The EMBO Journal Vol.17 No.7 pp.2122–2136, 1998

Transposase makes critical contacts with, and is stimulated by, single-stranded DNA at the P element termini *in vitro*

Eileen L.Beall and Donald C.Rio[1]

Department of Molecular and Cell Biology, 401 Barker Hall, University of California, Berkeley, CA 94720-3204, USA

[1]Corresponding author
e-mail: don_rio@uclink4.berkeley.edu

P elements transpose by a cut-and-paste mechanism. Donor DNA cleavage mediated by transposase generates 17 nucleotide (nt) 3′ single-strand extensions at the P element termini which, when present on oligonucleotide substrates, stimulate both the strand-transfer and disintegration reactions *in vitro*. A significant amount of the strand-transfer products are the result of double-ended integration. Chemical DNA modification–interference experiments indicate that during the strand-transfer reaction, P element transposase contacts regions of the substrate DNA that include the transposase binding site and the duplex portion of the 31 bp inverted repeat, as well as regions of the terminal 17 nt single-stranded DNA. Together these data suggest that the P element transposase protein contains two DNA-binding sites and that the active oligomeric form of the transposase protein is at least a dimer.

Keywords: disintegration/*Drosophila*/strand transfer/transposition/transposon

Introduction

The transposase/integrase family of proteins catalyzes the phosphodiester bond cleavage and joining reactions that occur during DNA transposition and the integration of retroviral proviruses (Mizuuchi, 1992a,b; Craig, 1995). Biochemically, transposition can occur either by replicative or cut-and-paste mechanisms (Craig, 1997). The fundamental chemical steps of transposition and retroviral integration are similar in that water is used as the nucleophile by the transposase/integrase protein during the first step of the reaction, donor DNA cleavage. During the second step of the reaction, strand transfer, the 3′ hydroxyl groups on the cleaved termini of the transposable element or proviral genome are used as the nucleophiles when the element is inserted into a new target site. Recently, a reaction termed disintegration has been demonstrated to occur by both the retroviral integrases (Chow *et al.*, 1992; Chow and Brown, 1994a) and the Tc-1 transposase protein (Vos and Plasterk, 1994). During the disintegration reaction, a substrate that mimics the strand-transfer intermediate is cleaved to give products that represent the cleaved donor DNA and linear double-stranded target site DNA. The disintegration reaction has been useful for detecting catalytic activity of wild-type and a variety of mutant integrase proteins (Katz and Skalka, 1994). Hydrolysis of a high-energy cofactor is not required during transposition or retroviral integration. In addition, the chemical transesterification steps of the reaction are isoenergetic, occur without the formation of a covalent protein–DNA intermediate and are reversible.

Structural studies of the catalytic domains of the retroviral integrases and the Mu transposase protein reveal that despite very little primary sequence homology, there is a remarkable overall structural similarity between these proteins. The similarities have been extended to include members of the RNase H and Ruv C class of nucleases (Dyda *et al.*, 1994; Bujacz *et al.*, 1995, 1996; Rice and Mizuuchi, 1995; Rice *et al.*, 1996). Previous biochemical studies, mutagenesis experiments and sequence comparisons suggest that the integrase family of proteins uses a conserved set of acidic amino acids (the catalytic triad or DDE motif) to coordinate the metal ions that are involved in catalyzing the phosphodiester transesterification reactions (Kulkosky *et al.*, 1992; Baker and Luo, 1994; Katz and Skalka, 1994; Bolland and Kleckner, 1996; Sarnovsky *et al.*, 1996). Metal ion coordination is critical for other types of phosphodiester bond reactions performed by polymerases, nucleases and ribozymes (Linn *et al.*, 1993; Pyle, 1993; Steitz and Steitz, 1993). Mutation of any one of the three DDE catalytic residues in several bacterial transposases or retroviral integrases results in substantial or complete loss of activity.

P transposable elements in *Drosophila melanogaster* transpose by a cut-and-paste mechanism (Engels *et al.*, 1990; Kaufman and Rio, 1992). The 87 kDa P element-encoded transposase protein requires the cofactors GTP and Mg^{2+} for activity (Kaufman and Rio, 1992; Mul and Rio, 1997). During P element transposition, transposase binds to sequences near the P element termini (Kaufman *et al.*, 1989) and catalyzes the cleavage and strand-transfer steps of the transposition reaction (Kaufman and Rio, 1992; Beall and Rio, 1997). Approximately 150 bp of DNA are required at each end of the P element for efficient transposition to occur *in vivo* (Mullins *et al.*, 1989). This region of DNA contains the terminal 31 bp inverted repeats, the internal transposase binding sites and the internal 11 bp inverted repeat transposition enhancer elements. Several genetic and biochemical studies indicate that following P element transposition, double-stranded DNA break repair functions are required that include the participation of the Ku and DNA-dependent protein kinase (DNA-PK) proteins (Beall *et al.*, 1994; Weaver, 1995; Beall and Rio, 1996). Recent biochemical studies of the P element transposase-catalyzed donor DNA cleavage reaction demonstrate that transposase makes a staggered cleavage at the P element termini to generate 17 nucleotide (nt) 3′ extensions at each P element end (Beall and Rio, 1997). This novel cleavage pattern explains the nature of chromosomal P element excision events because small amounts of P element sequence are often left behind at

the cleavage site following P element excision *in vivo* (Takasu-Ishikawa *et al.*, 1992; Staveley *et al.*, 1995). In addition, the cleaved termini may trigger DNA repair by recruitment of the Ku–DNA-PK complex, since short DNA overhangs can serve as binding sites for the mammalian Ku protein *in vitro* (Mimori and Hardin, 1986; Jackson, 1996).

Because of the unusual nature of the cleaved termini following P element excision, we wanted to test whether substrates that mimic the authentic cleaved termini were efficient substrates for transposase during the strand-transfer step of the transposition reaction. Here we demonstrate that oligonucleotide substrates containing a transposase binding site and the 31 bp terminal inverted repeat are used as substrates during the strand-transfer reaction *in vitro*. We show that substrates containing 17 nt single-stranded extensions, but not flush or smaller 3′ extensions, stimulate both the transposase-mediated strand-transfer and disintegration reactions *in vitro*, indicating that the DNA representing the cleaved P element terminus is an intermediate in the transposition reaction. Using chemical DNA modification–interference methods to determine both the phosphate and the base-specific transposase–DNA contacts during the strand-transfer reaction, we show that transposase contacts both the double- and the single-stranded regions of the substrate DNA. The unusually long, single-stranded nature of the cleaved P element termini appears to be critical for efficient transposase-mediated strand transfer and may trigger DNA repair following P element excision and integration *in vivo*.

Results

Expression and purification of P element transposase

In order to facilitate the biochemical analysis of the P element transposition reaction, we expressed and purified transposase from two different eukaryotic expression systems. Recently we found that expression vectors containing a chemically resynthesized N-terminal fragment of the transposase-protein coding sequence allowed for higher levels of transposase expression in *Drosophila* cell culture (Lee *et al.*, 1996; Mul and Rio, 1997). In addition, the involvement of the *Drosophila* Ku–DNA-PK in P element transposition suggests that post-translational modification(s) may be required for transposase activity (Beall and Rio, 1996). Therefore, we expressed and purified P element transposase from *Drosophila* cell culture and as a maltose binding protein (MBP)-fusion from Sf9 insect cells infected with a recombinant baculovirus (Figure 1). The recombinant transposase proteins were purified using conventional and/or affinity chromatography (see Materials and methods). The purity of each transposase preparation was determined by SDS–polyacrylamide gel electrophoresis and silver or Coomassie Blue staining (Figure 1).

Disintegration is stimulated by oligonucleotide substrates containing single-stranded DNA at the transposon terminus

The phosphodiester bond transesterification reactions of transposition do not require high-energy cofactor hydrolysis or involve covalent protein–DNA intermediates (Cech, 1987; Mizuuchi, 1992a,b; Craig, 1995). Using

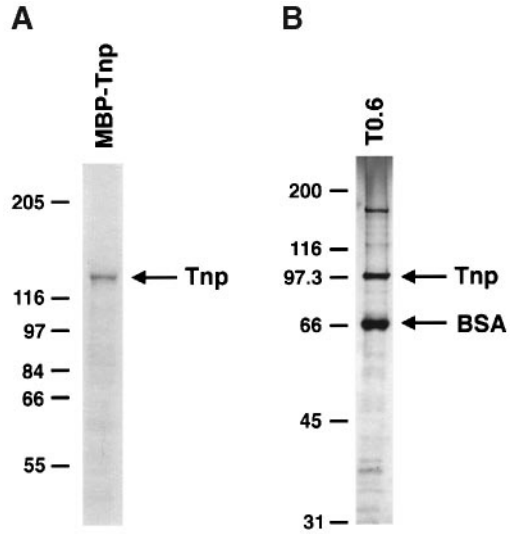
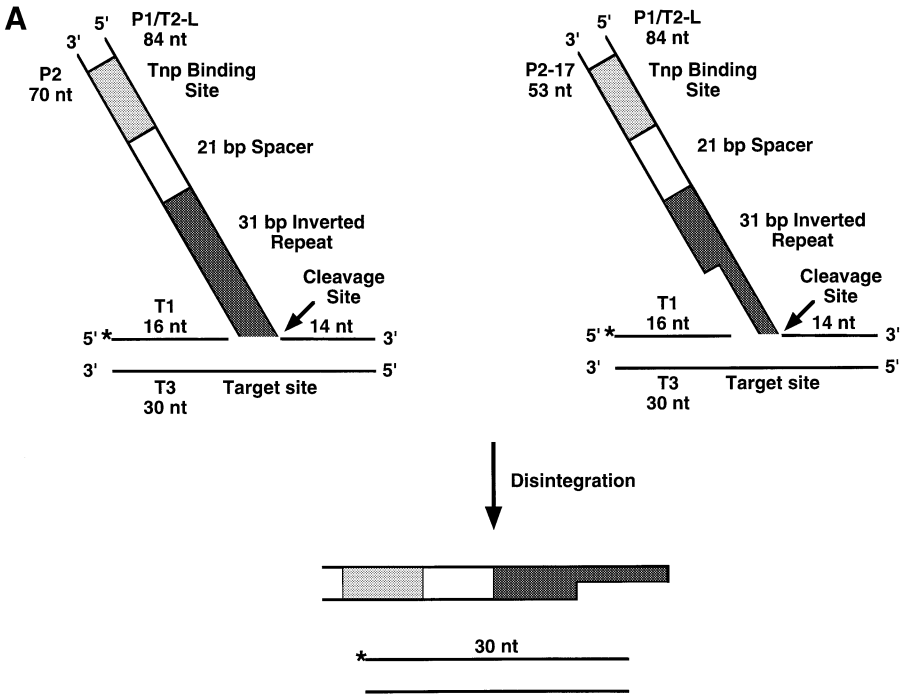


Fig. 1. Transposase-containing protein fractions used to study *in vitro* activities. Analysis of the purified P element transposase-containing fractions by SDS–polyacrylamide gel electrophoresis on 7.5% acrylamide gels is shown. The gels were stained with either Coomassie Blue (MBP–Tnp) or silver (T0.6). Positions of the molecular weight (M_r) markers with M_r in kDa are indicated on the left of each panel. (**A**) A recombinant baculovirus encoding the transposase protein as a MBP fusion was used to express transposase in Sf9 cell culture (MBP–Tnp). Whole cell extracts were prepared from infected cells and MBP–transposase purified following chromatography on amylose agarose (see Materials and methods). Approximately one fiftieth of the total peak fraction is shown. (**B**) Transposase was purified from *Drosophila* cell culture nuclear extracts using heparin–agarose chromatography. The flow-through fraction was chromatographed on a nonspecific DNA-affinity column (TdT), and the protein eluted with increasing KCl. BSA was added to the fraction at 50 μg/ml. Approximately one fiftieth of the 0.6 M KCl fraction is shown.

model oligonucleotide substrates and Mn^{2+}, several retroviral integrases have been shown to catalyze a reaction termed disintegration (Chow *et al.*, 1992). During disintegration, the bond between the viral long terminal repeat (LTR) DNA and host DNA is broken and the target site is rejoined to give products that resemble an unintegrated viral LTR end and a duplex DNA target site, a schematic diagram of which is shown in Figure 2A for the P element end-containing disintegration substrates. Disintegration has been used as an indicator of the catalytic activity of several retroviral integrase proteins (Chow *et al.*, 1992; Katz and Skalka, 1994).

