

Structural Basis of *hAT* Transposon End Recognition by Hermes, an Octameric DNA Transposase from *Musca domestica*

Alison B. Hickman,¹ Hosam E. Ewis,² Xianghong Li,² Joshua A. Knapp,^{3,9} Thomas Laver,^{4,10} Anna-Louise Doss,⁵ Gökhan Tolun,⁷ Alasdair C. Steven,⁷ Alexander Grishaev,⁸ Ad Bax,⁸ Peter W. Atkinson,^{3,4,5,6} Nancy L. Craig,² and Fred Dyda^{1,*}

¹Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

²Howard Hughes Medical Institute, Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

³Graduate Program in Biochemistry and Molecular Biology

⁴Graduate Program in Genetics, Genomics, and Bioinformatics

⁵Graduate Program in Cell, Molecular, and Developmental Biology

⁶Department of Entomology and Institute for Integrative Genome Biology

University of California Riverside, Riverside, CA 92521, USA

⁷Laboratory of Structural Biology Research, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD 20892, USA

⁸Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

⁹Present address: Navy Drug Screening Laboratory, Naval Air Station, Jacksonville, FL 32212, USA

¹⁰Present address: Advanced Cell Diagnostics, Hayward, CA 94545, USA

*Correspondence: fred.dyda@nih.gov

<http://dx.doi.org/10.1016/j.cell.2014.05.037>

SUMMARY

Hermes is a member of the *hAT* transposon superfamily that has active representatives, including McClintock's archetypal *Ac* mobile genetic element, in many eukaryotic species. The crystal structure of the Hermes transposase-DNA complex reveals that Hermes forms an octameric ring organized as a tetramer of dimers. Although isolated dimers are active *in vitro* for all the chemical steps of transposition, only octamers are active *in vivo*. The octamer can provide not only multiple specific DNA-binding domains to recognize repeated subterminal sequences within the transposon ends, which are important for activity, but also multiple nonspecific DNA binding surfaces for target capture. The unusual assembly explains the basis of bipartite DNA recognition at *hAT* transposon ends, provides a rationale for transposon end asymmetry, and suggests how the avidity provided by multiple sites of interaction could allow a transposase to locate its transposon ends amidst a sea of chromosomal DNA.

INTRODUCTION

Transposable elements and their inactive remnants populate the genomes of almost all organisms that have been examined

(Biémont, 2010), and genes encoding the associated transposase proteins are the most abundant genes in nature (Aziz et al., 2010). Among eukaryotes, the portion of the genome arising from transposable elements ranges from <1% for the honeybee (Honeybee Genome Sequencing Consortium, 2006) to >85% in maize (Schnable et al., 2009). Although they are often silenced, the existence of active transposons, particularly among plants and insects, suggests that their continued mutagenic potential may be beneficial to their hosts (Huang et al., 2012). Furthermore, domesticated transposases can be the source of vital proteins such as the recombination activating protein-1 (RAG1), an essential component of the adaptive immune system that is believed to have originated from a *Transib* DNA transposase (Kapitonov and Jurka, 2005). Thus, understanding how transposon movement and amplification have changed genomes is intimately linked to understanding how genomic organization and remodeling have driven evolution.

Transposable elements are divided into two classes depending on whether they use RNA or DNA intermediates to move. Approximately 20 superfamilies of eukaryotic DNA transposons have been identified and classified based on the amino acid sequences of their transposases (Wicker et al., 2007; Yuan and Wessler, 2011). One of the largest superfamilies is comprised of the *hAT* transposons, named after three of the earliest discovered active transposons: *hobo* from *Drosophila melanogaster* (Streck et al., 1986; Blackman et al., 1989), *Ac* from maize (McClintock, 1950), and *Tam3* from the snapdragon (Hehl et al., 1991; Atkinson et al., 1993).

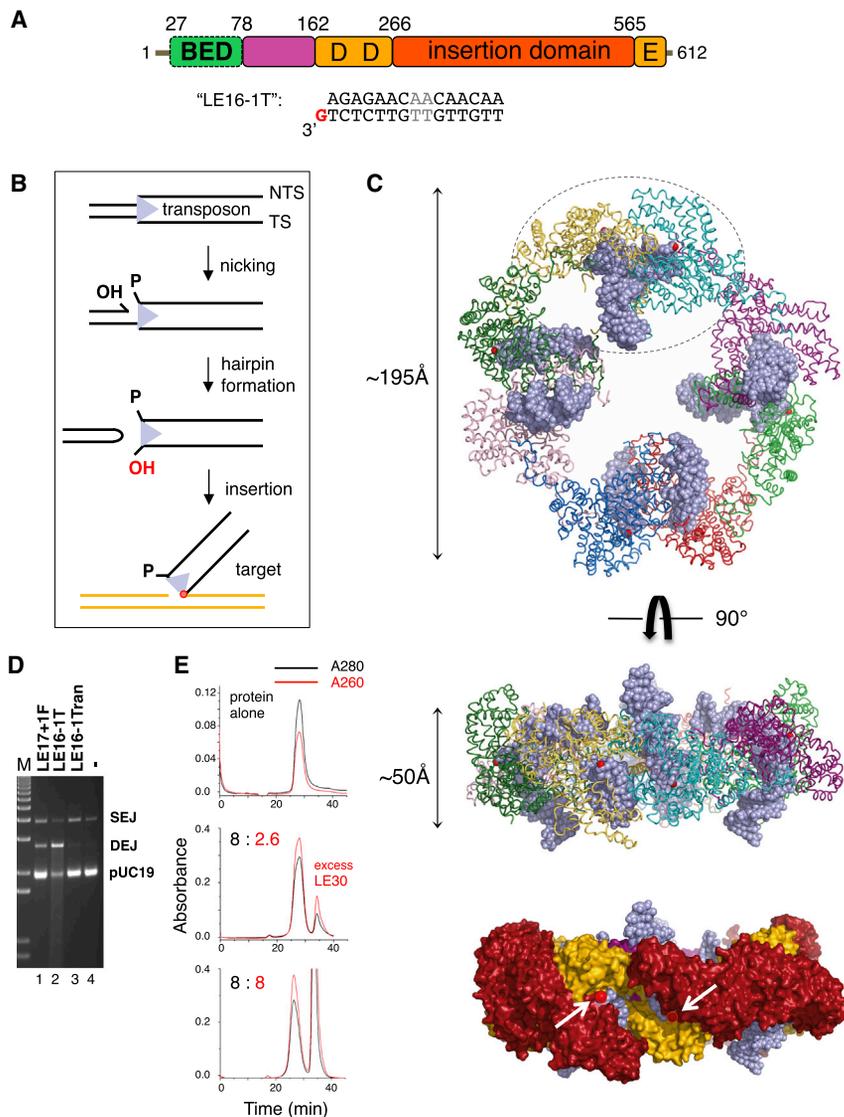


Figure 1. Hermes Overview and Structure

(A) Domain organization and "LE16-1T" DNA used for structure determination. BED domain is in green, intertwined dimerization domain in purple, RNaseH-like catalytic domain in orange, and insertion domain in red. The metal ion binding residues of the DDE motif (D180, D248, E572) are marked. Two AT bp in gray differ between the Left End (LE) and Right End (RE) 17-mer terminal inverted repeats (TIRs).

(B) Reaction scheme for *hAT* transposition. TS, transferred strand; NTS, nontransferred strand.

