantitation—All quantitation was done using Molecular Dynamics ImageQuant.

RESULTS

Hairpin Species Exist in *Tn5* Transposition—A key event in *Tn5* transposition is the blunt end cleavage of the transposon free from the donor backbone DNA. We have developed an *in*

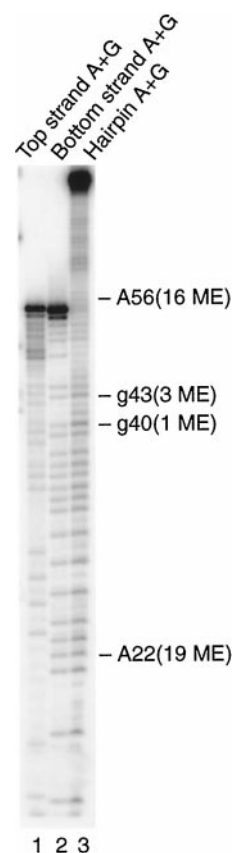


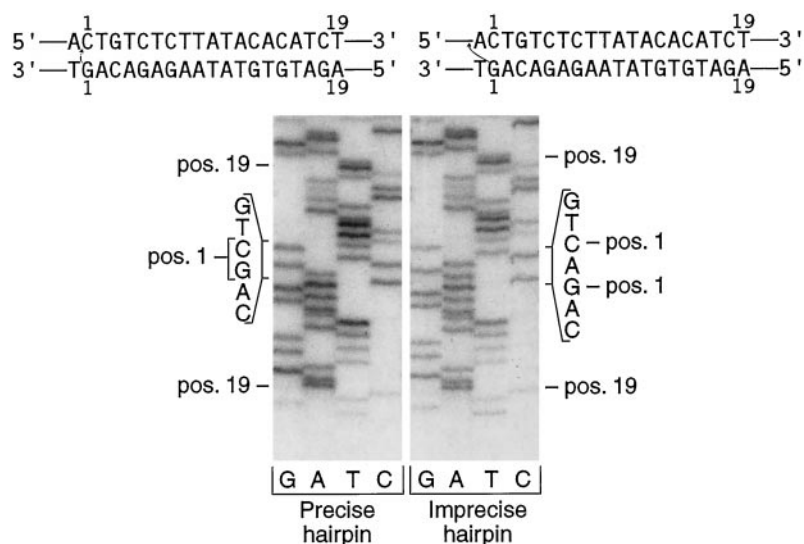
FIG. 4. **The 80-base species is a hairpin with two mosaic ends head to head.** The 80-base hairpin DNA was isolated from a denaturing polyacrylamide gel and subjected to Maxam-Gilbert A+G sequencing (lane 3). The individual strands of the substrate DNA were also sequenced (lanes 1 and 2). The sequencing reactions were electrophoresed on a 10% denaturing polyacrylamide gel. A few bases of the mosaic end DNA (ME) are marked.

vitro assay to study the cleavage step in detail. In this assay, we incubate short, radiolabeled DNAs containing the 19-bp mosaic end sequence with purified EK54/LP372 transposase in a potassium glutamate buffer in the presence of 10 mM Mg^{2+} acetate. The reactions are incubated at 37 °C to allow cleavage to occur, and the cleavage products are analyzed on denaturing polyacrylamide gels. Using a 53-bp DNA, we expected to see two cleavage products, a 40-base transposon and a 13-base donor DNA. In addition to the expected products, we detected a low mobility band migrating at 80 bases in a denaturing polyacrylamide gel (Fig. 3A). We hypothesized that this band was a hairpin intermediate.

To test whether the low mobility band was a hairpin, the DNA in the band was isolated from a denaturing gel and was electrophoresed on a native polyacrylamide gel. Under native conditions, the isolated DNA migrated faster, at 40 base pairs (Fig. 3B). This differential mobility under denaturing and non-denaturing conditions is characteristic of a hairpin.