In order to develop a disintegration assay for P element transposase, we designed oligonucleotide substrates that resemble an integrated P element end and contain either a flush (Figure 2A, left) or a 17 nt 5′ recessed, non-transferred strand (Figure 2A, right). The P2–17 substrate represents an authentic P element transposase-cleaved terminus that has been integrated into a new target site (Beall and Rio, 1997). The oligonucleotides were derived from the left (5′) P element end and contain the 31 bp inverted repeat and transposase binding sites attached to a natural P element target site. The branched oligomer disintegration substrates were gel purified following radiolabeling of the 16 nt T1 oligonucleotide and annealing of the remaining three oligonucleotides (T3, P2 and P1/T2-L). Transposase-mediated disintegration produces a 30 nt product that is detected following denaturing gel electrophoresis (Figure 2A).



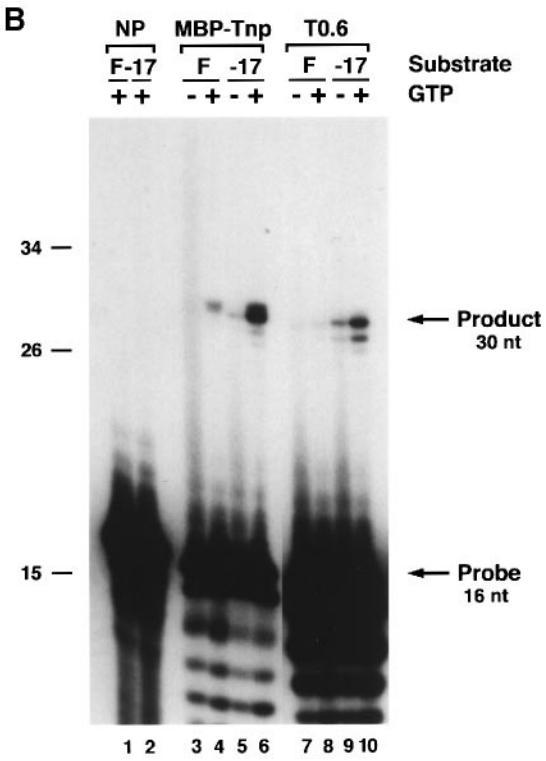


Fig. 2. Single-stranded DNA at the P element terminus stimulates disintegration. (**A**) Schematic diagram of the substrates used in the disintegration assay. Oligonucleotides derived from the left (5′) P element end (P1/T2-L and P2 or P2–17) that contain the transposase-binding site (light gray box) and the 31 bp inverted repeat (dark gray box) were annealed to oligonucleotides containing a natural target site (T1 and T3) to produce substrates that mimic an integrated P element terminus. Two different substrates were tested in the disintegration assay: either a flush-ended substrate (left side of figure) in which the P element end is flush with the target site junction or a 17 nt, 5′ recessed substrate (right side of figure) that represents an authentic cleaved P element terminus. The branched oligomer substrates contain a 5′ end radiolabeled oligonucleotide T1 (asterisk) and were gel-purified prior to use. Disintegration results in cleavage of the P element end from the target site (arrow) and ligation of the resulting nick at the target site junction to produce a 30 nt product that is detected following denaturing polyacrylamide–urea gel electrophoresis. For simplicity, only the products of the P2–17 disintegration reaction are shown at the bottom of the figure. (**B**) *Drosophila*- (T0.6) and baculovirus-produced (MBP–Tnp) transposase were tested in the disintegration assay as illustrated above. Reactions were performed at 30°C for 2 h, and products analyzed by denaturing polyacrylamide–urea gel electrophoresis and autoradiography. Shown is an autoradiogram from an experiment in which both of the branched oligomer substrates were tested either in the presence (+, even lanes) or absence (–, odd lanes) of the cofactor GTP. F, flush-ended substrate; –17, substrate containing the 17 nt, 5′ recessed substrate; NP, reactions in which transposase was omitted. The sizes expected for the free 16 nt T1 oligonucleotide and the 30 nt disintegration product are indicated on the right with arrows. Values from PhosphorImager quantitation of the products are reported in Table I. Chemical DNA sequencing analysis of the purified disintegration product from the MBP–Tnp reaction confirmed that the product of the reaction is authentic (data not shown). For each substrate, ~100 ng MBP–Tnp and ~10 ng T0.6 were tested for activity. Positions of the radiolabeled pBR322 MspI cleaved molecular weight marker DNAs are indicated on the left.

Both purified transposase protein fractions were incubated with the disintegration substrates and $MnCl_2$ in the presence or absence of GTP (Figure 2B). When $MgCl_2$ was used in the assay, transposase-induced disintegration was not observed (E.L.Beall and D.C.Rio, unpublished). Both the baculovirus- and *Drosophila*-produced transposase proteins were stimulated by the addition of GTP to produce a 30 nt product (Figure 2B). A low amount of disintegration was observed without added GTP, which may be due to low levels of GTP bound to the transposase protein. Chemical DNA sequencing of the 30 bp MBP–Tnp reaction product confirmed the accuracy of the dis-

integration reaction (E.L.Beall and D.C.Rio, unpublished). Disintegration was stimulated either 6-fold (MBP–Tnp) or 7-fold (T0.6) when the P2–17 substrate was used in the reaction (Figure 2B, compare lanes 4 and 6 and lanes 8 and 10). Titration experiments performed with the MBP–Tnp protein fraction showed decreased disintegration with decreased transposase concentration demonstrating that the MBP–Tnp is less efficient than T0.6 for disintegration. In addition, vaccinia virus- and *Escherichia coli*-produced transposase proteins were less efficient for disintegration (E.L.Beall and D.C.Rio, unpublished). Taken together, these data demonstrate that both the *Drosophila*- and baculovirus-produced transposase proteins catalyze disintegration in the presence of Mn^{2+} and GTP. In addition, substrates containing an authentic P element cleaved end (P2–17) stimulate transposase-mediated disintegration, suggesting that the DNA substrate structure is critical for catalysis by transposase.

Strand transfer is stimulated by oligonucleotide substrates containing single-stranded DNA at the transposon terminus

The ability of transposase to use oligonucleotide substrates during the disintegration reaction prompted us to examine whether transposase could also use terminal P element oligonucleotides during the strand-transfer step of the transposition reaction. We used several different oligonucleotide substrates that all contain a P element terminal 31 bp inverted repeat and a transposase-binding site derived from the right (3′) P element end (Figure 3A and B). Oligonucleotides derived from the left (5′) P element end were less efficient substrates for transposase during the strand-transfer reaction (E.L.Beall and D.C.Rio, unpublished). For each substrate, the oligonucleotide that is joined to the target DNA (P1) was radiolabeled at the 5′ end, annealed to the non-transferred strand oligonucleotide (P2), and incubated with transposase, GTP, Mg^{2+} and a tetramer of the plasmid, Bluescript, as the target DNA (Kaufman and Rio, 1992). The reaction products were analyzed by agarose gel electrophoresis and autoradiography. If transposase catalyzes the strand-transfer reaction, products resulting from either relaxed circular single-ended, or linear double-ended strand-transfer events will be detected by the transfer of the radioactivity from the free substrate oligonucleotide to that of the plasmid target DNA (Figure 3A).

The T0.6 transposase-containing protein fraction was tested in the strand-transfer assay using several different DNA substrates. The substrates were either flush (P2 flush), recessed 17 nt on the non-transferred strand (P2–17), recessed 9 nt on the non-transferred strand (P2–9), mutated at the transposase binding site (mutant), recessed 1 nt on the transferred strand (P1–1) or mutated at the nucleotide joined to the target DNA (P1 G to T) (Figure 3B). We observed reaction products corresponding to both linear, double-ended and relaxed circular, single-ended strand-transfer products, with a significant amount of the products resulting from double-ended strand transfer (Figure 3C). The transposase protein required GTP as a cofactor (Figure 3C, lane 1) and was stimulated ~25-fold by the substrate representing the authentic precleaved P element terminus (Figure 3C, compare lanes 2 and 3). Transposase showed an ~14-fold reduced activity when the P2–9 substrate was used (Figure 3C, compare lanes 3 and 4). Transposase also showed reduced strand transfer when either the mutated transposase binding site substrate (Figure 3C, lane 5 at ~18-fold reduced activity) or the P1 G to T substrate was used (Figure 3C, lane 7 at ~45-fold reduced activity). Transposase was unable to use a substrate that lacked the terminal transferred nucleotide (P1–1) (Figure 3C, lane 6). Strand-transfer products were not observed when the transposase protein was omitted from the reaction (Figure 3C, lanes 8–13). The MBP–transposase protein performed strand transfer at levels similar to, but lower than, the levels detected for the *Drosophila*-produced T0.6 protein fraction (Figure 3C, lanes 14 and 15). In addition, vaccinia virus- and *E.coli*-produced transposase proteins also demonstrated a low level of strand-transfer activity (E.L.Beall and D.C.Rio, unpublished). It is interesting that the mutated transposase binding site substrate did not completely abolish strand-transfer activity (Figure 3C, lane 5). Because this substrate still contains the authentic 17 nt 5′ recessed terminus, transposase may recognize not only the specific transposase-binding site nucleotide sequence, but also the unusual structure of the P element terminus during the strand-transfer reaction. In addition, these data suggest that transposase is functioning as a multimer during strand transfer since one reaction product, double-ended strand transfer, results from the concerted integration of two oligonucleotide substrates into the plasmid target DNA within close proximity.

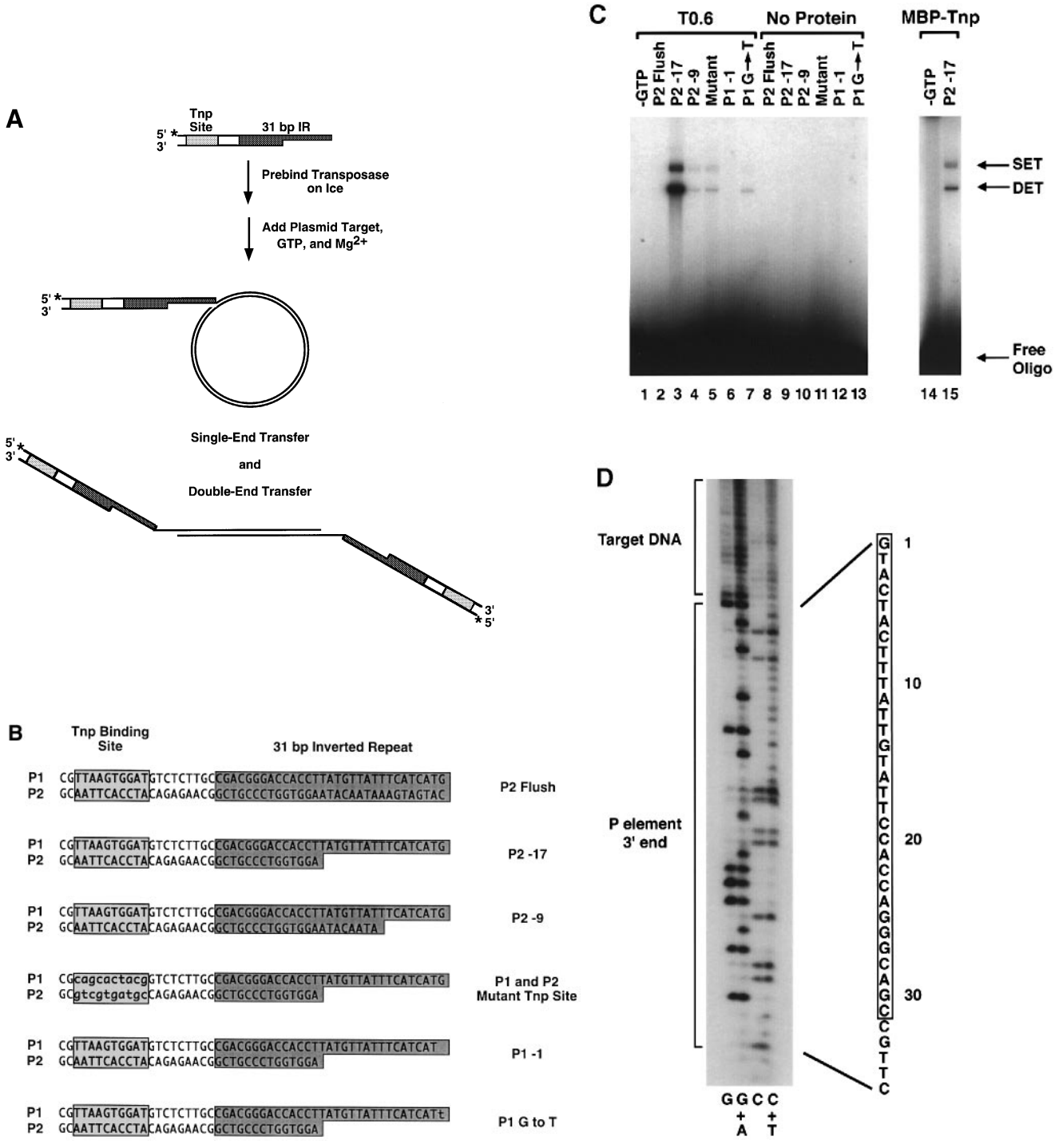
In order to confirm that the strand-transfer products were authentic, the T0.6 transposase-catalyzed products were subjected to chemical DNA-sequencing analysis. We found that the chemical sequencing ladder extended from the radiolabeled nucleotide to the transferred P element nucleotide (Figure 3D). Following the transferred G nucleotide, the sequence became a random distribution of bases that represented different target site DNA insertions demonstrating that the strand-transfer products were authentic. Interestingly, the P element end was predominantly transferred to target DNA sequences that began with a G (Figure 3D). A consensus target site DNA sequence was compiled from a variety of target sites selected by P element transposase *in vivo*, and was found to be the sequence 5′-GGCCAGAC (O'Hare and Rubin, 1983). Our data indicate that during the *in vitro* strand-transfer reaction, transposase used target sites similar to the sites used *in vivo* and that the preferred target site began with a G.