(C) Structure of Hermes79–612 bound to TIRs. In top and middle, each monomer is a different color, DNA is light blue, and red spheres mark the 3'-OH of each TS. In the bottom surface representation, domains are colored as in (A), and arrows point to the two 3'-OH groups (red spheres) within one dimer.

(D) Strand transfer assay using precleaved LE (28.6 nM) and Hermes 79–612 C519S (10 nM) at 30°C for 2 hr in standard buffer containing 150 mM NaCl. Lane 1: LE17 with one flanking 5'-phosphorylated base. Lane 2: oligonucleotide used for structural studies. Lane 3: randomized oligonucleotide of same length as LE16-1T. Lane 4: target plasmid pUC19 alone. SEJ, single-end joined products; DEJ, double-end joined products. The streak in lane 2 indicates repeated plasmid insertions, causing fragmentation.

(E) Size exclusion chromatography analysis of DNA binding by full-length Hermes. Top: Hermes alone. Middle: Hermes and LE30 mixed in an 8:2.6 ratio. There is some unbound DNA, and the unsymmetric main peak suggests that both the complex and free Hermes exist under these conditions. Bottom: Hermes and LE30 mixed in an 8:8 ratio. Relative to the 8:2.6 ratio, the complex peak is unchanged in size and 280/260 nm ratio, indicating that no more DNA has been bound, although the peak is more symmetrical.

See also [Figures S1, S2](#), and [Table S1](#).

Only a few *hAT* transposons and their transposases have been studied in detail, yet it seems likely that they share common mechanistic and structural features (for review, see [Rubin et al., 2001](#); [Arensburger et al., 2011](#)). Despite very limited sequence similarity, they have a common genetic organization in which a single ORF encoding the transposase is flanked by hundreds of bp of noncoding sequence. These regions 5' and 3' of the transposase gene relative to its direction of transcription are the transposon left end (LE) and right end (RE), respectively, and they terminate in short, ~11–24 bp terminal inverted repeats (TIRs). *hAT* transposases are ~600–800 amino acid multidomain proteins that catalyze transposon excision from one location and insertion into a new one using a cut-and-paste mechanism, integrating their transposons with characteristic 8 bp target site duplications (TSDs). *hAT* transposases do not appear to have mechanistic analogs among prokaryotes ([Hickman et al., 2010](#)), as they excise by generating double-strand breaks accompanied by the formation of DNA hairpins on flanking

DNA, the same mechanism used by RAG1/2 proteins responsible for the generation of vertebrate antigen receptors (for review, see [Schatz and Swanson, 2011](#)).

The only available structure of a *hAT* transposase, a portion of the *Hermes* transposase from the house fly *Musca domestica* ([Warren et al., 1994](#)), revealed an RNaseH-like catalytic domain interrupted by a large α -helical "insertion domain" and an N-terminal intertwined dimerization domain ([Hickman et al., 2005](#)) ([Figure 1A](#)). Together, these domains catalyze the chemical steps of DNA nicking, hairpin formation, and DNA strand transfer that comprise *hAT* transposition ([Figure 1B](#)) ([Zhou et al., 2004](#)).

The nucleoprotein assembly that carries out DNA transposition is known as a transpososome. To date, only one eukaryotic transpososome has been structurally characterized, that of the mariner *Mos1* transposon ([Richardson et al., 2009](#)) and it employs a catalytic mechanism that does not involve hairpin intermediates ([Dawson and Finnegan, 2003](#)). Here, we report the

structure and properties of the eukaryotic Hermes transpososome, providing insight into aspects of *hAT* transposition including transposon end recognition and the importance of sub-terminal repeats within *hAT* transposon ends.

RESULTS

The Overall Architecture of the Hermes Transpososome

Complexes of Hermes79–612 and a 16-mer oligonucleotide derived from the *Hermes* LE TIR were crystallized and the X-ray structure solved by single isomorphous replacement with anomalous scattering (SIRAS) at 3.4 Å resolution using a Ta₆Br₁₂ derivative (Figures S1A and S1B available online; Table S1). The complex is an octameric ring of monomers in which each monomer is bound to one TIR DNA (Figure 1C). The ring is ~195 Å in diameter, or almost twice that of a nucleosome core particle (Kornberg, 1977). The assembly is held together by alternating small and large interfaces: short domain-swapped helices at the periphery of the ring contributed by adjacent monomers alternate with an extensive interface projected toward the center of the ring formed by intertwined N-terminal dimerization domains from two monomers. The overall assembly is therefore a tetramer of dimers.

Within each dimer (one is circled in Figure 1C), two TIRs are oriented approximately perpendicular to each other. One DNA is bound such that its distal end points away from the plane of the ring and the other DNA end is more coplanar with the ring. This asymmetry that we observed in the crystal structure (see also Figures S1C and S1D) is a consequence of the small interface between the dimers at the outer edge of the ring. In the dimers, the TIRs point into a cleft on the rim of the ring that is of appropriate size and surface charge to bind target DNA. Thus, we have captured the state in which the ends of an excised transposon are bound and awaiting target capture.

The DNA used to obtain crystals is recessed by one nucleotide (nt) on the nontransferred strand (NTS) and binds more tightly than a blunt-ended 16-mer LE TIR to Hermes79–612. It is also more readily inserted into target DNA than a comparable authentic reaction intermediate (Figure 1D), which has two more 5' nt on the NTS. The observed 8:8 protein-to-DNA stoichiometry is a consequence of using a short TIR to facilitate crystallization, as complexes formed with a longer oligonucleotide containing the terminal 30 bp of the LE display the expected ~8:2 binding (Figure 1E). Binding of two transposon ends is a property of all DNA transpososomes that have been structurally characterized to date (Davies et al., 2000; Richardson et al., 2009; Montañó et al., 2012).

When the DNA-bound dimers within the octamer are compared with those of the apoprotein (Hickman et al., 2005), relative to the intertwined domain, the protein has unfurled and the RNaseH-like and insertion domains have swung out to accommodate the TIRs (Figure S1E). The intertwined dimerization domain is unchanged and can be superimposed with a root mean-square (rms) of 0.56 Å, including all α -carbons. However, in such an alignment, the overall rms is 6.4 Å over all α -carbons of residues 79–612 due mainly to catalytic and insertion domain movements and conformational changes within these domains.

Several lines of evidence indicate that full-length Hermes also forms an octamer. *Escherichia coli*-expressed full-length Hermes elutes at a position on size exclusion chromatography (SEC) consistent with an octamer (Figure S2A), and static light scattering (SEC-MALS) results indicate a monodisperse protein of molecular mass 544,000 Da (calculated octamer molecular mass is 70,110 Da \times 8 = 560,880 Da; Figure S2B). Solution X-ray scattering (SAXS) (Rambo and Tainer, 2013a) data acquired at three protein concentrations yield a molecular mass of 540 ± 80 kDa and a maximum assembly dimension of ~220 Å. When an optimized model for the full-length octamer was generated using AXES (Grishaev et al., 2010), comparison of the scattering intensity profile of the best fitting solution with the experimental scattering data indicated that the data are consistent with a monodisperse octameric scattering particle (Figure 2A).

Full-length Hermes expressed in eukaryotic cells such as budding yeast and Sf9 cells similarly forms a large multimeric species consistent with an octamer as judged by SEC (Figures S2C and S2D). Although we previously reported that full-length Hermes expressed in Sf9 cells is hexameric (Hickman et al., 2005), this was likely a misinterpretation of the preliminary data and our reliance on only one technique (SEC) for molecular weight estimation. Regardless of expression method, we have seen no evidence of subunit dissociation under different buffer conditions, with or without bound DNA, or an equilibrium between the octamer and any other multimeric state.