Maxam-Gilbert sequencing was used to verify that this species is a hairpin. The ^{32}P end-labeled putative hairpin DNA was isolated from a denaturing polyacrylamide gel and was subjected to Maxam-Gilbert A+G chemical sequencing (Fig. 4) (21). The sequence matched bottom strand substrate DNA until hitting position 1 of the 3' end of the mosaic end, and then the sequence began to match top strand substrate DNA beginning at position 1 of the 5' end of the mosaic end. Notice that the three sets of triplets beyond the A-56 position on the hairpin A+G sequence match the three sets of triplets found in the top strand A+G sequence (Fig. 4, lanes 1 and 3). This sequence of

FIG. 5. Sanger dideoxy sequences of cloned hairpins. The first sequence is the precise hairpin. The precise hairpin junction is 5'-CAGCTG-3'. The second sequence is the imprecise hairpin. The imprecise hairpin junction is 5'-CAGACTG-3'. Positions 1 and 19 are marked. The additional A in the imprecise hairpin is located between both position 1s.



head to head mosaic ends matches the predicted hairpin sequence. The sequence became more difficult to read after position 1 of the 3' end, indicating that multiple species existed in the isolated DNA.

Imprecise Hairpin Species Revealed by Sequencing—To further study the sequence of the hairpin species, we cloned putative hairpin DNAs and subjected individual clones to Sanger dideoxy sequencing. A DNA substrate was designed such that the hairpin would have a 4-bp *EcoRI* overhang at the 5' end, a 4-bp *AatII* overhang at the 3' end, and a properly spaced promoter -35 consensus sequence for direct selection during cloning. An *in vitro* cleavage assay was performed with this substrate, and the putative hairpin DNA was isolated from a denaturing polyacrylamide gel. This putative hairpin DNA was ligated as single-stranded DNA into an *EcoRI*-*AatII* digested pGREC1 plasmid, which is missing the -35 region of the tetracycline gene promoter. A successful ligation would provide the -35 region and allow for selection by tetracycline resistance. Upon transformation, the gapped DNA will be filled in by cellular functions. We recovered and dideoxy-sequenced 16 clones. Fourteen of the clones had sequences predicted for precise hairpins. Precise hairpins had position 1 of the 3' end of the mosaic end covalently linked to position 1 of the 5' end. This gave a hairpin junction of 5'-CAGCTG-3'. The other two clones derived from imprecise hairpins. In this case, the imprecise hairpins were formed from a covalent linkage between position 1 of the 3' end of the mosaic end and position -1 of the 5' end. The imprecise hairpins gave a hairpin junction of 5'-CAGACTG-3' (Fig. 5). Imprecise hairpins can also be seen on denaturing polyacrylamide gels (see time course experiment below).

A Hairpin Is an Intermediate in *Tn5* Transposition—We were interested in determining whether the hairpin is an intermediate in the *Tn5* transposition reaction or a by-product. A kinetic analysis of an *in vitro* cleavage reaction described above shows that the hairpin is an intermediate (Fig. 6A). The 80-base hairpin appeared quickly, within the first 5 min of the reaction, and continued to increase until it peaked at 25 min. At this peak, the hairpin was 1.6% of the total labeled molecules. After 25 min, the amount of hairpin in the reaction dropped to 0.5% at 60 min (Fig. 6C). The other species in the reaction behaved differently (Fig. 6B). The substrate decreased from 100 to 50% in 45 min; after 45 min, no more substrate was metabolized. The 40-base band represents both nicked and double strand break products. These products accumulated to approximately 25% of the total label at 25 min, and then a

plateau was reached. Donor DNA was released upon hairpin formation. The donor DNA released from precise hairpin formation was 13 bases, and was detectable at low levels at 5 min. This product accumulated to 20% at 45 min and then leveled off.

Interestingly, two hairpin species and two released donor DNA products were detectable. The second hairpin species is an 81-base imprecise hairpin. This species was hardly detectable until about 25 min; at this point, it made up just less than 1% of the total label. The imprecise hairpin persists through the end of the time course (150 min). Upon imprecise hairpin formation, a 12-base donor DNA was released. This species was clearly detectable, 1.3% of the total label, at 25 min. It increased to 5% of the label at 150 min. From 45 to 150 min, many other species of varying sizes were detectable. We speculate that they are strand transfer products generated by integration at various sites in unmetabolized substrate DNA.