Transposase contacts both single- and double-stranded regions of the substrate DNA during strand transfer

Protein–DNA interactions can be investigated using a variety of chemical probing methods (Wissmann and Hillen, 1991; Cao and Revzin, 1993) as well as using gel retardation or DNase I footprinting methods. While these latter methods are useful for studying proteins with a high binding-site specificity, they rarely provide information about critical transient DNA contacts that are made during the assembly of catalytically active nucleoprotein complexes. For the HIV integrase protein, chemical modification of the DNA substrate was successfully used to determine the sites of contact between the integrase protein

and the substrate DNA during strand transfer (Bushman and Craigie, 1992; Chow and Brown, 1994b). For Mu transposase, the DNase I footprinting pattern on single-ended Mu DNA was distinct from that of catalytically active transposase–DNA complexes (Mizuuchi *et al.*, 1991). DNase I footprinting analysis was used to determine that P element transposase binds to sites within the P element sequence that are either 40 or 52 bp away from the sites of cleavage (Kaufman *et al.*, 1989). We reasoned that additional contacts must be made between the transposase protein and the P element termini that are not detected by DNase I footprinting methods since transposase presumably contacts the terminal nucleotides during the DNA cleavage and strand-transfer reactions.

In order to investigate the base-specific sites of interaction between the transposase protein and the substrate DNA during strand transfer, we used a 'missing base' method (Brunelle and Schleif, 1987) in which either formic acid was used to depurinate or hydrazine was used to depyrimidinate the strand-transfer substrates prior to incubation with transposase. The sites of DNA modification were identified following piperidine cleavage. In addition, we investigated the phosphate-specific contacts between the transposase protein and the strand-transfer



substrate DNA by using ethylation interference, in which ethylnitrosourea (ENU) was used to modify nonbridging phosphoryl oxygens on the substrate DNA prior to incubation with transposase (Siebenlist and Gilbert, 1980). The sites of phosphate modification were identified following cleavage of the labile phosphotriester bond with alkali. We prepared the radiolabeled, modified oligonucleotide substrates and performed large scale strand-transfer reactions using the T0.6 protein fraction as the transposase source (Figure 4). Separate reactions were performed to examine contacts made on both the transferred (P1) or non-transferred (P2–17) strand using the authentic, 17 nt 5′ recessed strand-transfer substrate. The transferred products were purified from an agarose gel, chemically cleaved at the modified positions, and subjected to denaturing polyacrylamide gel electrophoresis (Figure 4). By comparing reactions that were performed in the presence or absence of transposase, critical sites of contact were determined by the reduction or absence of a product fragment in the lane that contained transposase during the reaction.

Using the missing base method and a radiolabeled transferred (P1) strand substrate, several positions were identified in which base removal resulted in a reduced recovery of the modified DNA at those positions in the presence of transposase (Figure 5A and B). These positions include: regions within the single-stranded DNA near the strand-transferred terminus (Figures 5A and 7; G52 and T51 which showed <10% activity when bases were removed at these positions), other sites within the 17 nt single-stranded extension and the duplex region containing the 31 bp inverted repeat (Figures 5A and 7) and every position within the transposase binding site (Figures 5B and 7). Our strand-transfer data, demonstrating that either a change in the transferred nucleotide from G to T or removal of the terminal G nucleotide reduced or abolished transposase activity (Figure 3C), is consistent with the critical nature of the terminal nucleotide in strand transfer. In addition, previous mutagenesis experiments determined that the duplex portion of the 31 bp inverted repeat, as well as the transposase binding site and single-stranded region of the 31 bp inverted repeat, are critical for P element transposition to occur *in vivo* (Mullins *et al.*, 1989). Interestingly, increased strand transfer (>200%) was observed with the removal of a base at position T38, T36, T35 or C34 (Figures 5A and 7). Since these positions are opposite the cleavage site, it is possible that base removal at these sites causes a local distortion in the DNA that increases the affinity of transposase for these modified DNAs (see Discussion). When the experiment was repeated using a radiolabeled non-transferred strand (P2–17) substrate, positions within the 31 bp inverted repeat (Figures 5D and 7) and the transposase-binding site (Figures 5C and 7) were identified which inhibited the strand-transfer reaction when the bases were removed at those sites. For both strands, varying degrees of interference were detected at each site which presumably reflect the different levels of transposase inhibition at each position of the modified strand-transfer substrate. These data indicate that in addition to the previously identified sites of contact at the transposase binding site, transposase makes critical base contacts within the duplex region of the 31 bp inverted repeat and the 17 nt single-stranded DNA segment during assembly and/or catalysis of the strand-transfer reaction.

Using radiolabeled, ENU-treated substrate DNA, phosphate groups that are important for contact with transposase during the strand-transfer reaction were identified (Figure 6). Ethylation at several transferred strand (P1) phosphate positions resulted in reduced recovery of the DNA ethylated at those positions in the presence of transposase (Figure 6A and B). These positions include: regions within the single-stranded DNA near the strand-transferred terminus (Figures 6A and 7; for example pG52, pT51 and pA50), several sites within the 17 nt single-stranded extension and the duplex region containing the 31 bp inverted repeat (Figures 6A and 7) and several positions within the transposase binding site (Figures 6B and 7). When the experiment was repeated using the radiolabeled non-transferred strand (P2–17) substrate DNA, only one position within the transposase binding site and four sites within the 31 bp inverted repeat were found to be important for contact with transposase during strand transfer (Figure 6C and D). Interestingly, several positions showed

Fig. 3. Single-stranded DNA at the P element terminus stimulates strand transfer. (**A**) Schematic diagram of the strand-transfer assay. Oligonucleotides derived from the right (3′) P element end that contain the transposase binding site (light gray box) and the 31 bp inverted repeat (dark gray box) were annealed to produce substrates that mimic a cleaved P element terminus. Transposase-containing fractions were incubated with radiolabeled (asterisk) substrate DNA for 15 min on ice. The strand-transfer reaction was initiated by adding GTP, Mg^{2+} and the plasmid DNA target. After 2 h at 30°C the reactions were proteinase K-treated, deproteinized, ethanol-precipitated and analyzed by agarose gel electrophoresis. Strand transfer of the free substrate oligonucleotide to the plasmid DNA target results in both relaxed circular single-ended or linearized double-ended transfer products, as shown. (**B**) Schematic diagram of the different oligonucleotide strand-transfer substrate DNAs. Substrates all contain the transposase-binding site (light gray boxes) and the 31 bp inverted repeat (dark gray boxes) separated by a 9 bp spacer sequence. P1, transferred strand; P2, non-transferred strand. The substrates were either flush (P2 flush), recessed 17 nt on the non-transferred strand (P2–17), recessed 9 nt on the non-transferred strand (P2–9), mutated at the transposase binding site (mutant), recessed 1 nt on the transferred strand (P1–1) or mutated at the nucleotide joined to the target DNA (P1 G to T). The mutated nucleotides are indicated by lower case letters. (**C**) *Drosophila* T0.6 (lanes 1–7) and baculovirus MBP–Tnp (lanes 14 and 15) were tested in the strand-transfer assay as described above. Shown is an autoradiogram from an experiment in which all of the different strand-transfer substrates described in Figure 3B were tested. All reactions contained the cofactors GTP and Mg^{2+} unless otherwise indicated. All substrates were radiolabeled at the 5′ end of the transferred (P1) strand. The positions of the strand-transfer products are indicated by arrows. SET, single-ended transfer; DET, double-ended transfer; No protein, reactions in which transposase was omitted. Values from PhosphorImager quantitation of the products are reported in Table II. For each reaction, 0.5 pmol of radiolabeled substrate DNA, ~10 ng of the T0.6 and ~100 ng of the MBP–Tnp transposase-containing fractions were tested. The flush-ended substrate was used in the reactions lacking GTP (lane 1). Identical results were obtained for all of the different substrates when GTP was omitted from the reaction (data not shown). (**D**) Large scale strand-transfer reactions were performed using the T0.6 transposase fraction and the radiolabeled P2–17 substrate DNA. Reaction products corresponding to both the SET and DET were isolated from an agarose gel. Chemical DNA sequencing analysis of the purified products is shown. The relevant sequence is indicated, with P element 31 bp inverted repeat sequence boxed and numbered from the terminal P element nucleotide. Authentic strand-transfer products display a readily discernible sequencing ladder up to the transferred nucleotide that is followed by a random sequence array representing different target site insertions.

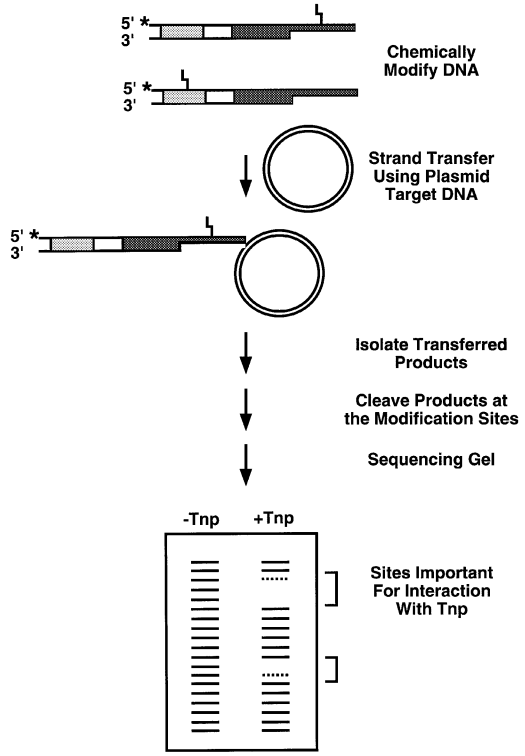


Fig. 4. Schematic diagram of the experiments used to determine the critical contacts made between the transposase protein and the substrate DNA during strand transfer. The radiolabeled P2–17 strand-transfer substrate was treated with ENU to ethylate the nonbridging phosphoryl oxygens, hydrazine to depyrimidinate or formic acid to depurinate the DNA prior to strand transfer. For simplicity, sites of ethylation or base removal are represented by a bent line. In separate experiments, either the transferred strand (P1) was radiolabeled at the 5′ end (asterisk), or the non-transferred strand (P2) was radiolabeled to investigate transposase–DNA interactions with both strands independently. Following chemical treatment, the DNAs were used in large scale strand-transfer assays with the T0.6 transposase fraction. Strand-transfer products (both single- and double-ended transfers) were isolated from agarose gels and the purified DNA treated with either NaOH (ethylated DNA) or piperidine (missing base DNA) to cleave the DNA at the site of chemical modification. The cleaved products were analyzed by denaturing gel electrophoresis and autoradiography. Positions at which the chemical modification interferes with the ability of transposase to use that molecule during strand transfer are revealed as a reduction or a gap in the ladder (+Tnp lane) when the products are compared with reactions in which transposase is omitted (–Tnp lane).

increased strand transfer when ethylated (Figure 6A, C and D). The reason for the enhancement is not understood, but may be due to removal of the negative charge on the phosphate or a change in the DNA conformation at the modified position that increases the affinity of transposase for the modified DNA.