The form of Hermes used for crystallization lacks 78 amino acids at the N terminus, which contains a BED-finger domain (Figure 1A) (Aravind, 2000). BED domains have been proposed to bind DNA and to coordinate Zn²⁺ through a conserved CCHH or CCHC motif (Aravind, 2000). All *hAT* transposases possess a predicted BED domain, the vast majority of which are the CCHH type (Arensburger et al., 2011).

In the structure, the N termini of all eight monomers are in the center of the ring, and it seems likely this is where the BED domains would reside. To test this hypothesis, we performed negative staining EM using purified full-length Hermes. The fields contained many particles of the appropriate dimensions for an octamer (Figure S3B), and these data were enhanced by picking these particles and performing image averaging (Figure S3C). Averaged top and side-views are shown in Figure 2B. The top view accurately reproduced the crystal structure when it was limited to the same resolution, but had additional positive density in the center, suggesting that this is where the BED domains are located; these features were confirmed in a 3D density map (Figure 2B). Indeed, eight BED domains can be readily modeled within the center of the octamer using the structure of a homology model generated by I-TASSER (Zhang, 2008) so that each BED C terminus is close to and paired with an observed N terminus (Figure 2C). In such a model, the BED domains cluster within the ring without steric clashes to either the rest of the protein or the TIRs.

Hermes-DNA Interactions within the Transpososome

The bound TIRs are essentially B-form DNA except at the active site where the minor groove is widened due to interaction with the α helix that bears E572 of the DDE catalytic motif. Residues

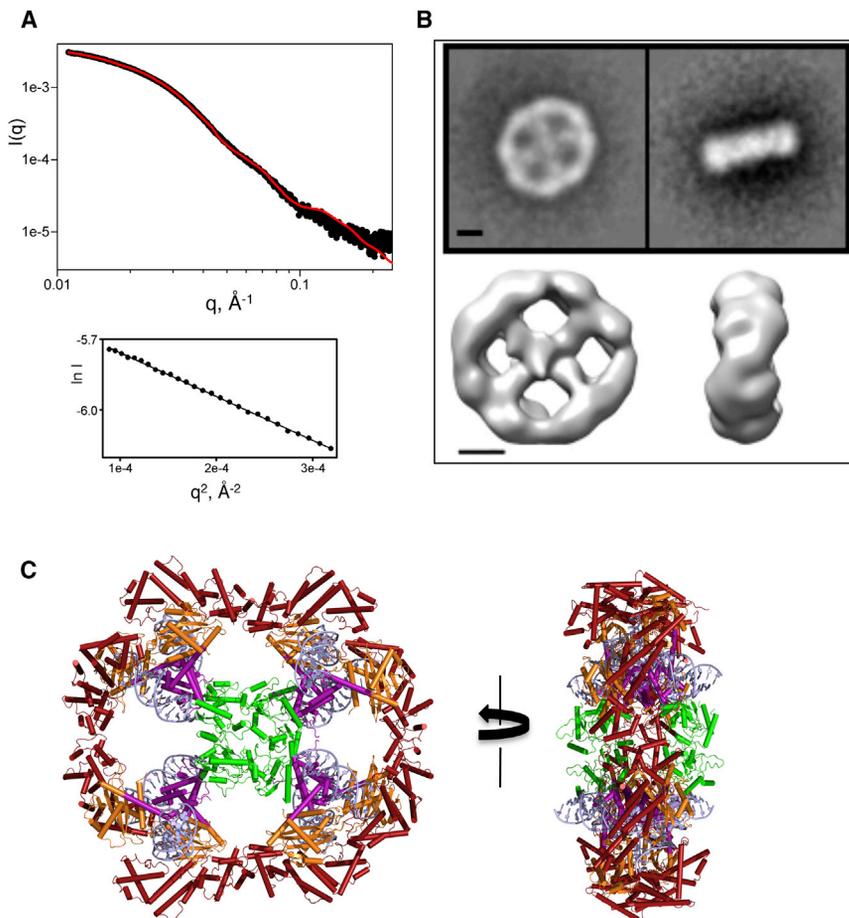


Figure 2. The Full-Length Hermes Octamer

(A) Top: SAXS experimental scattering profile (black, with uncertainty values indicated by error bars) and scattering intensities of the best-fitting model of a full-length Hermes octamer (red). Bottom: Guinier region.

(B) Class-average images from negatively-stained full-length Hermes. Top left: top view average of 303 particles. Top right: side view average of 104 particles. Bottom panels show surface renderings of a density map generated from negatively stained particles, in top and side views. Scale bars represent 50 Å.

(C) Modeled location of eight BED domains in green. Each BED C terminus (red) was constrained to be close to one of the observed Hermes N termini.

See also Figure S3.

contributed by the three domains present in the crystallized form of Hermes participate in TIR binding (Figures 3A and 3B). Close to the TIR tip, a network of amino acids specifically recognize the first 5 bp; beyond those, we observe only one other base-specific interaction, that between base G11 and R149. Residues 585–588, located in the turn between the α helix bearing E572 and the final helix of the protein, contribute interactions within the minor groove. A second cluster of interactions more interior to the TIR tip occur in the major groove and involve residues in the region 138–149; most of these interactions are with the DNA phosphate backbone.

Hermes LE and RE TIRs differ by only two bp (Figure 1A), and in the structure, these bp are not contacted by protein. Furthermore, there are no protein-DNA interactions involving bp 12–16 of the TIR. We observe a bound Na^+ , counterion to a phosphate group of the DNA backbone, coordinated by the main chain carbonyls of E138, E139, and L141.

Several residues that are very highly conserved across the *hAT* superfamily (Arensburger et al., 2011) play important structural roles at the active site where the catalytic residues D180, D248, and E572 converge. For example, W319 stacks against base G1 of the transferred strand (TS) and caps the transposon 3' end (Figure 3C). W319 is important for one or more steps before strand transfer as the W319A mutant protein is severely

impaired for in vitro hairpin formation (Figure 3D), but readily forms single-end joined (SEJ) and double-end joined (DEJ) products when provided with pre-cleaved transposon ends (Hickman et al., 2005). Mutating W319 to either F or Y results in close to wild-type hairpin formation activities with prenicked ends, although W319F appears to be less accurate as it generates more than one hairpin species (Figure 3D, compare lanes 2 and 7). Collectively, these results indicate that the role of W319 is to correctly guide hairpin formation, and other aromatic residues can accomplish

the same task. Among *hAT* transposons, the TS terminal base is almost invariably G or A (Rubin et al., 2001; Kunze and Weil, 2002), suggesting that only purine bases can properly stack with tryptophan during hairpin formation. Another conserved W residue close to the active site is W182 that occupies a pocket between the highly conserved F575 and the peptide plane at T183/D184. W182A is deficient in hairpin formation activity (Figure 3D), yet W182F and W182Y are partially active when the NTS is prenicked, although with a similar inaccuracy as for W319F. From its position in the structure, it appears likely that W182 is important for the correct positioning of the catalytic residue D180.