Preformed Hairpins Can Be Cleaved—Because a kinetic analysis of an *in vitro* cleavage reaction showed that the hairpin is an intermediate, we hypothesized that preformed hairpins should be a substrate for transposase cleavage. The hairpin oligonucleotides were designed so that upon proper cleavage, a 28-base fragment will be released. Preformed precise and imprecise hairpin oligonucleotides were annealed, 5' end-labeled, used as substrates in *in vitro* cleavage reactions, and analyzed by denaturing polyacrylamide gel electrophoresis. Both precise and imprecise hairpins can be cleaved correctly at the 3' +1 position, as seen by release of the 28-base product (Fig. 7). This processing of the hairpin is further evidence that the hairpin is an intermediate in the cleavage reaction. The imprecise hairpin is cleaved 10-fold less well than the precise hairpin. Although 50% of the precise hairpins were cleaved, only 6% of the imprecise hairpins were cleaved. We speculate that the additional band at approximately 33 bases in the precise hairpin reaction is a strand transfer product. It is likely that this band was not seen in the imprecise hairpin reaction because cleavage of the imprecise hairpin is less efficient.

Preliminary evidence shows that precise and imprecise hairpins can undergo strand transfer. One imprecise hairpin strand transfer product has been sequenced, and it has the precise transposon/duplication junction, as well as the trademark 9-bp duplication. The additional base in the imprecise hairpin is not found in the strand transfer product (data not shown).

Hairpin Formation Is the Principal Mechanism for 5' End

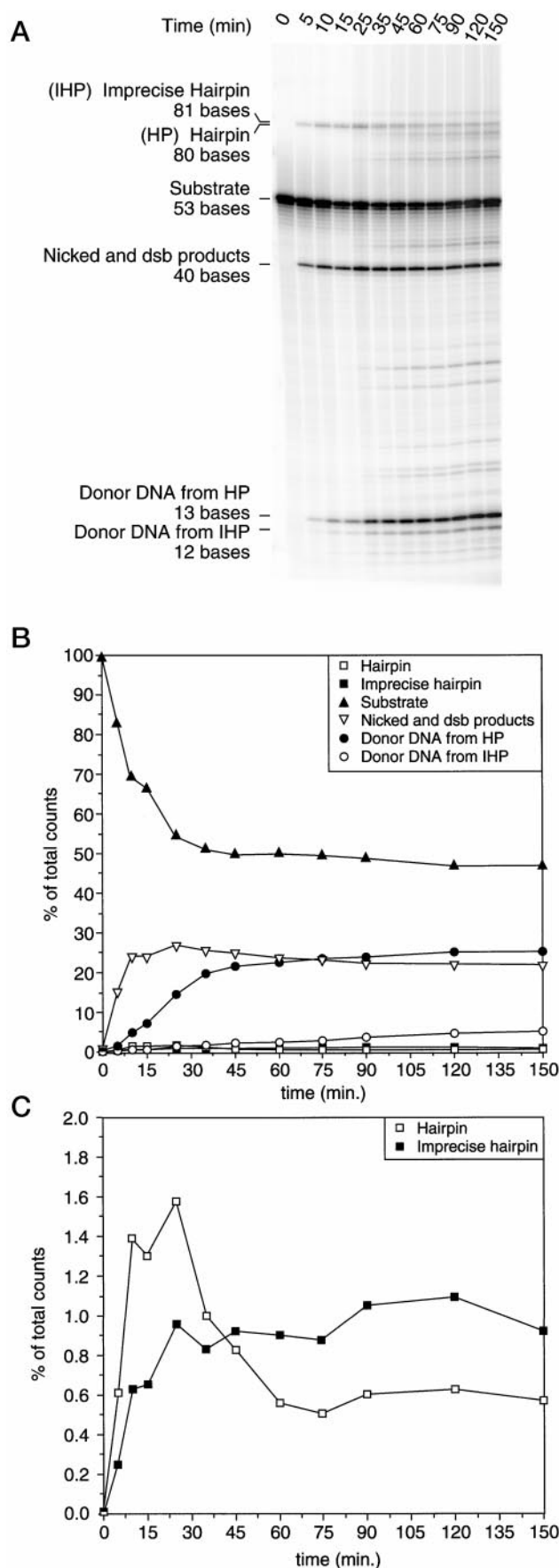


FIG. 6. A, time course of an *in vitro* cleavage reaction. A 40- μ l *in vitro* cleavage reaction consisting of 71.5 nM radiolabeled 53-base mosaic end and 377 nM EK54/LP372 transposase (see under "Experimental Procedures" for reaction conditions) was incubated at 37 °C. Aliquots were