The results from the missing base and ethylation interference experiments are summarized in Figure 7A and as a B-form DNA model of the oligonucleotide substrate in Figure 7B. It is clear from the DNA model that specific phosphate contacts cluster within the transposase-binding site and the duplex portion of the 31 bp inverted repeat, and that the clusters lie on similar faces of the DNA helix (Figure 7B, white spheres). In addition, the majority of base- and phosphate-specific contacts appear to be located within the major groove of the DNA helix at both regions of the substrate DNA. The cluster of specific phosphate contacts within the single-stranded DNA region near the transferred nucleotide may lie on the same face of the DNA helix as the transposase binding site. For simplicity, the single-stranded region of the substrate DNA is shown as a helical form in Figure 7B. However, it is not known whether the single-stranded region remains helical after the donor DNA cleavage reaction. Taken together, these data indicate that the transposase protein must contact both single- and double-stranded regions of the substrate DNA during the strand-transfer reaction, and that the catalytically active form of transposase may be a multimer (see Discussion). Some positions of modification appear to enhance reactivity, presumably due to effects on the structure or charge of the substrate DNA. The contacts that show the most dramatic reduction in transposase activity occur at the terminal nucleotides (Figure 7), in which both base- and phosphate-specific contacts presumably allow for proper substrate recognition, alignment and strand transfer.

Discussion

The results presented here demonstrate that P element transposase can use model oligonucleotide substrates during the strand-transfer step of the transposition reaction. The 17 nt single-stranded extension normally produced by transposase during the cleavage step of the transposition reaction stimulates both the strand transfer and the disintegration reactions *in vitro*. In addition, a series of chemical modification–interference experiments indicate that transposase recognizes the terminal single-stranded region as well as two duplex regions of the substrate DNA during the strand-transfer reaction. Our data support and extend previous *in vivo* studies of the P element transposition reaction. We find that target site selection is similar to target sites selected *in vivo* (O'Hare and Rubin, 1983). In addition, some of the contacts made between the transposase protein and the substrate DNA outside of the consensus transposase binding site map to regions within the terminal 31 bp inverted repeats that had previously been identified as being required for transposition by P element mutagenesis experiments *in vivo* (Mullins *et al.*, 1989). Together, these data provide a better understanding of how P element transposase–DNA interactions occur to carry out the different chemical steps of the P element transposition reaction.

DNA structure is critical for transposase

The use of single transposon end strand-transfer oligonucleotide substrates allowed us to examine the role of the 17 nt single-stranded region during both the strand-transfer and disintegration reactions. The data presented here show that the single-stranded region is important for both reactions. In addition, substrates containing the 17 nt single-stranded region represent a true intermediate in the transposition reaction since flush-ended substrate DNAs were not efficiently used during strand transfer (Figure 3). Our data suggest that following cleavage, the transposon termini do not remain as a duplex prior to strand transfer. Our data also suggest that the P element transposase 'active site' accommodates both duplex DNA during donor DNA cleavage and single-stranded DNA during strand transfer. This flexibility within the active site of the

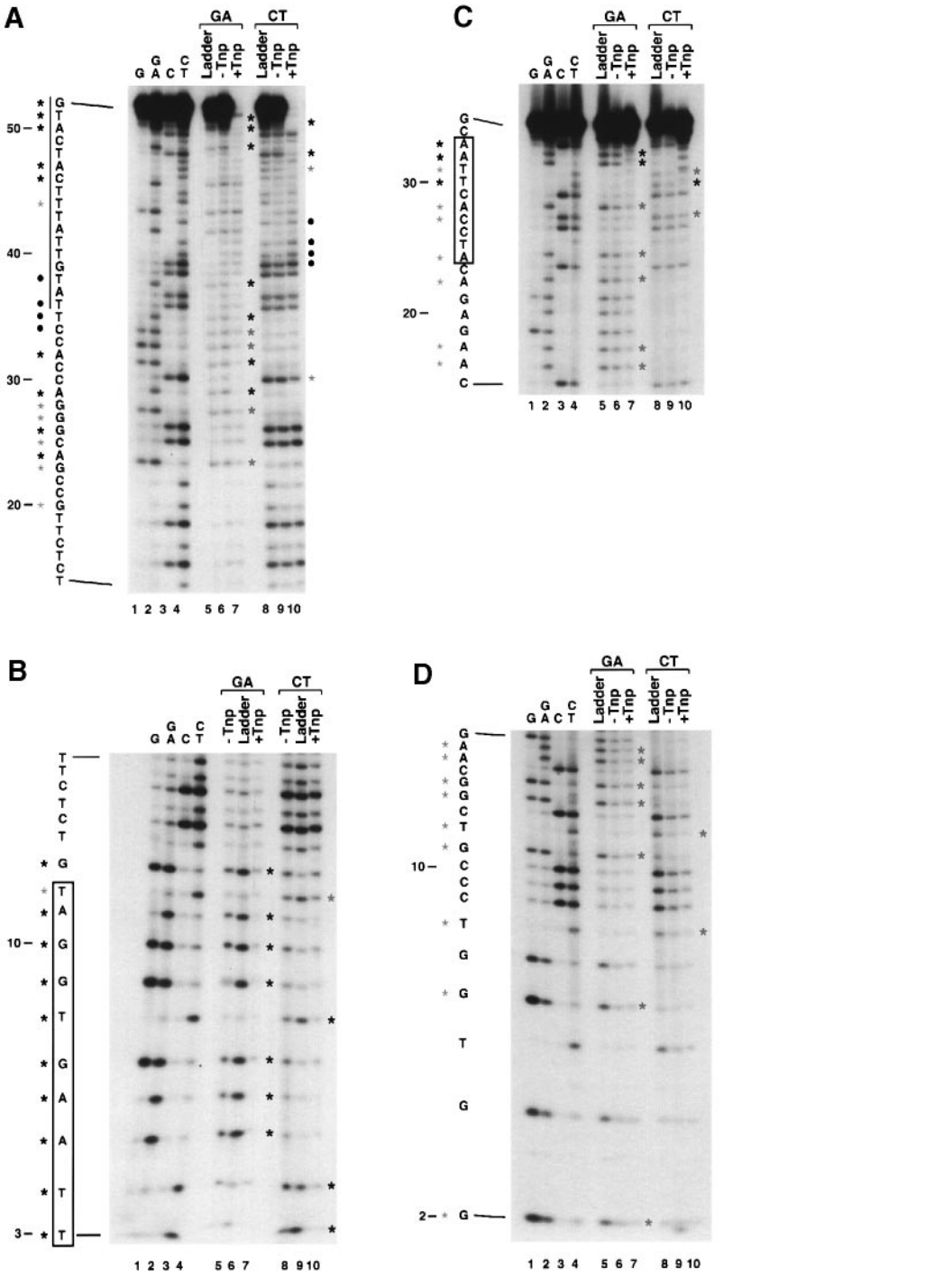


Fig. 5. Transposase makes several base contacts within the single-stranded and duplex regions of the substrate DNA during strand transfer. Shown are the piperidine-cleaved products from the missing base strand-transfer substrate DNA reactions. Maxam–Gilbert chemical sequencing reactions (lanes 1–4 for each panel) are shown as markers. For each strand, the sequence is numbered from the 5′ end of the radiolabeled oligonucleotide. The transposase-binding site sequence is boxed and the 17 nt single-stranded region is indicated by a line. GA, reactions in which the substrate DNA was depurinated prior to strand transfer (lanes 5–7 for each panel); CT, reactions in which the substrate DNA was depyrimidinated prior to strand transfer (lanes 8–10 for each panel). PhosphorImager quantitation was performed on every band for each of the lanes marked –Tnp and +Tnp (data not shown) and these data used to determine the sites of contact between the transposase protein and the substrate DNA. Black asterisk, positions in the +Tnp lane that showed from 0–40% of the –Tnp levels when the base was removed; gray asterisk, positions in the +Tnp lane that showed 40–60% of the –Tnp levels when the base was removed; black dots, positions in the +Tnp lane that showed >200% of the –Tnp levels when the base was removed. (**A** and **B**) Identification of the base contacts made between the transposase protein and the transferred (P1) strand. Shown are autoradiograms from reactions in which the transferred (P1) oligonucleotide was radiolabeled and the products analyzed on 12% (A) or 23% (B) polyacrylamide–urea sequencing gels. +Tnp, piperidine-cleaved, transposase-mediated strand-transfer products; –Tnp, piperidine-cleaved substrate DNA; ladder, lanes in which the relevant DNA sequencing reactions were loaded for comparison (either GA or CT Maxam–Gilbert DNA sequencing reactions). (**C** and **D**) Identification of the base contacts made between the transposase protein and the non-transferred (P2–17) strand. Shown are autoradiograms from reactions in which the non-transferred (P2–17) oligonucleotide was radiolabeled and the products analyzed on 12% (C) or 23% (D) polyacrylamide–urea sequencing gels. +Tnp, piperidine-cleaved, transposase-mediated strand-transfer products; –Tnp, piperidine-cleaved substrate DNA; ladder, lanes in which the relevant sequencing reactions were loaded for comparison (either GA or CT Maxam–Gilbert sequencing reactions).