Two other conserved residues within the active site are H268 and R318. In the apoprotein structure (Figure 3E, left), the R318 side chain was observed close to those of D180 and E572 and was part of a hydrophobic sandwich in which the side chains of R318, W319, and H268 stack against each other. In the presence of TIRs (Figure 3E, right), R318 swings ~ 12 Å out of the active site and forms an ion pair with a DNA backbone phosphate on the TS, 3 bp away from the transposon tip, and now H268 is located within charge-neutralization distance of D180 and E572. H268 is part of a C/DxxH motif that is found in other eukaryotic superfamilies such as the *MULE* (*Mutator* and *Rehavkus*), *P* element, and *Kolobok* transposons; a CxxC motif

is similarly located in *CACTA/Mirage/Chapaev* (CMC) and *Transib* transposases (Yuan and Wessler, 2011).

Relative to the active site into which a transposon end leads, the intertwined dimerization domain interacts with DNA both in *cis* and in *trans* and hence brings together, or synapses, a pair of TIRs. As has been noted, the intertwined domain is topologically similar to the nonamer binding domain (NBD) of RAG1 of the V(D)J recombination system (Yin et al., 2009). The intertwined domains of Hermes and RAG1 appear to be obligate dimers, ensuring the multimerization of the full-length proteins. In both proteins, helices of one monomer wind around each other to form a U-shape cavity in which a helix of the second monomer is buried (Figure S4). Both intertwined domains recognize DNA interior to the sites of cleavage yet, despite this functional parallel, the structures of the two intertwined domain-DNA complexes are only superficially similar. Whereas the two DNA molecules synapsed by the NBD are essentially antiparallel, *Hermes* TIRs are almost perpendicular to each other.

A Hermes Dimer Is the Fundamental Catalytic Unit

The small interface that links dimers into a ring consists of reciprocal interactions in which a short helix (residues 499–505) of one monomer packs into a pocket on the surface of the insertion domain of another (Figures 4A and 4B). Introducing three point mutations into the interface disrupts the octamer, and the resulting dimers (“HermesTM” for triple mutant) are catalytically active in vitro (Hickman et al., 2005). We have also generated dimers by deleting the helix and surrounding residues (“Hermes Δ 497–516”).

Hermes Δ 497–516 dimers can catalyze all of the catalytic steps we measure in vitro for the wild-type (WT) protein. In a plasmid cleavage assay at low ionic strength, dimers are hyperactive relative to WT Hermes when the terminal 30 bp of each transposon end are present (Figures 4C, compare lanes 3–5 and 10–12, and S5); they are also hyperactive when full transposon ends are used (data not shown). Thus, NTS nicking and hairpin formation are not impaired. Furthermore, Hermes Δ 497–516 dimers insert precleaved 30-mer LE oligonucleotides into a plasmid target with substantially higher activity than WT Hermes at low ionic strength (Figure 4D). We sequenced the DEJ products of in vitro plasmid insertion and determined that Hermes Δ 497–516 dimers coordinate the insertion of two precleaved ends with the correct 8 bp target joining spacing (data not shown).

Although Hermes dimers are hyperactive in vitro at low ionic strength, their activities are severely reduced under more physiologically relevant conditions. As shown in Figure 4C, WT Hermes is active for plasmid cleavage over a broad range of salt concentration (0.05–0.3 M NaCl) whereas Hermes Δ 497–516 is barely active once the ionic strength is raised above 150 mM (compare lanes 7–9 with 14–16).

Hermes Dimers Are Inactive In Vivo

Consistent with the lack of in vitro activity at physiological ionic strength, Hermes dimers are inactive in vivo in all of the cell types we have tested. In an interplasmid transposition assay in fly embryos (Sarkar et al., 1997), WT Hermes had a transposition frequency of 5×10^{-5} (piggyBac control was 2.4×10^{-5}), whereas

there were no transposition events for HermesTM. In an excision assay in *Drosophila* S2R+ cells, activity was undetectable for both HermesTM and Hermes Δ 497–516 (Figure S6A); the result was the same in HEK293 cells (data not shown). Excision activity was similarly undetectable in *Saccharomyces cerevisiae* (Figure S6B). We have verified that HermesTM and Hermes Δ 497–516 are as stable as WT Hermes in vivo (Figure S6C) and are transported into the nucleus of HEK293 cells (Figure S6D).

The Target Binding Cleft

In the structure, two TIRs are synapsed by each dimer and appear correctly oriented for strand transfer into target DNA. The obvious cleft on the rim of the ring where pairs of transposon TIRs converge is lined with positively charged residues contributed by the insertion and RNaseH-like domains (Figures 5A and 5B), many of which are well-conserved in the *Ac* family of *hAT* transposases (Arensburger et al., 2011). Single point mutation data indicate that residues in this region are important for in vivo transposition (Figures 5C and 5D; Tables S2 and S3). For example, the transposase mutant K299A is severely impaired for germ-line transposition in *D. melanogaster* and has >100 \times lower frequency of transposition in the somatic nuclei of developing *D. melanogaster* embryos. R306A has \sim 8 \times lower frequency of somatic transposition yet is able to generate transgenic offspring at a frequency comparable to WT Hermes. K292A and K300A also show reduced transposition frequency in somatic transposition assays and are impaired at generating transgenic offspring.

The cleft is \sim 80 Å from end to end, a consequence of the angle at which it traverses the \sim 50-Å-wide octameric ring (Figure 5A), and has a distinct inward arc suggesting that target DNA is bent when bound. Within each dimer, the two terminal 3'-OH groups of the TS are located 35 Å apart, too far for coordinated insertion into opposite target strands 8 bp apart if target DNA is straight B-form DNA (in which case nucleophilic attack would occur on phosphate atoms \sim 29.5 Å apart, and the attacking -OH groups should only be \sim 26 Å apart), but appropriately spaced if target DNA is bent. Kinked target DNA from the PFV intasome (Maertens et al., 2010) can be docked into the cleft and in the resulting model, the two 3'-OH groups approach opposite strands \sim 8 bp apart (Figure 5E). Although this model approximates how target DNA might be bound, we suspect that kinks will occur at the two bp represented by the T/A bp of the nTnnnnAn sequence that is the target site preference of Hermes (Gangadharan et al., 2010). Bent target DNA, observed in the MuA transpososome (Montaño et al., 2012) and shown to be important for IS231A and Tn10 transposition (Hallet et al., 1994; Pribil and Haniford, 2003), has been proposed to be a conserved feature of DNA transposition that serves to drive the reaction forward by preventing reversal of strand transfer (Cherepanov et al., 2011; Montaño et al., 2012).