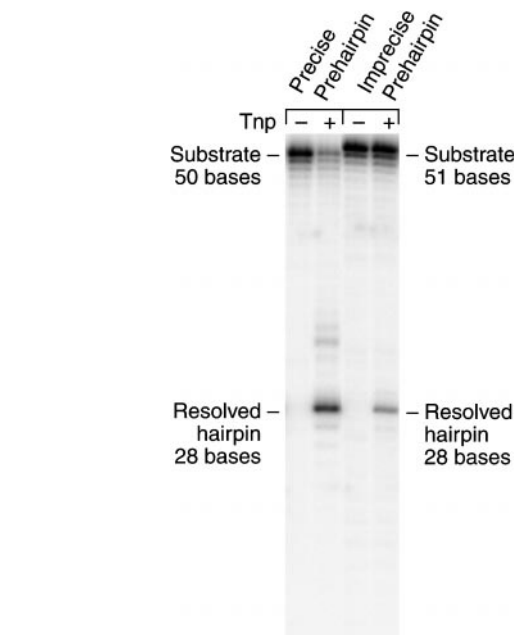


FIG. 7. **Preformed precise and imprecise hairpins can be cleaved by transposase to resolve the hairpin.** The 28-base product is the result of hairpin cleavage. The imprecise hairpin is cleaved 10-fold less well than the precise hairpin. Although 50% of the precise hairpins are cleaved, only 6% of the imprecise hairpins are cleaved. The *in vitro* cleavage reactions consisted of 13 nM hairpin DNA and 377 nM EK54/LP372 transposase (see under "Experimental Procedures" for reaction conditions) and were incubated at 37 °C for 30 min.

Cleavage—We hypothesized that 5' end cleavage requires formation of a hairpin on the transposon end and hairpin formation requires a 3'OH at position 1 of the bottom strand to attack the top strand. To test this, we designed prenicked mosaic end DNA substrates that had either a deoxyguanosine or a dideoxyguanosine at position 1 of the bottom strand. In addition, we used a pregapped mosaic end DNA substrate missing position 1 of the bottom strand (Fig. 8A). If hairpin formation requires a 3'OH, the dideoxy containing substrate would be unable to support hairpin formation and 5' end cleavage. The prenicked substrate with deoxyguanosine (40dG) (Fig. 8B, lane 6) was efficiently metabolized, as seen by strong signals for hairpin formation (at 80 bases) (16% of total labeled molecules) and donor DNA release (at 20 bases) (26% of total label). A smearing of the DNA toward the upper half of the gel indicates that a variety of integration events occurred.

Interestingly, the pregapped substrate was also metabolized, although much less efficiently (39dA) (Fig. 8B, lane 2). Hairpins were formed 150-fold less well (only 0.1% of the total label) with the pregapped substrate than with the prenicked substrate with deoxyguanosine. Donor backbone release was also much reduced to only 2% of the total label. It appears that the 3'OH on position 2 of the bottom strand attacks the proper

taken at times 0, 5, 10, 15, 25, 35, 45, 60, 75, 90, 120, and 150 min. Reactions were stopped, ethanol-precipitated, resuspended in 50% formamide loading dye, boiled for 5 min, and electrophoresed through a 15% denaturing polyacrylamide gel. A doublet representing precise and imprecise hairpins is seen at bases 80 and 81. The 40-base band represents nicked and double strand break products. The 13- and 12-base bands are released donor backbones as a result of precise and imprecise hairpin formation, respectively. B, graphical representation of all the products of the time course reaction. C, enlargement of a portion of B to show the hairpins. The 80-base precise hairpin increases with time, peaking at 25 min and 1.6% of the total counts. After this point, the amount of hairpin formation drops off to 0.5% of the total counts at 60 min. The precise hairpin appears and disappears over time, displaying the kinetics of an intermediate.