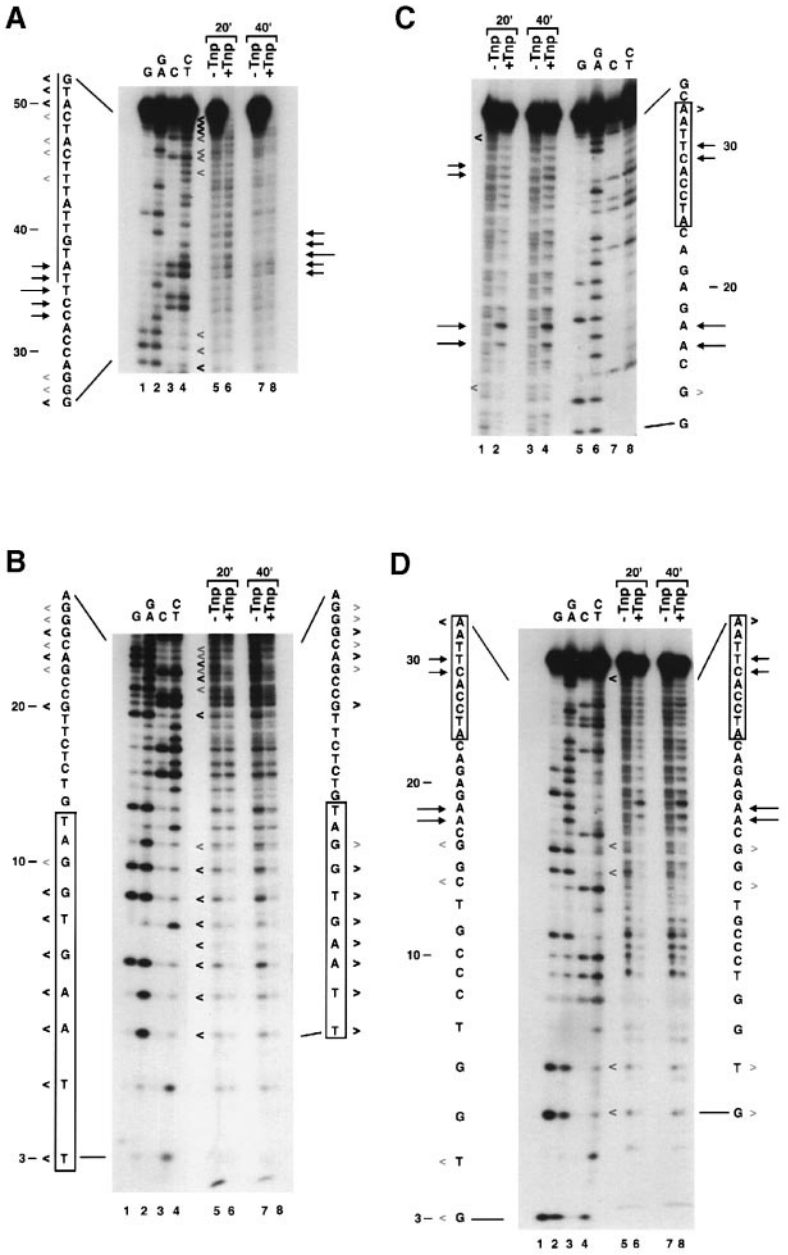


Fig. 6. Transposase makes several phosphate contacts within the single-stranded and duplex regions of the substrate DNA during strand transfer. Shown are the NaOH-cleaved products from reactions that were performed with ethylated strand-transfer substrate DNAs. Maxam–Gilbert chemical DNA sequencing reactions (lanes 1–4 in A, B and D and lanes 5–8 in C) are shown as markers. For each strand, the sequence is numbered from the 5′ end of the radiolabeled oligonucleotide. The transposase binding site sequence is boxed and the 17 nt single-stranded region is indicated by a line. 20′, reactions in which the substrate DNA was treated with ENU for 20 min prior to strand transfer (lanes 5 and 6 in A, B and D and lanes 1 and 2 in C); 40′, reactions in which the substrate DNA was treated with ENU for 40 min prior to strand transfer (lanes 7 and 8 in A, B and D and lanes 1 and 2 in C). PhosphorImager quantitation was performed on every band for each of the lanes marked –Tnp and +Tnp for the 20 min ethylation only (data not shown) and these data were used to determine the sites of contact between the transposase protein and the substrate DNA. Black <, phosphate positions in the +Tnp lane that showed from 0–40% of the –Tnp levels when ethylated; gray <, phosphate positions in the +Tnp lane that showed 40–60% of the –Tnp levels when ethylated; small arrows, phosphate positions in the +Tnp lane that showed 110–130% of the –Tnp levels when ethylated; large arrows, phosphate positions in the +Tnp lane that showed >130% of the –Tnp levels when ethylated. Note that NaOH cleavage at ethylated phosphates generates both hydroxyl- and ethyl-phosphate 3′ termini that are represented as multiple bands that migrate more slowly than the Maxam–Gilbert DNA sequencing ladder markers (Siebenlist and Gilbert, 1980). Therefore, the phosphates were assigned to the base in the sequencing ladder that migrates just below the phosphate band. For the 23% acrylamide gels shown in (B) and (D), phosphate assignments were based on comparisons to synthetic, radiolabeled, small oligonucleotides (data not shown; see Materials and methods) and the corresponding assignments indicated on the right side of each panel. For both panels, anomalous migrations relative to the Maxam–Gilbert sequencing markers were only observed for the hydroxyl-containing product forms <12 nt in length. (**A** and **B**) Identification of the phosphate contacts made between the transposase protein and the transferred (P1) strand. Shown are autoradiograms from reactions in which the transferred (P1) oligonucleotide was radiolabeled and the products analyzed on 12% (A) or 23% (B) polyacrylamide–urea sequencing gels. +Tnp, NaOH-cleaved, transposase-mediated strand-transfer products; –Tnp, NaOH-cleaved substrate DNA. (**C** and **D**) Identification of the phosphate contacts made between the transposase protein and the non-transferred (P2–17) strand. Shown are autoradiograms from reactions in which the non-transferred (P2–17) oligonucleotide was radiolabeled and the products analyzed on 12% (C) or 23% (D) polyacrylamide–urea sequencing gels. +Tnp, NaOH-cleaved, transposase-mediated strand-transfer products; –Tnp, NaOH-cleaved substrate DNA.

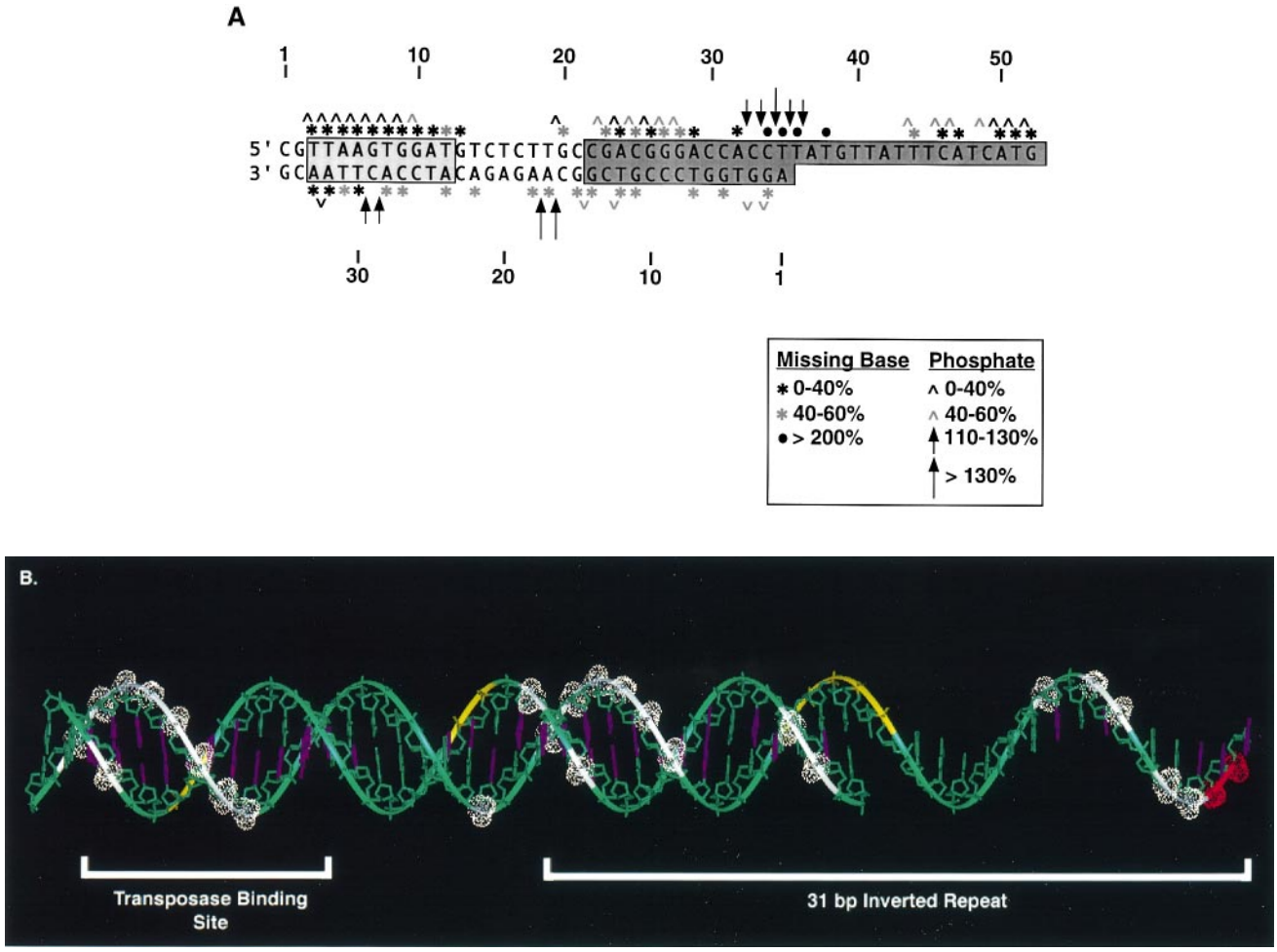


Fig. 7. A model of the contacts made between transposase and the substrate DNA during strand transfer. (**A**) A schematic representation of both the base- and phosphate-specific contacts made between transposase and the strand-transfer substrate DNA. All of the symbols are as described in Figures 5 and 6. The transposase-binding site is boxed in light gray, and the 31 bp inverted repeat is boxed in dark gray. Both the transferred (upper) and non-transferred (lower) strands are numbered from the 5′ ends. (**B**) B-form DNA model of the base- and phosphate-specific contacts made between transposase and the strand-transfer substrate DNA. Both strands of the substrate DNA are shown in green, and the phosphate backbones are drawn as ribbons. For simplicity, the 17 nt single-stranded region is also depicted as B-form DNA. Whether the 17 nt single-stranded region is helical in the substrate DNA is unknown. White spheres, phosphates that show 10–60% strand-transfer activity when ethylated (both black and gray in Figure 7A); red spheres, phosphates that show <10% strand-transfer activity when ethylated; yellow, phosphates that show enhanced reactivity (>110%) when ethylated (small and large arrows in Figure 7A); purple bases, bases that show <60% strand-transfer activity when removed (black and gray asterisks in Figure 7A).

recombinase may be a general feature of the transposase/integrase protein family since several recombinases are able to use substrates that contain duplex, mispaired or single-stranded termini, such as RAG1/RAG2, Mu transposase and the retroviral integrases (Savilahti *et al.*, 1995; Cuomo *et al.*, 1996; Ramsden *et al.*, 1996; Scottoline *et al.*, 1997). However, P element transposase is unique in that the single-stranded nature of the naturally cleaved termini is essential for efficient strand transfer.

Previous studies using bacteriophage Mu transposase (Savilahti *et al.*, 1995), retroviral integrases (Scottoline *et al.*, 1997) and the Rag-1/Rag-2 proteins during V(D)J recombination (Cuomo *et al.*, 1996; Ramsden *et al.*, 1996) suggest that distortion of the DNA duplex near the site of DNA cleavage or strand transfer stimulates the reaction. These experiments imply that the recombinases may facilitate the chemical steps of the reaction by altering the conformation of the substrate DNA during the reaction. Our data using chemically modified DNA substrates during the strand-transfer reaction suggest that P element transposase may also alter the conformation of the DNA during the cleavage reaction. The 31 bp inverted repeat sequence is invariant among all P elements, yet base recognition at the bases opposite the cleavage site TT does not appear to be critical for transposase-catalyzed strand transfer (Figures 5A and 7). In fact, increased strand transfer was observed with substrates that lacked a base or negative phosphate charge at positions opposite the cleavage site TT (Figures 5A and 6A) and may be due to a more accurate representation of the strand-transfer intermediate in these DNAs. Increased strand transfer upon removal of a base in the substrate DNA implies that a structural constraint was mediated by the removed nucleotide. This finding suggests that transposase may be stimulated by, or cause local distortions in, the DNA helix at or near the site of cleavage during transposition. Missing base contact studies of the bovine papillomavirus type 1 (BPV-1) E2 protein revealed increased DNA-binding

activity when bases were removed from a spacer sequence between the E2 half-site binding sites (Li *et al.*, 1989). The increase in activity was ascribed to a bend in the DNA which facilitated binding of monomers within the protein dimer which was later confirmed by X-ray crystallography (Hegde *et al.*, 1992). Furthermore, DNA flexibility has been shown to enhance the interaction of RNA polymerase with sites in an *E.coli* promoter sequence (Werel *et al.*, 1991), as well as enhance cooperative interactions between λ repressor protein dimers and the operator binding sites (Hochschild and Ptashne, 1986). The finding that significant amounts of the strand-transfer products are double-ended (Figure 3C), and that there are two regions of contact made by transposase protein within the strand-transfer substrate that are located on similar faces of the DNA helix (Figure 7), suggests that the active form of transposase is at least a dimer (see below). It is possible that flexibility in, or distortion of, the DNA helix surrounding the cleavage site for transposase may be facilitated by protein–protein interactions following assembly of the transposition complex containing two P element ends.