The Importance of the Subterminal Repeats

One notable feature of *hAT* transposon ends are multiple copies of short 5–6 bp subterminal sequences, apparently haphazardly arrayed in both orientations without evident periodicity (Kunze and Starlinger, 1989; Kim et al., 2011; Liu and Crawford, 1998; Liu et al., 2001; Urasaki et al., 2006). Accumulating data indicate that *hAT* transposases recognize their transposon tips in a

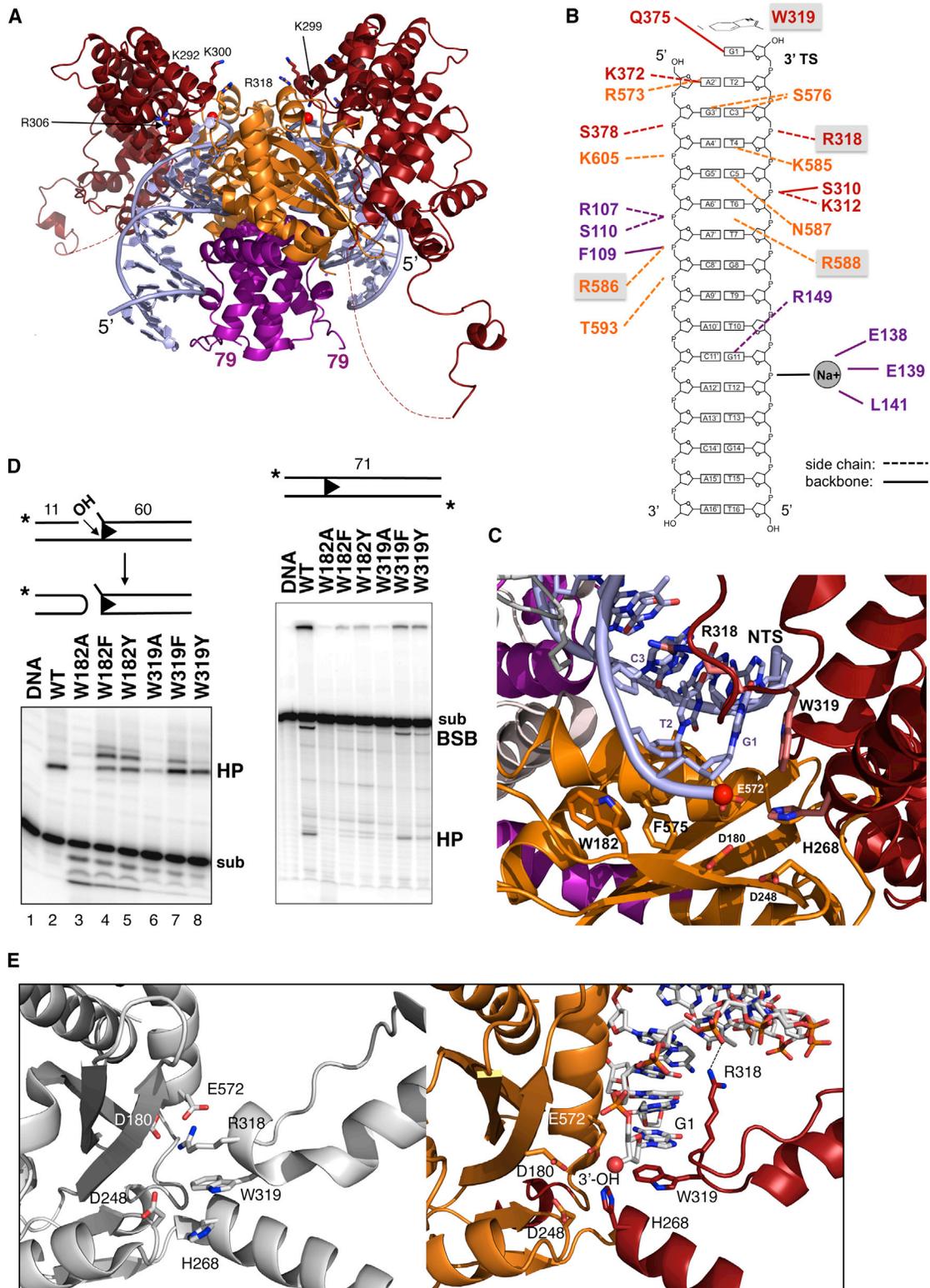


Figure 3. Hermes-DNA Interactions

(A) View of one dimer within the octamer. Red spheres indicate the 3'-OH groups that converge in a cleft lined with basic residues, some of which are in ball-and-stick representation and labeled. Dashed lines indicate the disordered loop between residues 464 and 493.

(legend continued on next page)

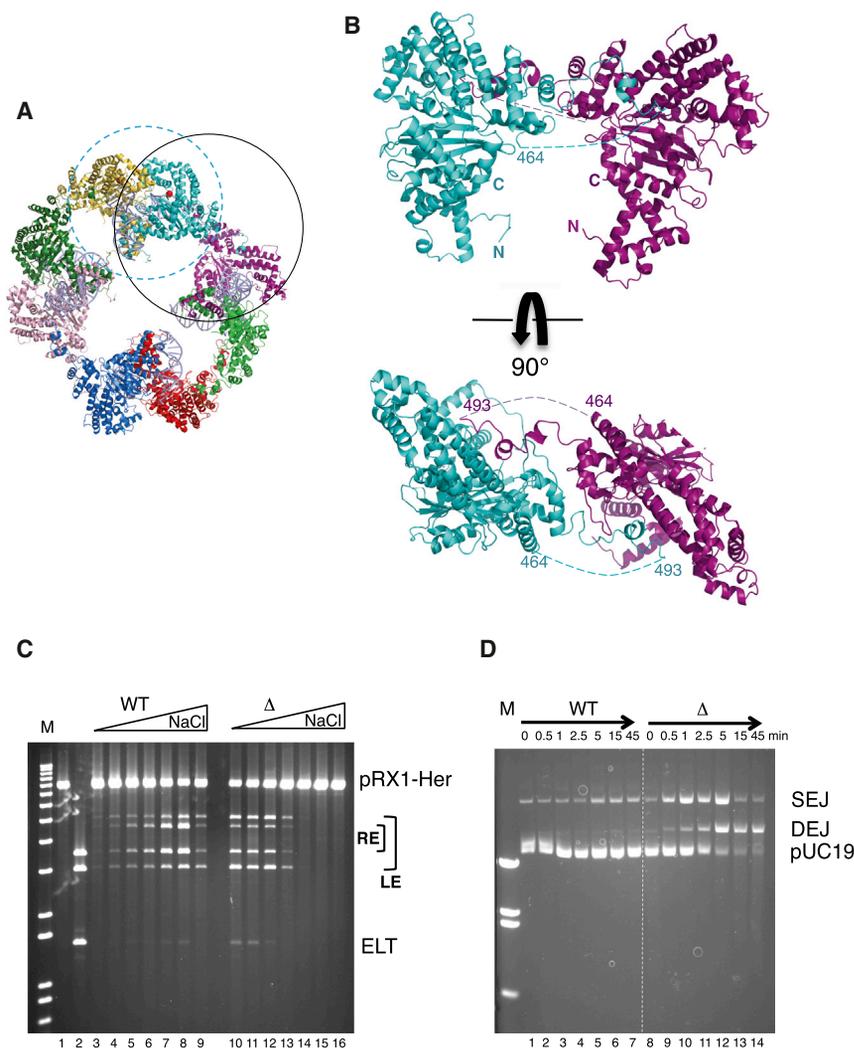


Figure 4. A Dimer Is the Catalytic Unit

(A) In the octamer, the dimer where two monomers contribute to the small interface is circled in solid black, and the catalytically active dimer formed by the intertwined domain is circled in dashed blue.

(B) Views of the small interface. Dashed lines indicate the disordered loops.

(C) Plasmid cleavage activity of WT Hermes and Hermes Δ 497–516 as a function of [NaCl] (0–0.3 M). Reactions were at 30°C for 60 min with 8.6 nM protein and 1 nM pRX1-Her (Figure S5). After restriction digest, bands indicate LE and RE cleavage as marked. Cleavage at both ends results in an excised linear transposon (ELT). Lanes 3 and 10: 0.1 mM NaCl. Lanes 4 and 11: 50 mM. Lanes 5 and 12: 100 mM. Lanes 6 and 13: 150 mM. Lanes 7 and 14: 200 mM. Lanes 8 and 15: 250 mM. Lanes 9 and 16: 300 mM.