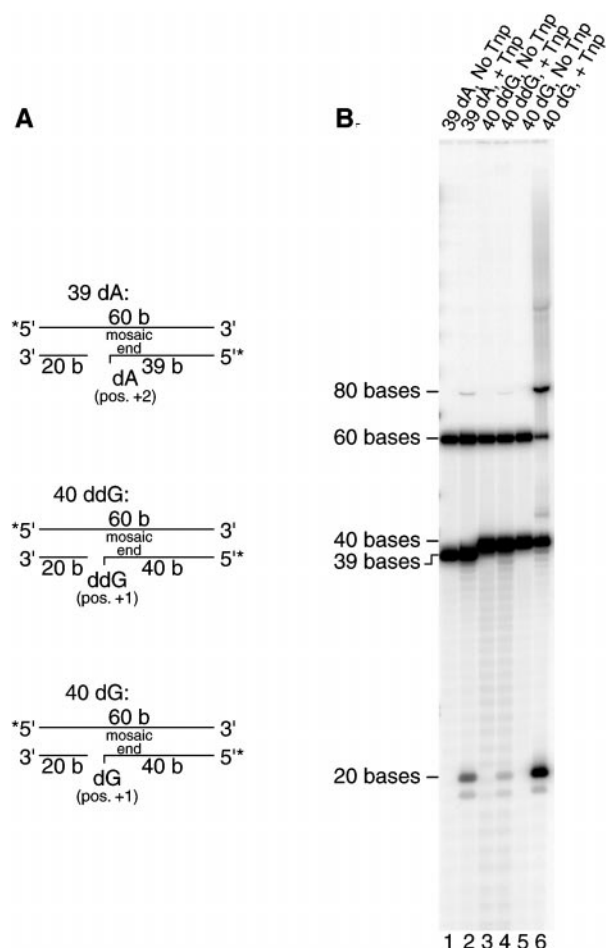


FIG. 8. A, the three different substrates used are shown. 40dG has a nick and a dG (3'OH) at position 1 on the bottom strand of the mosaic end. This substrate has bypassed this first chemical step in transposition. 40ddG has a nick and a ddG at position 1 on the bottom strand of the mosaic end. This substrate has no free position 1 3'OH. 39dA is the pregapped substrate missing position 1 on the bottom strand of the mosaic end. This substrate has a 3'OH at position 2 on the bottom strand of the mosaic end. B, the three substrates described above were used in *in vitro* cleavage reactions (see under "Experimental Procedures" for reaction conditions). Reactions were incubated at 37 °C for 30 min. Lanes 1, 3, and 5 have no transposase added, whereas lanes 2, 4, and 6 all contain EK54/LP372 transposase. The 40dG substrate (lane 6) is efficiently metabolized. The hairpin makes up 16% of the total counts, and released donor DNA makes up 26% of the total counts. The pregapped substrate (39dA) (lane 2) was also metabolized, although significantly less efficiently. The 40ddG substrate (lane 4) supported some minor hairpin formation (only 0.04% of the total counts).

position 1 on the top strand, as indicated by release of the 20-base donor DNA and formation of the 79-base hairpin on this gel. Because no 21-base product was seen, we can conclude that the 3'OH on position 2 of the bottom strand does not attack position 2 on the top strand. This result indicates that although the 3' position 1 is required for efficient hairpin formation, DNA lacking this position does support some hairpin formation.

Unexpectedly, the prenicked dideoxyguanosine substrate did support some minor hairpin formation (only 0.04% of total label) (40ddG) (Fig. 8B, lane 4), as well as some minor donor backbone release (0.48% of the total label). The dideoxyguanosine was added to a 39-base mosaic end containing oligonucleotide by terminal transferase. This reaction was then tailed with dNTPs and terminal transferase, and the ddG oligonucleotide was polyacrylamide gel-purified. Therefore, we do not believe that the hairpin formation is a result of contaminating dG at position 1 of the bottom strand.

We suspect that our EK54/LP372 transposase protein preparation is contaminated with some very minor 3'–5' exonuclease activity (see under "Experimental Procedures"). We believe that the exonuclease removed the 3' ddG at position 1 in a portion of the 40ddG substrate. With this position removed, the 40 ddG substrate is identical to the pregapped (39dA) substrate. The contaminating pregapped substrate is probably forming the 79-base hairpin seen in this reaction. The dramatic 53-fold reduction in the 5' cleavage product (seen as donor DNA release) formed with the 40ddG substrate (*versus* the 40dG) indicates that hairpin formation is the principal mechanism for 5' cleavage.

In all the reactions, a 19-base band was detectable. This could result from imprecise hairpin formation. The 81- and 80-base imprecise hairpins are weak in comparison to the precise hairpins and could not be seen on this gel. The difference between the amount of imprecise hairpins detected here and in the time course could be due to variations in the length and sequence of the donor DNAs.