The unique nature of the cleaved P element termini

After P elements are excised from chromosomal locations *in vivo*, most or all of both the 17 nt extensions are frequently retained at the excision site (Takasu-Ishikawa *et al.*, 1992; Staveley *et al.*, 1995). The extensions may act as a signal that a transposition event has been initiated so that the DNA end-joining repair machinery is targeted to the site of damage. In mammalian cells, several proteins are known to be involved in double-strand DNA break end-joining repair which include the Ku (p70 and p80) and DNA-PKcs (p465) proteins. Cell lines containing mutations in either of the Ku subunits or the DNA-PKcs protein display sensitivity to DNA damaging agents and defects in V(D)J recombination (Taccioli *et al.*, 1993, 1994; Blunt *et al.*, 1995; Boubnov *et al.*, 1995; Kirchgessner *et al.*, 1995; Errami *et al.*, 1996; Gu *et al.*, 1997). Genetic and biochemical data suggest that the analogous *Drosophila* Ku–DNA-PK complex is required for repair of double-strand DNA breaks following P element transposition (Beall and Rio, 1996). In fact, the P element cleavage site occurs directly adjacent to the *Drosophila* Ku p70 (IRBP)-binding site (Rio and Rubin, 1988) such that the single-stranded DNA extensions contain the complete IRBP-binding site. The mammalian Ku–DNA-PK complex recognizes single-double-stranded DNA junctions, gaps and hairpins (Mimori and Hardin, 1986; Paillard and Strauss, 1991; Griffith *et al.*, 1992). It is possible that the single-stranded P element sequences left at the donor site after transposase-mediated cleavage may already contain, or are subsequently bound by, the *Drosophila* Ku–DNA-PK complex. Subsequently, phosphorylation of downstream targets by the activated Ku–DNA-PK and/or induction of the Ku ATPase and helicase activities (Cao *et al.*, 1994; Tuteja *et al.*, 1994) may trigger the double-stranded DNA break repair mechanisms. Targeting the Ku–DNA-PK complex to the sites of P element transposase-cleaved DNA would allow efficient donor DNA repair following P element excision *in vivo*.

While other cut-and-paste transposons, such as Tn7, Tn10 and Tc1/3, cleave the non-transferred strand within a few nucleotides of the transposon terminus (Bainton *et al.*, 1991; Benjamin and Kleckner, 1992; van Luenen *et al.*, 1994; Gary *et al.*, 1996; Vos *et al.*, 1996), P elements contain extensive single-stranded DNA at the termini following cleavage. The data presented here suggest that the single-stranded nature of the P element termini is critical for efficient strand transfer or integration of the excised transposon. Since the *Drosophila* Ku p70 DNA-binding site will also be retained on the cleaved element termini, Ku–DNA-PK bound at the element ends may facilitate the integration into, and DNA repair at, the new target site. Development of *in vitro* cleavage (Beall and Rio, 1997) and strand-transfer assays should allow direct tests of the effects of the Ku–DNA-PK complex on P element transposase activity.

The active form of the transposase protein is at least a dimer

Previous DNase I protection experiments did not indicate that the P element transposase protein interacts with the terminal 31 bp inverted repeats (Kaufman *et al.*, 1989). These experiments did not allow for the examination of transposase–DNA contacts made throughout the course of the transposition reaction. The chemical modification–interference experiments presented here indicate that the transposase protein makes critical contacts with the single-stranded and two distinct double-stranded regions of the substrate oligonucleotides during strand transfer. Similarly for the Mu transposase and retroviral integrase proteins, important protein–DNA interactions were found to occur at the element termini when functional activity assays rather than simple DNA-binding experiments were used to assess such interactions (Mizuuchi *et al.*, 1991; Bushman and Craigie, 1992; Chow and Brown, 1994b). Consistent with this notion, we found that UV cross-linking experiments failed to reveal differential binding of transposase to modified DNAs carrying 5-iodouracil substitutions within or outside the regions of protein contact identified by the modification–interference studies (E.L.Beall and D.C.Rio, unpublished). It is possible that specific transposase–DNA interactions within the 31 bp inverted repeats will be revealed only under functional assay conditions.

Our results indicate that multiple complex DNA–protein interactions must take place to catalyze the transesterification reaction of strand transfer. Studies with the retroviral integrase proteins reveal that single-ended integration predominantly occurs *in vitro*, yet the protein itself oligomerizes (Engelman *et al.*, 1993; van Gent *et al.*, 1993; Ellison and Brown, 1994; Ellison *et al.*, 1995; Zheng *et al.*, 1996). Studies of the Mu transposase protein using oligonucleotide substrates reveal that the strand-transfer reaction is most efficient when precleaved DNA derived from the right Mu DNA end is used in the presence of DMSO (Craigie and Mizuuchi, 1987; Mizuuchi and Mizuuchi, 1989; Namgoong *et al.*, 1994). Under these conditions, the two ends are stably bound by a tetramer of the Mu transposase protein (Mizuuchi and Adzuma, 1991; Mizuuchi *et al.*, 1991) and nearly equal amounts of both single-ended and double-ended strand-transfer products are observed (Aldaz *et al.*, 1996; Yang *et al.*, 1996). Unlike Mu transposase and the retroviral integrases,

we found that the P element transposase protein generates significant amounts of the two-ended product during strand transfer when the oligonucleotide substrate contains the authentic cleaved terminus. However, when DMSO and zwitterionic detergents were included in the reaction we found an increased ratio of single-ended:double-ended strand-transfer products (E.L.Beall and D.C.Rio, unpublished). It appears that under modified reaction conditions the requirements for nucleoprotein complex assembly leading to concerted strand transfer are relaxed. The observation that a majority of the strand-transfer products are double-ended (to generate a linear product; Figure 3A) suggests that double-ended strand transfer is probably promoted by oligomerization of the transposase protein. Oligomerization may also be critical to juxtapose the active sites of individual transposase monomers to generate the 17 nt staggered double-stranded DNA breaks during the P element cleavage reaction (Beall and Rio, 1997).

The finding that there are two clusters of critical phosphate contacts within the duplex portion of the strand-transfer substrate that lie on similar faces of the DNA helix (Figure 7B) is also consistent with an oligomer (possibly a dimer) of transposase catalyzing the strand-transfer reaction. The transposase site-specific DNA-binding domain is located within the N-terminal 88 amino acids of the protein (Lee *et al.*, 1996; C.Lee, E.Beall and D.Rio, submitted), whereas the catalytic domain lies within the C-terminus (Y.Mul, B.Wang and D.C.Rio, unpublished). Studies with bacterially expressed N-terminal transposase protein fragments indicate that the DNA-binding domain of the transposase protein has some affinity for the 31 bp inverted repeats (Lee *et al.*, 1996; C.Lee, E.Beall and D.Rio, submitted). By having the DNA-binding domain separated from the catalytic domain, a high degree of flexibility within the transposase protein can be achieved such that one monomer (or dimer) of transposase bound to each transposase-binding site at the P element termini can readily interact with the cleavage and/or strand-transfer site either 40 or 52 bp away. However, whether transposase is acting as an oligomer during P element transposition has yet to be determined. The ability of P element transposase to catalyze double-ended strand transfer should provide a way to study both the active oligomeric state of the transposase protein and the interactions of the transposase protein with the target site DNA during the reaction. The ability of transposase to use simple oligonucleotide substrates during strand transfer *in vitro* should facilitate future studies examining the biochemistry of P element transposition.

Materials and methods

Preparation of oligonucleotide substrates

All oligonucleotides were synthesized on an ABI model 392 DNA synthesizer. After deprotection, the oligonucleotides were purified by denaturing urea-polyacrylamide gel electrophoresis, eluted, concentrated with n-butanol and precipitated with ethanol.

Preparation of recombinant transposase proteins

A recombinant baculovirus encoding the transposase protein was created from the transposase cDNA (Mul and Rio, 1997) and the MBP coding sequence obtained by PCR from the plasmid pMAL-c2 (New England Biolabs). The MBP–transposase fragment was inserted into the baculovirus transfer vector, pVL1393 (Pharmingen), and transfected into Sf9 cells along with baculogold viral DNA (Pharmingen) by liposome-mediated transfection using standard methods. After plaque purification and amplification, 10–15 cm dishes of Sf9 cells were infected at an m.o.i. of 10. Sixty hours after infection, the cells were harvested and used to prepare whole-cell sonicated extracts (Chen *et al.*, 1994). The MBP–transposase fusion protein was purified using amylose agarose chromatography as described (Ausubel *et al.*, 1987). The peak fraction contained ~100 ng/μl transposase protein as judged by Coomassie Blue-stained SDS–polyacrylamide gels containing known amounts of bovine serum albumin.

P element transposase was purified from the *Drosophila* Schneider L2 stable cell line pUChygMT-Tnp, following $CuSO_4$ induction as described (Mul and Rio, 1997). Briefly, nuclear extracts were precipitated with ammonium sulfate and fractions containing transposase pooled, dialyzed and chromatographed on heparin–agarose (Kaufman *et al.*, 1989). All buffers contained 50 mM NaF to inhibit phosphatase activity. The flow-through (H0.1FT) contained highly active transposase, as determined by a genetic-based plasmid assay (Kaufman and Rio, 1992). The H0.1FT transposase-containing fraction was chromatographed on a nonspecific DNA affinity resin (TdT) essentially as described (Kaufman *et al.*, 1989; Beall and Rio, 1997). The transposase protein was eluted with increasing KCl and the 0.6 M KCl peak fraction (T0.6) was used in the strand-transfer and disintegration assays. Both fractions contained ~10 ng/μl transposase protein as judged by silver-stained SDS–polyacrylamide gels containing known amounts of bovine serum albumin.

In vitro disintegration assay

The DNA oligonucleotides used for the disintegration assays are as follows:

T1: 5′-GTCTGACAGTCCAAGG

P1/T2L: 5′-CACTGAATTTAAGTGTATACTTCGGTAAGCTTCGGC-TTTCGACGGGACCACCTTATGTTATTTCATCATGGCCAGCGGGT-CTCG

P2: 5′-CATGATGAAATAACATAAGGTGGTCCCGTCGAAAGCCG-AAGCTTACCGAAGTATACACTTAAATTCAGTG

T3: 5′-CGAGACCCGCTGGCCCTTGGACTGTCAGAC

P2–17: 5′-AGGTGGTCCCGTCGAAAGCCGAAGCTTACCGAAGT-ATACACTTAAATTCAGTG.

The oligonucleotides were gel-purified and 50 pmol of T1 was radiolabeled at the 5′ end using T4 polynucleotide kinase and [γ-^{32}P]ATP (6000 Ci/mmol; ICN). 250 pmol each of P1/T2-L, P2 and T3 were annealed to the radiolabeled T1 oligonucleotide and the branched oligomer disintegration substrate gel-purified following electrophoresis on a 10% native polyacrylamide gel as described (Chow *et al.*, 1992).

Reaction conditions for the disintegration assay were as follows: ~100 fmol of radiolabeled disintegration substrate was incubated with the transposase-containing fractions in a volume of 15 μl in chromatography buffer (HGKED: 20 mM HEPES–KOH, pH 7.6, 20% glycerol, 100 mM KCl, 0.5 mM EGTA, 0.5 mM EDTA, 1 mM DTT, 0.2 mM PMSF, with the addition of 100 μg/ml bovine serum albumin) containing 10 mM $MnCl_2$ and 2 mM GTP such that the final [KCl]⩽35 mM. Reactions were performed at 30°C for 2 h and terminated by the addition of 15 μl deionized formamide/30 mM EDTA/loading dyes. For the reactions using the T0.6 transposase-containing fraction, poly-dIdC was included to compete with nuclease activity present in this fraction. One fifth of the total reaction products were analyzed on 15% polyacrylamide–urea sequencing gels. Products were quantitated by PhosphorImage analysis (Fuji) and corrected for loading variations by comparison of the total radioactivity loaded in each lane. The corrected values were used to calculate the percentage of disintegration activity for each substrate and transposase protein fraction and the ratio of the P2–17:Flush (P2) activity was used to calculate the fold-activation. Table I shows the results from data presented in Figure 2B.