(D) LE30 strand transfer activity of WT Hermes and Hermes Δ 497–516 as a function of time at 23°C in buffer containing 10 nM protein, 22.9 nM LE30, and 50 mM NaCl.

See also Figures S5 and S6.

bipartite manner, with weaker transposase binding to the TIRs and stronger binding by an N-terminal domain to these subterminal repeat sequences (Kunze and Starlinger, 1989; Becker and Kunze, 1997; Mack and Crawford, 2001; Urasaki et al., 2006; Kahlon et al., 2011; Kim et al., 2011).

A 5'-GTGGC repeat has been previously identified within *Hermes* ends where one repeat on each end overlaps with the TIR, and it has been proposed that some of these repeats are likely important for DNA binding (Kim et al., 2011). There are seven such repeats within *Hermes* (Figure 6A), and the location of the outermost repeat on each end, defined here as 13 bp encompassing the GTGGC repeat and designated “LE₁” (bp 13–25

from the tip of LE) and “RE₁” (bp 13–25 from the RE tip), taken in light of the structure, suggests that these are bound by the BED-finger domain: Hermes79–612 does not interact with the TIR after bp 11, and if TIR DNA were extended toward the center of the octamer, the BED domains would likely encounter DNA subterminal to the TIRs.

To determine whether the repeats within *Hermes* ends are important for in vitro activity, we mutated the ends in several ways. As shown in Figure 6B, mutation of either LE₁ or RE₁ to random

sequence leads to loss of plasmid cleavage activity on that end. When the position of LE₁ is shifted relative to the transposon tip, a shift of one bp toward the tip is tolerated (Figure 6C; lanes 4–6), whereas a one bp shift in the other direction is less so (lanes 7–9). Inserting 3 or 5 bp between the TIR and LE₁ leads to a complete loss of plasmid cleavage activity on LE (lanes 10–15), indicating that the phasing of these two regions is crucial. In this plasmid cleavage assay, LE cleavage can occur without RE cleavage and vice versa, and BED-deleted Hermes79–612 is barely active under these experimental conditions (Figure 6B, lanes 2–4). In a strand transfer assay, mutation of LE₁ and RE₁ also leads to a loss of activity (Figure 6D), and a similar

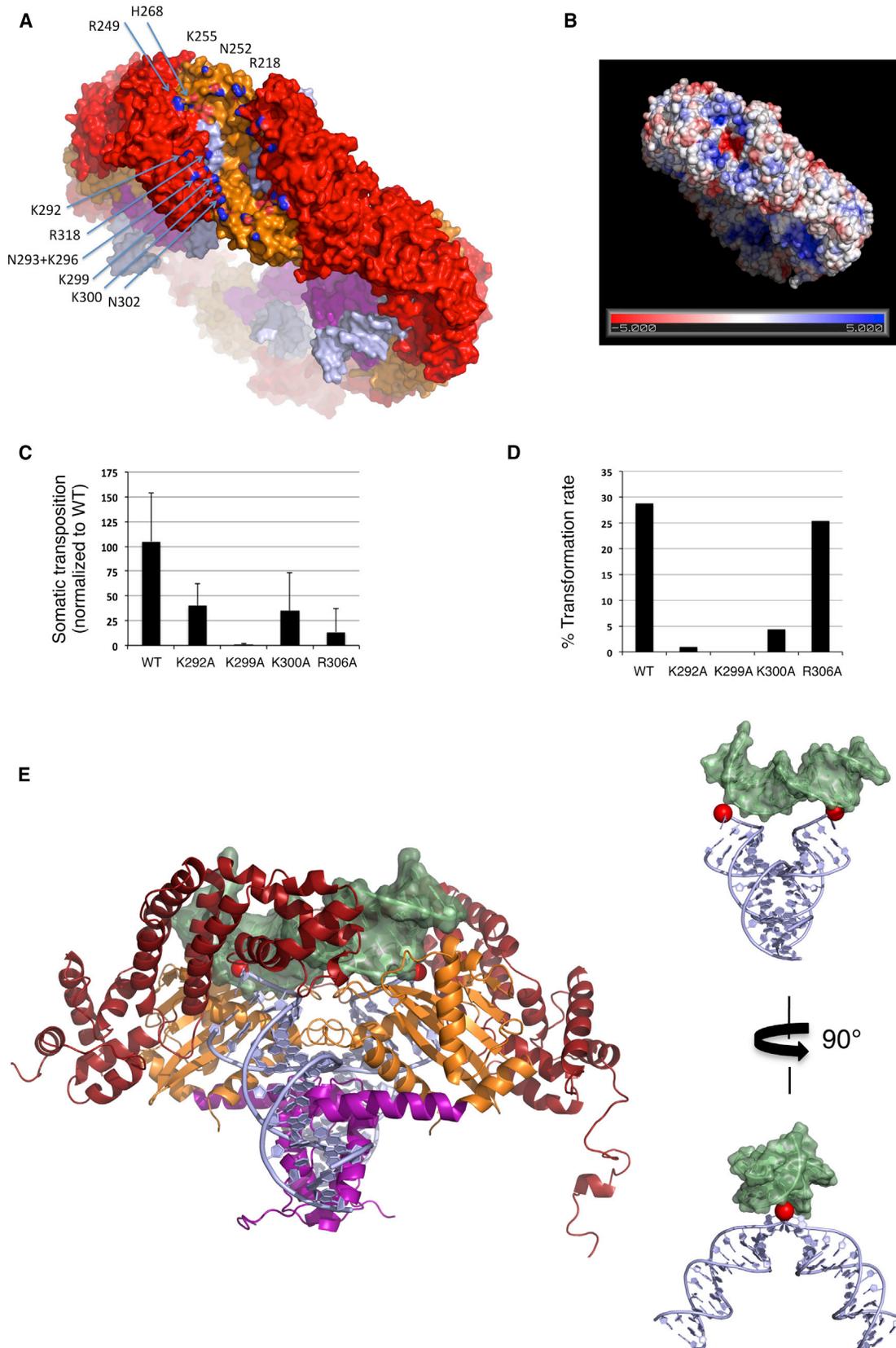
(B) Summary of protein-DNA interactions. Letter color corresponds to the domain color in Figure 1A. Boxed residues are highly conserved across the *hAT* superfamily.

(C) Close-up of active site.

(D) Effect of mutating W182 and W319 on in vitro cleavage and hairpin formation. sub, substrate; HP, hairpin; BSB, bottom strand break. BSB results from TS cleavage upon hairpin formation.

(E) Active site without (left) and with (right) bound TIR.

See also Figure S4.



(legend on next page)

loss of activity was observed for Hermes79–612 with an unmutated LE30 end. Thus, it appears that the BED domain interacts with the outermost subterminal repeat on each end and this is important for both cleavage and strand transfer.

To more finely map the role of the DNA sequence just subterminal of the TIR, we measured strand transfer activity as a function of LE length. As the length was increased from 17 bp to include more of LE₁, we observed a dramatic increase in activity once the DNA length exceeded 21–22 bp (Figure S7B and S7C). This strongly suggests that the first 9 bp of the LE₁ subterminal repeat are sufficient to engage the BED domain.

Finally, we mapped the interaction of Hermes with LE₁ using a competition binding assay based on the gel shift of a LE30 probe. When the nuclear extract from *Drosophila* S2 cells expressing Hermes (Figure S7D) was incubated with LE30, one major band was observed with several faint bands above and below (Figure 6E). This binding could be competed off with specific LE30 competitor, and we tested the ability of mutated probes, each with a single transversion mutation at each position, to perturb LE30 binding. The results map the region of strongest interaction to bp 14–24 (Figure 6F), essentially spanning the LE₁ subterminal repeat.