DISCUSSION

Transposition Occurs via a Hairpin Intermediate—The data presented here demonstrate that *Tn5* transposes via a DNA hairpin intermediate. An *in vitro* cleavage assay led to the discovery of a low mobility DNA species. Through native gel analysis and chemical and dideoxy sequencing, we verified that the low mobility species is a hairpin formed at the transposon end sequences. We also verified that the hairpin is an intermediate in the transposition reaction. A time course showed the hairpin increasing and then decreasing over time displaying the kinetics of an intermediate. In addition, preformed hairpins can be further processed to blunt end transposons competent for strand transfer. Hairpin formation is the primary mechanism for 5' end cleavage, as seen in the dideoxy experiment.

Discovery of this hairpin intermediate provides evidence for a model for the cleavage step in *Tn5* transposition. The first step involves transposase-catalyzed 3' hydrolytic nicks at position 1 of the end sequences to release a pair of 3'OHs at both ends of the transposon. These free 3'OHs attack the phosphodiester bond connecting position –1 and +1 on the end sequences of the nontransferred strands (5' ends) forming hairpins on the ends of the transposon DNA. The donor DNA is released during the hairpin formation step. This step is key, as it serves as the 5' cleavage step. In order for the transposon to integrate, its hairpin ends need to be resolved. Transposase hydrolytically cleaves (opens) the hairpins in a step identical to the first chemical step (3' nicking), releasing a pair of 3'OHs. These 3'OHs are capable of strand transfer into the target DNA, a chemical step similar to hairpin formation. In this scenario, the cleavage step is reduced to two repetitive chemistries, nicking and strand transfer. This model is very attractive because it provides a mechanism by which one transposase active site can catalyze all the chemical reactions at one end without massive protein/DNA rearrangements. In addition, this mechanism also favors the two-metal ion mechanism of cleavage similar to that used by ribozymes and ribonucleases (19, 22).

Imprecise Hairpin Formation—Interestingly, we found that imprecise hairpins are formed along with precise hairpins. Imprecise hairpins are hairpins that form when the 3'OH on the transferred strand attacks a position other than +1 of the end sequence on the nontransferred strand. The imprecise hairpin we have found is formed by the 3'OH attacking at the phosphodiester bond between position –2 and –1 (1 base into the donor DNA) on the nontransferred strand. This leads to a hairpin with an extra base between the head-to-head end sequences. We also found that the imprecise hairpin was able to be resolved by transposase, although 10-fold less well than the

precise hairpin. This suggests that transposase recognizes/interacts with a specific stem-loop structure of the hairpin. Altering the stem-loop structure (*i.e.* by increasing the loop size) could alter the recognition/interaction of transposase with the stem-loop, thereby changing the rate of cleavage of the hairpin intermediate. We presume that the imprecise hairpin is resolved at a slower rate than the precise hairpin. This was seen in the time course experiment; the precise hairpin was resolved by 25 min, whereas the imprecise hairpin persisted (Fig. 6C).

Preliminary evidence shows that a resolved imprecise hairpin can undergo strand transfer. The strand transfer product has the characteristic 9-bp duplication, and the extra base (5' overhang) is not incorporated into the product. There are a few different processes that could potentially remove the extra base. First, after integration, the extra base would be in the gapped region of the strand transfer product. It is currently thought that DNA polymerase fills in the gaps generated in the target DNA after transposition (9, 19). The 5'–3' exonuclease activity of DNA polymerase could remove the extra base. In addition, another cellular 5'–3' exonuclease activity could remove the extra base before or after integration. It is also possible that transposase is responsible for removal of the extra base prior to integration. It is conceivable that the 3'OH released after hairpin cleavage could again attack the nontransferred strand, but precisely this time. This attack would lead to hairpin formation again and loss of the 5' single base overhang. This precise hairpin can be resolved and integrated. The possibility that an imprecise hairpin must undergo a second cleavage reaction before integration would lead to an overall slower rate of integration.

Likewise, in retroviral integration, the viral DNA ends are processed such that the viral DNA ends have two-nucleotide 5' overhangs. This structure resembles the cleaved imprecise hairpin, which has a one-nucleotide 5' overhang. It has been suggested that the viral overhangs are removed by host repair mechanisms and/or by the DNA splicing mechanism of integrase (23, 24).