Table I. Corrected values from PhosphorImager quantitation used in disintegration assay

	Corrected values for		
	Flush (P2)	P2–17	Fold-activation
MBP–transposase	26.50	152.21	6
T0.6	18.84	140.31	7

Table II. Values from PhosphorImager quantitation of products in the strand-transfer assay

Protein sample	Substrate	Corrected values	Difference
T0.6	–GTP	79.52	
	Flush	170.02	
	P2–17	4230.76	25× increase
	P2–9	308.97	14× decrease
	mutant	236.41	18× decrease
	P1–1	26.06	162× decrease
	P1 G to T	45.2	45× decrease

In vitro strand-transfer assay

The DNA oligonucleotides used for the strand-transfer assays are as follows:

P1: 5′-CGTTAAGTGGATGTCTCTTGCCGACGGGACCACCTTATG-TTATTTCATCATG

P1-mutant: 5′-CGCAGCACTACGGTCTCTTGCCGACGGGACCAC-CTTATGTTATTTCATCATG

P1–1: 5′-CGTTAAGTGGATGTCTCTTGCCGACGGGACCACCTTA-TGTTATTTCATCAT

P1 G to T: 5′-CGTTAAGTGGATGTCTCTTGCCGACGGGACCACC-TTATGTTATTTCATCATT

P2: 5′-CATGATGAAATAACATAAGGTGGTCCCGTCGGCAAGAG-ACATCCACTTAACG

P2–17: 5′-AGGTGGTCCCGTCGGCAAGAGACATCCACTTAACG

P2–9: 5′-ATAACATAAGGTGGTCCCGTCGGCAAGAGACATCCAC-TTAACG

P2-mutant: 5′-AGGTGGTCCCGTCGGCAAGAGACCGTAGTGCT-GCG.

The oligonucleotides were gel-purified and 11.6 pmol of P1 (or variants of P1 described above), radiolabeled at the 5′ end using T4 polynucleotide kinase and [γ-^{32}P]ATP (6000 Ci/mmol, ICN) 11.6 pmol of P2 (or variants of P2 described above), was annealed to the radiolabeled P1 oligonucleotide, and the duplex DNA strand-transfer substrate removed from the unincorporated [γ-^{32}P]ATP using Probe-Quant G-50 microcolumns (Pharmacia).

Reaction conditions for the strand-transfer assay were as follows: 0.5 pmol of radiolabeled strand-transfer substrate was incubated with ~100 ng MBP–Tnp or ~10 ng T0.6 transposase-containing fractions in a volume of 6 µl in chromatography buffer (HGKED: 20 mM HEPES–KOH, pH 7.6, 20% glycerol, 100 mM KCl, 0.5 mM EGTA, 0.5 mM EDTA, 1 mM DTT, 0.2 mM PMSF, with the addition of 100 µg/ml bovine serum albumin). Binding was carried out on ice for 15 min. The reaction was initiated by addition of: 0.35×HGKED (0 M KCl), 5 mM $Mg(OAc)_2$, 2 mM GTP and 100 ng Bluescript tetramer target DNA to make a total volume of 20 µl, and [KCl]≤35 mM. Reactions were performed at 30°C for 2 h, terminated by the addition of 125 µl of 50 mM Tris–HCl, pH 7.5, 10 mM EDTA, 0.3 M NaCl, 1% SDS, 250 µg/ml yeast RNA and incubated at 37°C for 30 min with 0.1 mg/ml proteinase K. The reactions were extracted with 25:24:1 phenol:chloroform:isoamyl alcohol and ethanol-precipitated. The pellets were resuspended in 10 µl TE containing 100 µg/ml RNase A and analyzed by agarose gel electrophoresis. The gels were dried onto Whatman DE-81 paper to ensure that the unreacted oligonucleotide substrate DNA was retained and the dried gels were analyzed by autoradiography. Products were quantitated by PhosphorImage analysis (Fuji) and corrected for loading variations by comparison of the total radioactivity loaded in each lane. The corrected values were derived from the total amount of radioactivity for both the single- and double-ended strand-transfer products. The ratio of the P2–17:Flush activity was used to calculate the fold-increase in transposase activity for the P2–17 substrate. The ratio of the P2–17:mutated substrate activity was used to calculate the fold-decrease in transposase activity for each mutant substrate. Table II shows the results from the data presented in Figure 3C.

For the experiment described in Figure 3D, large scale (5× increased) strand-transfer reactions were performed using the T0.6 transposase-containing fraction and radiolabeled P1 oligonucleotide annealed to unlabeled P2–17. The combined single- and double-ended strand-transfer products were excised from agarose gel slices following electrophoresis and autoradiography. The DNA was purified using Qiagen Gel Extraction Kit and the eluted DNA was dried. The purified DNA was subjected to Maxam–Gilbert chemical sequencing (Ausubel *et al.*, 1987) and the products analyzed on a 10% polyacrylamide–urea sequencing gel.

Chemical modification–interference assays

The P2–17 strand-transfer substrate was prepared as described above following radiolabeling of either the P1 or the P2–17 oligonucleotide. Annealed substrate DNAs were treated with ENU for either 20 or 40 min for the ethylation interference experiments (Wissmann and Hillen, 1991). Annealed substrate DNAs were treated with hydrazine (to depyrimidinate) or formic acid (to depurinate) for the missing base experiments (Cao and Revzin, 1993). Following ethanol-precipitation, the substrate DNAs were resuspended in TE/0.1 M NaCl and the strands reannealed before use. For the sequencing ladder markers, radiolabeled oligonucleotides (either P1 or P2–17) were subjected to Maxam–Gilbert sequencing reactions using standard protocols (Ausubel *et al.*, 1987).

Large scale (10× increased) strand-transfer reactions were performed with the chemically modified strand-transfer substrate DNAs. The combined single- and double-ended strand-transfer products were excised from an agarose gel following electrophoresis and autoradiography. The DNA was purified using GeneClean (Bio-101 kit) and the eluted DNA ethanol precipitated with 10 µg glycogen as a carrier.

For the ENU-treated samples, the DNA was resuspended in 60 µl 10 mM $NaPO_4$/1 mM EDTA, pH 7.0. 5 µl of 1 M NaOH was added to 30 µl of the DNA sample prior to heating at 90°C for 30 min. 5 µl of 1 M HCl was added and the products ethanol precipitated using 10 µg glycogen as a carrier. The precipitated DNA was washed extensively with 70% ethanol and dried. The pellet was resuspended in 0.1 M NaOH/1 mM EDTA and an equal volume of deionized formamide/loading dyes. The products were analyzed on both 12 and 23% polyacrylamide–urea sequencing gels.

For the products analyzed on 23% denaturing gels, the Maxam–Gilbert sequencing ladder cannot be used to assign positions for the very small DNAs (Bushman and Craigie, 1992). Therefore, size markers were synthesized that correspond to the expected 3′-hydroxyl-containing alkali-cleaved products for the P1-transferred strand small DNAs:

5′ pCpG
5′ pCpGpT
5′ pCpGpTpT
5′ pCpGpTpTpApA
5′ pCpGpTpTpApApGpT
5′ pCpGpTpTpApApGpTpGpG

and for the P2 nontransferred strand small DNAs:

5′ pApG
5′ ApGpG
5′ pApGpGpTpGpG.

The oligonucleotides were radiolabeled at the 5′ end using T4 polynucleotide kinase and [γ-^{32}P]ATP (6000 Ci/mmol; ICN). [α-^{32}P]dATP and [α-^{32}P]dCTP (3000 Ci/mmol; Amersham) were treated with snake venom phosphodiesterase (Pharmacia) to produce pA and pC markers. Alkali-cleaved, ENU-treated DNAs were analyzed alongside radiolabeled marker DNAs on 23% polyacrylamide–urea sequencing gels and phosphate assignments based on alignment with the standard DNAs (data not shown). The ethylphosphate-containing alkali-cleaved products were not identified for the small DNAs because of the absence of a suitable marker.

For the missing base samples, the DNA was resuspended in 50 µl 1 M piperidine and heated at 90°C for 30 min. 300 µl of dH_2O was added and the samples lyophilyzed. The lyophilyzation was repeated twice and the pellets resuspended in 0.1 M NaOH/1 mM EDTA and an equal volume of deionized formamide/loading dyes. The products were analyzed on both 12 and 23% polyacrylamide–urea sequencing gels.

Every band in Figures 5 and 6 for the + and – transposase reactions was quantitated by PhosphorImager analysis (Fuji) and corrected for loading variations by comparing the total radioactivity loaded in each lane. Comparison of the corrected values for each band between the +Tnp and –Tnp lanes was used to determine whether transposase could use the DNA modified at each position during strand transfer. A quantitation value in the +Tnp lane that was <60% of the –Tnp value was considered important for contact by transposase. Weaker effects that were not interpreted may still be of importance for interaction between transposase and the substrate DNA.

The computer model of the B-form DNA transposase-substrate contact sites (Figure 7B) was generated using Bio-Sim Insight II, version 95.0.4 software.

Acknowledgements

We thank S.Robertson for help in preparing the computer-generated model of the transposase-DNA contact sites. We thank M.Adams, K.Ahrens, M.Botchan, K.Collins, A.Copeland, R.MacDiarmid, P.Kaufman, C.Lee, S.Roche and J.Sekelsky for many helpful discussions and suggestions on the manuscript. We also thank an anonymous reviewer for suggestions on the ethylation interference experiments. This work was supported by a grant from the National Institutes of Health (R01GM48862).

References

Aldaz,H., Schuster,E. and Baker,T.A. (1996) The interwoven architecture of the Mu transposase couples DNA synapsis to catalysis. *Cell*, **85**, 257–269.

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1987) *Current Protocols in Molecular Biology*. Wiley and Sons, New York, NY.

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.

Baker,T.A. and Luo,L. (1994) Identification of residues in the Mu transposase essential for catalysis. *Proc. Natl Acad. Sci. USA*, **91**, 6654–6658.

Beall,E.L. and Rio,D.C. (1996) *Drosophila* IRBP/Ku p70 corresponds to the mutagen-sensitive *mus309* gene and is involved in P element excision *in vivo*. *Genes Dev.*, **10**, 921–933.

Beall,E.L. and Rio,D.C. (1997) *Drosophila* P element transposase is a novel site-specific endonuclease. *Genes Dev.*, **11**, 2137–2151.

Beall,E.L., Admon,A. and Rio,D.C. (1994) A *Drosophila* protein homologous to the human p70 Ku autoimmune antigen interacts with the P transposable element inverted repeats. *Proc. Natl Acad. Sci. USA*, **91**, 12681–12685.

Benjamin,H.W. and Kleckner,N. (1992) Excision of Tn10 from the donor site during transposition occurs by flush double-strand cleavages at the transposon termini. *Proc. Natl Acad. Sci. USA*, **89**, 4648–4652.

Blunt,T. *et al.* (1995) Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA repair defects associated with the murine scid mutation. *Cell*, **80**, 813–823.

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.

Boubnov,N.V. *et al.* (1995) Complementation of the ionizing radiation sensitivity, DNA end binding, and V(D)J recombination defects of double-strand break repair mutants by the p86 Ku autoantigen. *Proc. Natl Acad. Sci. USA*, **92**, 890–894.

Brunelle,A. and Schleif,R.F. (1987) Missing contact probing of DNA–protein interactions. *Proc. Natl Acad. Sci. USA*, **84**, 6673–6676.

Bujacz,G., Jaskolski,M., Alexandratos,J., Wlodawer,A., Merkel,G., Katz,R.A. and Skalka,A.M. (1995) High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J. Mol. Biol.*, **253**, 333–346.

Bujacz,G., Jaskolski,M., Alexandratos,J., Wlodawer,A., Merkel,G., Katz,R.A. and Skalka,A.M. (1996) The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations. *Structure*, **4**, 89–96.

Bushman,F.D. and Craigie,R. (1992) Integration of human immunodeficiency virus DNA: adduct interference analysis of required DNA sites. *Proc. Natl Acad. Sci. USA*, **89**, 3458–3462.

Cao,J. and Revzin,A. (1993) In Revzin,A. (ed.), *Footprinting of Nucleic Acid–Protein Complexes*. Academic Press, San Diego, USA, pp. 173–188.