The Importance of Long Ends

In *Hermes*, the transposase gene is flanked by 449 bp of LE and 464 bp of RE DNA (Warren et al., 1994). For *hobo*, whose transposase is 55% identical in sequence to Hermes, in vivo transposition requires at least 140 bp of its LE and 65 bp of RE (Kim et al., 2011). To establish whether Hermes also has an end length requirement, we tested transposition in vivo as a function of transposon length (Table S4). In a *Drosophila* cell line stably expressing Hermes, there was an ~2.4-fold decrease in transposition when the LE was reduced from 444 to 305 bp and the RE simultaneously reduced from 384 to 307 bp. Notably, no transposition was seen if only 30 bp of *Hermes* LE and RE were present, indicating that sequences beyond the TIRs and the first subterminal repeat are needed. End asymmetry was also important as no transposition was observed when the ends were symmetrized by replacing RE with 305 bp of LE, suggesting that both the LE and RE are required for transposition.

Why Are There Two Binding Sites at the *Hermes* Transposon Tips?

To understand why Hermes recognizes both its TIR and a subterminal repeat, we used fuzznuc from the EMBOSS suite (Rice et al., 2000) that searches for short nucleotide sequences in genomes. As there is no full genome assembly for *M. domestica*, a

natural host of *Hermes*, we searched the *D. melanogaster* genome (120 Mb; release 5) for the sequence 5'-CAGAGnnnnnC, the TIR bases that are specifically contacted by Hermes where “n” is any base. This sequence occurs 2,421 times, indicating that the TIRs are not sufficient to uniquely direct Hermes to its transposon ends. The 8 bp sequence present in all four of the LE *Hermes* subterminal repeats, 5'-CAAGTGGC (Figure 6A), occurs 3,828 times. However, when the two sequences were combined with the correct TIR/LE₁ spacing, this longer sequence is present only once. As the *M. domestica* genome is ~1.6-fold larger than that of *D. melanogaster* (Gao and Scott, 2006), this suggests that the TIR and a subterminal repeat together are enough to allow Hermes to find its transposon ends within the genomic background of its host.

Identification of Alternate Cleavage Sites at RE

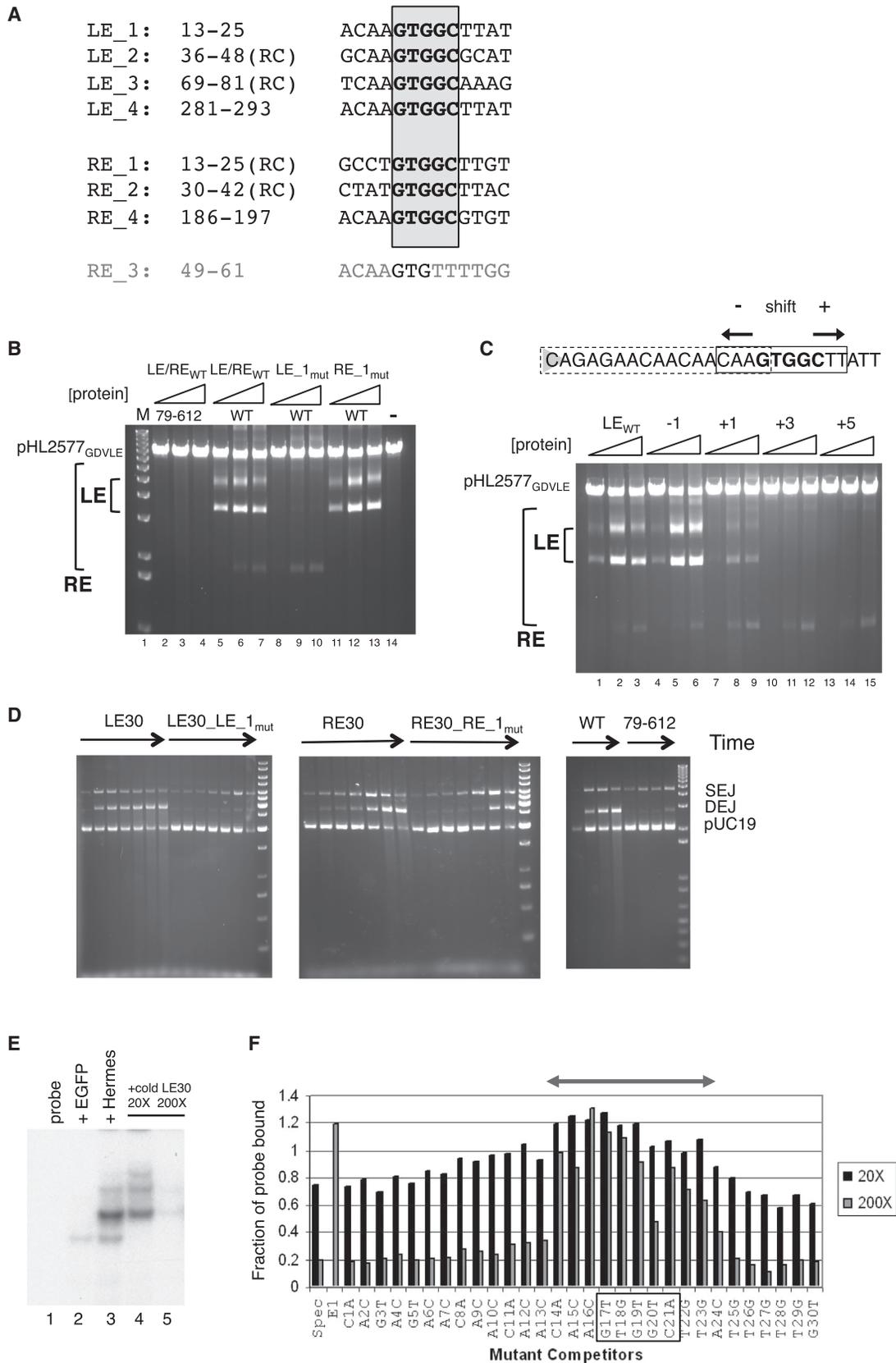
When we sequenced and characterized 63 Hermes-catalyzed in vitro transposition events into a target plasmid, most were perfect, i.e., only *Hermes* DNA had inserted into the target, all *Hermes* DNA was intact, and there were 8 bp TSDs (Table S5). Curiously, among the imperfect transposition events, TIR deletions occurred only at the RE and usually involved 18 bp or 37 bp deletions. This discrete deletion size could arise from alternate binding of Hermes to its RE coinciding with cryptic TIR sequences, in which Hermes mistakes RE₂ or another region (designated RE₃) for RE₁ (Figure 7A). This seems possible as the bases conserved in the cryptic TIRs correspond to those which formed specific contacts in the Hermes structure: four of the five terminal bp and C11 on the NTS. This suggests that there may be one more RE subterminal repeat that was not initially identified as it does not have a GTGGC repeat yet is clearly related to the sequences of LE₁ and LE₄. Thus, we can identify four subterminal repeats on *Hermes* LE and four on RE, asymmetrically arrayed.