Efficient Hairpin Formation Requires Position 1 of the End Sequence on the Transferred Strand—Surprisingly, the pregapped substrate missing position 1 of the end sequence on the transferred strand formed a detectable amount of hairpins *in vitro* (Fig. 8B), although hairpins were formed 150-fold less well with the pregapped substrate than with the prenicked substrate. It is interesting that the 3'OH of position 2 is capable of forming hairpins. This data leads to the conclusion that the lack of position 1 does not completely abolish hairpin formation, but position 1 is required for efficient hairpin formation. This correlates well with Tn10 and V(D)J recombination data. In V(D)J recombination, the conserved heptamer of the recognition signal sequence is adjacent to the coding DNA. Positions 1 and 2 of the heptamer are positionally equivalent to positions 1 and 2 of the Tn5 mosaic end sequence. Mutations at positions 1 and 2 of the heptamer specifically block hairpin formation; thus, positions 1 and 2 are required for hairpin formation (25). Similarly, Tn10 data suggests that mutations at positions 1 and 2 of the end sequence affect steps after interaction of the ends but before full excision (26). Perhaps removing both positions 1 and 2 of the Tn5 mosaic end sequence would entirely abolish hairpin formation.

Interestingly, the 3'OH of position 2 on the pregapped substrate attacks position 1 (the proper transposon-donor DNA junction) on the top (5') strand rather than position 2 (the directly opposing phosphodiester bond). Likewise, in V(D)J recombination, imprecise nicks (displaced 1, 2, or 3 bases into the coding DNA) occur. The 3'OHs from these displaced nicks also attack at the proper heptamer-coding junction (25). Somehow,

transposase and the RAG proteins can accommodate a displaced 3'OH in their active site in such a way as to perform the correct nucleophilic attack. This could hint at potential flexibility in the DNA and/or protein.

Implications of Transposase Binding to Hairpin DNA—The ability of Tn5 transposase to cleave hairpin DNA implies that transposase binds hairpin DNA in addition to normal double-stranded end sequence DNA (during the initial step in transposition). Because transposase binds these two very different DNA structures, it is possible that transposase undergoes a conformational change in order to interact with the hairpin DNA, and this conformational change “fits” with the structure of the loop. It is believed that throughout transposition, transposase and the DNA are likely undergoing many conformational changes, which lead to a variety of nucleoprotein structures. Oligomers of transposase stably bound to two double-stranded end sequence DNAs (in the absence of Mg²⁺) form the first higher order nucleoprotein structure in Tn5 transposition, the synaptic complex. Transposase bound to hairpin DNA is probably a higher order nucleoprotein structure different than the synaptic complex, albeit a normal intermediate structure. The idea of transposition involving many nucleoprotein structures is not a new one. In Mu transposition, four stable nucleoprotein structures have been identified (27). Tn5 transposition presumably also proceeds through multiple complex transposase-DNA configurations.

Hairpin Formation as a Paradigm in Transposition—The finding of hairpins in transposition and transposition-like systems is becoming more prevalent, suggesting a new paradigm in transposition. Recently, Tn10 was found to transpose via a hairpin intermediate (19). IS911 proceeds by a mechanism similar to hairpin formation but uses a circle intermediate instead. In this system, a 3' nick is made at one transposon end. The freed 3'OH attacks the opposite transposon end (in a trans reaction) on the same strand, forming a single-stranded circle. The complementary strand still has donor DNA attached. Host machinery could resolve this structure for strand transfer (28). In Tn5, a circle intermediate does not form, but rather hairpins are formed on both transposon ends (*cis* reaction). By mixing DNAs of different transposon lengths with transposase and looking for different length hairpins, we have found that the released 3'OH attacks the opposite phosphodiester bond on its own end sequence (*cis* reaction) rather than on the partner end sequence (*trans* reaction).⁴

The V(D)J recombination system is chemically very similar to transposition. In fact, recently, the V(D)J recombination proteins RAG1 and RAG2 have been shown to carry out transposition (29, 30). V(D)J recombination also proceeds via a hairpin intermediate, although their hairpins are on the coding DNA (equivalent to donor DNA) and not on the recognition signal sequence DNA (equivalent to transposon DNA). In V(D)J recombination, a 5' nick is made adjacent to the recognition signal sequence. The free 3'OH on the coding DNA end performs a nucleophilic attack on the bottom strand of the coding DNA releasing the recognition signal sequence. In this system, the nucleophilic attack occurs on the phosphodiester bond directly opposite (18). In Ascot-1, a *hAT* family transposon from the fungus *Ascobolus immersus*, a hairpin mechanism is proposed to explain the excision footprints left after transposition. The hairpin mechanism proposed is almost identical to V(D)J recombination. 5' nicks are made at the phosphodiester bond one base outside the transposon. Interestingly, the released 3'OH does not attack the phosphodiester bond directly oppo-

⁴ A. Bhasin, I. Y. Goryshin, M. Steiniger-White, D. York, and W. S. Reznikoff, manuscript in preparation.

site, but rather attacks the phosphodiester bond immediately 5' to it (31). This is essentially the same as the imprecise hairpin formation that we found in *Tn5*. These transposons likely utilize a hairpin mechanism so that the transposase does not have to undergo major conformational changes in order to perform a variety of chemical reactions.