Cao,Q.P., Pitt,S., Leszyk,J. and Baril,E.F. (1994) DNA-dependent ATPase from HeLa cells is related to human Ku autoantigen. *Biochemistry*, **33**, 8548–8557.

Cech,T.R. (1987) The chemistry of self-splicing RNA and RNA enzymes. *Science*, **236**, 1532–1539.

Chen,J.L., Attardi,L.D., Verrijzer,C.P., Yokomori,K. and Tjian,R. (1994) Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell*, **79**, 93–105.

Chow,S.A. and Brown,P.O. (1994a) Juxtaposition of two viral DNA ends in a bimolecular disintegration reaction mediated by multimers of human immunodeficiency virus type 1 or murine leukemia virus integrase. *J. Virol.*, **68**, 7869–7878.

Chow,S.A. and Brown,P.O. (1994b) Substrate features important for recognition and catalysis by human immunodeficiency virus type 1 integrase identified by using novel DNA substrates. *J. Virol.*, **68**, 3896–3907.

Chow,S.A., Vincent,K.A., Ellison,V. and Brown,P.O. (1992) Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science*, **255**, 723–726.

Craig,N.L. (1995) Unity in transposition reactions. *Science*, **270**, 253–254.

Craig,N.L. (1997) Target site selection in transposition. *Annu. Rev. Biochem.*, **66**, 437–474.

Craigie,R. and Mizuuchi,K. (1987) Transposition of Mu DNA: joining of Mu to target DNA can be uncoupled from cleavage at the ends of Mu. *Cell*, **51**, 493–501.

Cuomo,C.A., Mundy,C.L. and Oettinger,M.A. (1996) DNA sequence and structure requirements for cleavage of V(D)J recombination signal sequences. *Mol. Cell. Biol.*, **16**, 5683–5690.

Dyda,F., Hickman,A.B., Jenkins,T.M., Engelman,A., Craigie,R. and Davies,D.R. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science*, **266**, 1981–1986.

Ellison,V. and Brown,P.O. (1994) A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration *in vitro*. *Proc. Natl Acad. Sci. USA*, **91**, 7316–7320.

Ellison,V., Gerton,J., Vincent,K.A. and Brown,P.O. (1995) An essential interaction between distinct domains of HIV-1 integrase mediates assembly of the active multimer. *J. Biol. Chem.*, **270**, 3320–3326.

Engelman,A., Bushman,F.D. and Craigie,R. (1993) Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *EMBO J.*, **12**, 3269–3275.

Engels,W.R., Johnson-Schlitz,D.M., Eggleston,W.B. and Sved,J. (1990) High-frequency P element loss in *Drosophila* is homolog dependent. *Cell*, **62**, 515–525.

Errami,A., Smider,V., Rathmell,W.K., He,D.M., Hendrickson,E.A., Zdzienicka,M.Z. and Chu,G. (1996) Ku86 defines the genetic defect and restores X-ray resistance and V(D)J recombination to complementation group 5 hamster cell mutants. *Mol. Cell. Biol.*, **16**, 1519–1526.

Gary,P.A., Biery,M.C., Bainton,R.J. and Craig,N.L. (1996) Multiple DNA processing reactions underlie Tn7 transposition. *J. Mol. Biol.*, **257**, 301–316.

Griffith,A.J., Blier,P.R., Mimori,T. and Hardin,J.A. (1992) Ku polypeptides synthesized *in vitro* assemble into complexes which recognize ends of double-stranded DNA. *J. Biol. Chem.*, **267**, 331–338.

Gu,Y., Jin,S., Gao,Y., Weaver,D.T. and Alt,F.A. (1997) Ku70-deficient embryonic stem cells have increased ionizing radiosensitivity, defective DNA end-binding activity, and inability to support V(D)J recombination. *Proc. Natl Acad. Sci. USA*, **94**, 8076–8081.

Hegde,R.S., Grossman,S.R., Laimins,L.A. and Sigler,P.B. (1992) Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature*, **359**, 505–512.

Hochschild,A. and Ptashne,M. (1986) Cooperative binding of λ repressors to sites separated by integral turns of the DNA helix. *Cell*, **44**, 681–687.

Jackson,S.P. (1996) The recognition of DNA damage. *Curr. Opin. Genet. Dev.*, **6**, 19–25.

Katz,R.A. and Skalka,A.M. (1994) The retroviral enzymes. *Annu. Rev. Biochem.*, **63**, 133–173.

Kaufman,P.D. and Rio,D.C. (1992) P element transposition *in vitro* proceeds by a cut-and-paste mechanism and uses GTP as a cofactor. *Cell*, **69**, 27–39.

Kaufman,P.D., Doll,R.F. and Rio,D.C. (1989) *Drosophila* P element transposase recognizes internal P element DNA sequences. *Cell*, **59**, 359–371.

Kirchgessner,C.U., Patil,C.K., Evans,J.W., Cuomo,C.A., Fried,L.M., Carter,T., Oettinger,M.A. and Brown,J.M. (1995) DNA-dependent kinase (*p350*) as a candidate gene for the murine SCID defect. *Science*, **267**, 1178–1183.

Kulkosky,J., Jones,K.S., Katz,R.A., Mack,J.P. 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.

Lee,C.C., Mul,Y.M. and Rio,D.C. (1996) The *Drosophila* P-element KP repressor protein dimerizes and interacts with multiple sites on P-element DNA. *Mol. Cell. Biol.*, **16**, 5616–5622.

Li,R., Knight,J., Bream,G., Stenlund,A. and Botchan,M. (1989) Specific recognition nucleotides and their DNA context determine the affinity

of E2 protein for 17 binding sites in the BPV-1 genome. *Genes Dev.*, **3**, 510–526.
Linn,S.M., Lloyd,R.S. and Roberts,R.J. (1993) *Nucleases*. Cold Spring Harbor Laboratory Press, Plainview, USA.
Mimori,T. and Hardin,J.A. (1986) Mechanism of interaction between Ku protein and DNA. *J. Biol. Chem.*, **261**, 10375–10379.
Mizuuchi,K. (1992a) Polynucleotidyl transfer reactions in transpositional DNA recombination. *J. Biol. Chem.*, **267**, 21273–21276.
Mizuuchi,K. (1992b) Transpositional recombination: mechanistic insights from studies of Mu and other elements. *Annu. Rev. Biochem.*, **61**, 1011–1051.
Mizuuchi,K. and Adzuma,K. (1991) Inversion of the phosphate chirality at the target site of Mu DNA strand transfer: evidence for a one-step transesterification mechanism. *Cell*, **66**, 129–140.
Mizuuchi,M. and Mizuuchi,K. (1989) Efficient Mu transposition requires interaction of transposase with a DNA sequence at the Mu operator: implications for regulation. *Cell*, **58**, 399–408.
Mizuuchi,M., Baker,T.A. and Mizuuchi,K. (1991) DNase protection analysis of the stable synaptic complexes involved in Mu transposition. *Proc. Natl Acad. Sci. USA*, **88**, 9031–9035.
Mul,Y.M. and Rio,D.C. (1997) Reprogramming the purine nucleotide cofactor requirement of *Drosophila* P element transposase *in vivo*. *EMBO J.*, **16**, 4441–4447.
Mullins,M.C., Rio,D.C. and Rubin,G.M. (1989) *cis*-acting DNA sequence requirements for P-element transposition. *Genes Dev.*, **3**, 729–738.
Namgoong,S.Y., Jayaram,M., Kim,K. and Harshey,R.M. (1994) DNA–protein cooperativity in the assembly and stabilization of Mu strand transfer complex. Relevance of DNA phasing and ATT site cleavage. *J. Mol. Biol.*, **238**, 514–527.
O'Hare,K. and Rubin,G.M. (1983) Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. *Cell*, **34**, 25–35.
Paillard,S. and Strauss,F. (1991) Analysis of the mechanism of interaction of simian Ku protein with DNA. *Nucleic Acids Res.*, **19**, 5619–5624.
Pyle,A.M. (1993) Ribozymes: a distinct class of metalloenzymes. *Science*, **261**, 709–714.
Ramsden,D.A., McBlane,J.F., van Gent,D.C. and Gellert,M. (1996) Distinct DNA sequence and structure requirements for the two steps of V(D)J recombination signal cleavage. *EMBO J.*, **15**, 3197–3206.
Rice,P. and Mizuuchi,K. (1995) Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell*, **82**, 209–220.
Rice,P., Craigie,R. and Davies,D.R. (1996) Retroviral integrases and their cousins. *Curr. Opin. Struct. Biol.*, **6**, 76–83.
Rio,D.C. and Rubin,G.M. (1988) Identification and purification of a *Drosophila* protein that binds to the terminal 31-base-pair inverted repeats of the P transposable element. *Proc. Natl Acad. Sci. USA*, **85**, 8929–8933.
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.
Savilahti,H., Rice,P.A. and Mizuuchi,K. (1995) The phage Mu transpososome core: DNA requirements for assembly and function. *EMBO J.*, **14**, 4893–4903.
Scottoline,B.P., Chow,S., Ellison,V. and Brown,P.O. (1997) Disruption of the terminal base pairs of retroviral DNA during integration. *Genes Dev.*, **11**, 371–382.
Siebenlist,U. and Gilbert,W. (1980) Contacts between *Escherichia coli* RNA polymerase and an early promoter of phage T7. *Proc. Natl Acad. Sci. USA*, **77**, 122–126.
Staveley,B.E., Heslip,T.R., Hodgetts,R.B. and Bell,J.B. (1995) Protected P-element termini suggest a role for inverted-repeat-binding protein in transposase-induced gap repair in *Drosophila melanogaster*. *Genetics*, **139**, 1321–1329.
Steitz,T.A. and Steitz,J.A. (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl Acad. Sci. USA*, **90**, 6498–6502.
Taccioli,G.E., Rathbun,G., Oltz,E., Stamato,T., Jeggo,P.A. and Alt,F.W. (1993) Impairment of V(D)J recombination in double-strand break repair mutants. *Science*, **260**, 207–210.
Taccioli,G.E. *et al.* (1994) Ku80: product of the *XRCC5* gene and its role in DNA repair and V(D)J recombination. *Science*, **265**, 1442–1445.
Takasu-Ishikawa,E., Yoshihara,M. and Hotta,Y. (1992) Extra sequences found at P element excision sites in *Drosophila melanogaster*. *Mol. Gen. Genet.*, **232**, 17–23.
Tuteja,N. *et al.* (1994) Human DNA helicase II: a novel DNA unwinding enzyme identified as the Ku autoantigen. *EMBO J.*, **13**, 4991–5001.
van Gent,D.C., Vink,C., Groeneger,A.A. and Plasterk,R.H. (1993) Complementation between HIV integrase proteins mutated in different domains. *EMBO J.*, **12**, 3261–3267.
van Luenen,H.G., Colloms,S.D. and Plasterk,R.H. (1994) The mechanism of transposition of Tc3 in *C.elegans*. *Cell*, **79**, 293–301.
Vos,J.C. and Plasterk,R.H. (1994) Tc1 transposase of *Caenorhabditis elegans* is an endonuclease with a bipartite DNA binding domain. *EMBO J.*, **13**, 6125–6132.
Vos,J.C., De Baere,I. and Plasterk,R.H. (1996) Transposase is the only nematode protein required for *in vitro* transposition of Tc1. *Genes Dev.*, **10**, 755–761.
Weaver,D.T. (1995) What to do at an end: DNA double-strand-break repair. *Trends Genet.*, **11**, 388–392.
Werel,W., Schickor,P. and Heumann,H. (1991) Flexibility of the DNA enhances promoter affinity of *Escherichia coli* RNA polymerase. *EMBO J.*, **10**, 2589–2594.
Wissmann,A. and Hillen,W. (1991) DNA contacts probed by modification protection and interference studies. *Methods Enzymol.*, **208**, 365–379.
Yang,J.Y., Jayaram,M. and Harshey,R.M. (1996) Positional information within the Mu transposase tetramer: catalytic contributions of individual monomers. *Cell*, **85**, 447–455.
Zheng,R., Jenkins,T.M. and Craigie,R. (1996) Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc. Natl Acad. Sci. USA*, **93**, 13659–13664.

Received November 10, 1997; revised and accepted February 4, 1998