Acknowledgments—We thank Todd Naumann, Mindy Steiniger-White, and Sally Twining for valuable insights during discussions and Barb Schriver for technical support. We also thank Laura Vanderploeg, Robin Davies, and Adam Steinberg for excellent work on the figures.

REFERENCES

1. Berg, D. E., and Howe, M. M. (1989) *Mobile DNA*, pp. 185–210, American Society for Microbiology, Washington, D. C.
2. Reznikoff, W. S. (1993) *Annu. Rev. Microbiol.* **47**, 945–963
3. Goryshin, I. Y., Reznikoff, W. S. (1998) *J. Biol. Chem.* **273**, 7367–7374
4. Zhou, M., Bhasin, A., and Reznikoff, W. S. (1998) *J. Mol. Biol.* **276**, 913–925
5. Mizuuchi, M., Baker, T. A., and Mizuuchi, K. (1992) *Cell* **70**, 303–311
6. Sakai, J., Chalmers, R. M., and Kleckner, N. (1995) *EMBO J.* **14**, 4374–4383
7. Mizuuchi, K., and Adzuma, K. (1991) *Cell* **66**, 129–140
8. Mizuuchi, K. (1992) *Annu. Rev. Biochem.* **61**, 1011–1051
9. Sasakawa, C., Uno, Y., Yoshikawa, M. (1981) *Mol. Gen. Genet.* **182**, 19–24
10. Syvanen, M., Hopkins, J. D., and Clements, M. (1982) *J. Mol. Biol.* **158**, 203–212
11. Mizuuchi, K. (1983) *Cell* **35**, 785–794
12. Morisato, D., and Kleckner, N. (1984) *Cell* **39**, 181–190
13. Bainton, R., Gamas, P., and Craig, N. L. (1991) *Cell* **65**, 805–816
14. Bolland, S., and Kleckner, N. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 7814–7818
15. Namgoong, S.-Y., and Harshey, R., M. (1998) *EMBO J.* **17**, 3775–3785
16. Bolland, S., and Kleckner, N. (1996) *Cell* **84**, 223–233
17. Sarnovsky, R. J., May, E. W., and Craig, N. L. (1996) *EMBO J.* **15**, 6348–6361
18. McBlane, J. F., van Gent, D. C., Ramsden, D. A., Romeo, C., Cuomo, C. A., Gellert, M., and Oettinger, M. A. (1995) *Cell* **83**, 387–395
19. Kennedy, A. K., Guhathakurta, A., Kleckner, N., and Haniford, D. B. (1998) *Cell* **95**, 125–134
20. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, p. 7.76, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
21. Maxam, A. M., and Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560
22. Steitz, T. A., and Steitz, J. A. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 6498–6502
23. Chow, S. A., Vincent, K. A., Ellison, V., Brown, P. O. (1992) *Science* **255**, 723–726
24. Kulkosky, J., Katz, R. A., Merkel, G., Skalka, A. M. (1995) *Virology* **206**, 448–456
25. Ramsden, D. A., McBlane, J. F., van Gent, D. C., and Gellert, M. (1996) *EMBO J.* **15**, 3197–3206
26. Haniford, D., and Kleckner, N. (1994) *EMBO J.* **13**, 3401–3411
27. Watson, M. A., and Chaconas, G. (1996) *Cell* **85**, 435–445
28. Ton-Hoang, B., Polard, P., and Chandler, M. (1998) *EMBO J.* **17**, 1169–1181
29. Agrawal, A., Eastman, Q. M., and Schatz, D. G. (1998) *Nature* **394**, 744–751
30. Hiom, K., Melek, M., and Gellert, M. (1998) *Cell* **94**, 463–470
31. Colot, V., Haedens, V., and Rossignol, J.-L. (1998) *Mol. Cell. Biol.* **18**, 4337–4346