DISCUSSION

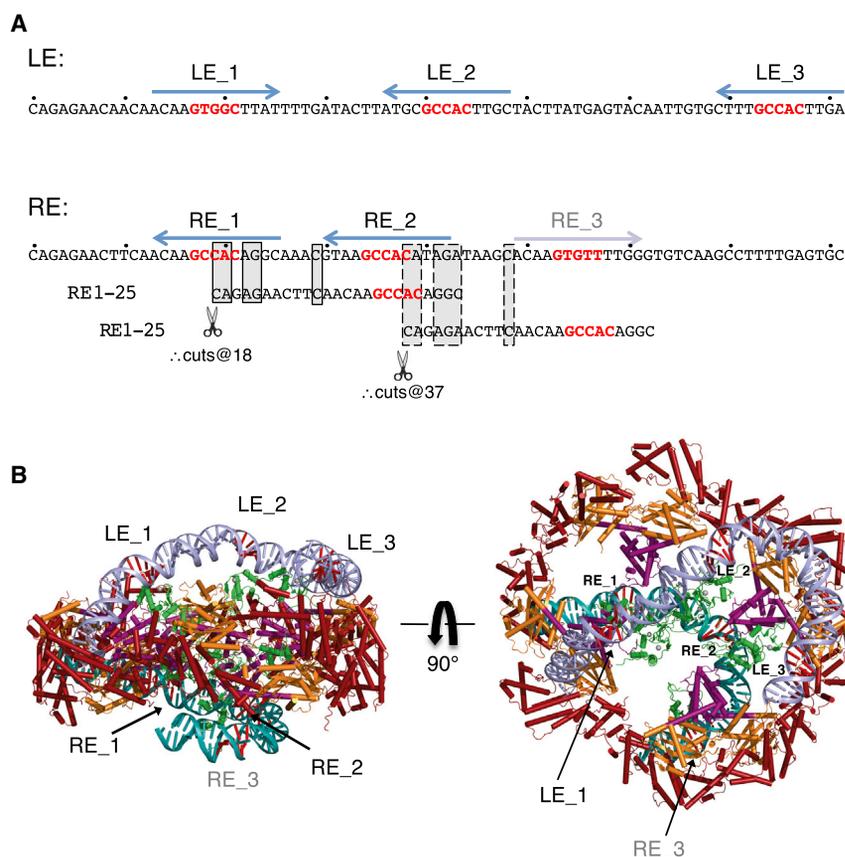
The most surprising aspect of the *Hermes hAT* transposase structure is that it is an octamer. Some other transpososomes exhibit an overabundance of monomers yet the octameric ring of Hermes seems excessive. The MuA transposase is a tetramer in its active form (Lavoie et al., 1991), and the “extra” monomers provide additional DNA binding domains for multiple MuA binding sites located on phage genome ends; the related retroviral intasomes are also obligate tetramers for reasons not yet clear (Li et al., 2006; Cherepanov et al., 2011). Because Hermes dimers generated by mutagenesis are not functional in vivo but

Figure 5. Target Binding

- (A) Surface representation of the octamer rim and selected residues lining the cleft; only those of one monomer are labeled (N atoms are colored blue).
 (B) Electrostatic potential calculated using APBS (Baker et al., 2001) with only one dimer complexed with DNA and two active site Mn²⁺ at 150 mM salt.
 (C) Effect of single point mutations on somatic transposition frequency in *D. melanogaster* embryos (Table S2). Error bars represent SD in *Hermes* transposition frequency (corrected for a *piggyBac* internal control) for three replicate injections.
 (D) Effect of single point mutations on germline transformation rate in *D. melanogaster* calculated by dividing the number of transgenics by the number of fertile crosses (Table S3).
 (E) Left: Hermes dimer docked to target DNA (light green) of the PFV intasome, PDB ID 3OS1. Right: model of the DNAs alone with *Hermes* TIRs in light blue and PFV target DNA in green. Red spheres are the TS 3'-OH groups.
 See also Tables S2 and S3.



(legend on next page)



appear capable of all of the *in vitro* activities of a transposase, the octamer must confer a crucial property beyond catalysis. We have ruled out that the octamer simply provides intracellular stability to an otherwise unstable dimer or that nuclear localization is compromised.

The structure and biochemical analyses presented here suggest that one important aspect of the octamer is that it provides multiple BED domains for transposon end binding. Although we observed eight transposon TIRs bound in our crystallized complex, this is most likely an artifact of using short oligonucleotides to facilitate crystallization as the Hermes octamer is unable to bind more than two transposon ends *in vitro* once

The location of BED domains inferred by the EM data (missing in the crystal structure due to the obstinate behavior of the full-length protein) clearly places them so that two could interact with the outermost subterminal repeat on each end. Our observations that mutation of LE₁ or RE₁ abolishes both DNA cleavage and strand transfer and that the spacing of LE₁ relative to the tip of the transposon is critical, implicate these sequences as specific transposase binding sites.

One obvious attribute of an octamer is that, in addition to two BED domains for binding the outermost repeat at each transposon end, it possesses six more that can be used to bind additional subterminal repeats. This would explain a conserved

Figure 7. Hermes/DNA Recognition beyond the TIRs

(A) Arrangement of subterminal repeats within NTS of LE 1–81 and RE 1–81, where 5′-GTGGC (or its reverse complement) of each repeat is in red. Below the RE sequence are two alignments with RE bp 1–25 (common bases are boxed with solid lines for the alignment of RE₂ with RE₁ and dashed lines for the alignment of RE₃ with RE₁). Transposition assay results suggest that Hermes can mistake RE₂ and RE₃ for RE₁, resulting in aberrant cleavage.

(B) Modeled Hermes binding to its transposon ends. The structure of the first 16 bp of each end are those in the observed crystal structure. On LE (light blue), modeled DNA from bp 17–81 is bent to bring LE₂ and LE₃ close to the presumed locations of the BED domains (green; PDB ID 2CT5). RE (turquoise) bp 17–66 are modeled using 50 bp of nucleosomal DNA (PDB ID 1AOI). 5′-GTGGC of each subterminal repeat is in red. The modeled BED domains have been allowed to move within the ring relative to the model in Figure 2C. It is not clear how LE₄ and RE₄ might be bound. See also Table S5.

the oligonucleotide length is increased. Instead, we favor a model in which one catalytic dimer within the octamer binds two ends, and the remaining monomers contribute important interactions to the same ends.

Figure 6. Subterminal Repeats within Hermes Ends Are Recognized by the BED Domain

(A) Alignment of subterminal repeats containing 5′-GTGGC (black) where numbering is the distance in bp from the tip of either LE or RE. “RC” indicates that the repeat is in the opposite orientation relative to LE₁. Hermes contains no other 5′-GTGGC repeats. An eighth repeat “RE₃” (gray) was inferred from *in vitro* transposition reactions.

(B) Effect of mutating LE₁ or RE₁ on plasmid cleavage as a function of protein concentration. For each set, [protein] are 9.5 nM, 47 nM, and 95 nM.

(C) Effect on plasmid cleavage of shifting part of the LE₁ sequence (boxed) relative to the transposon tip.

(D) Time course of strand transfer upon mutating LE₁ or RE₁, and as a function of deleting the BED domain. For the first four sets, time points are 0, 0.5, 1, 2.5, 5, 15, and 45 min at 25°C in buffer containing 15 nM protein, 28.6 nM ends, and 200 mM NaCl; for the fifth set, comparison of LE30 strand transfer activity by WT Hermes to that of Hermes79–612, time points are 0, 1, 7, and 45 min.

(E) Electromobility shift assay with *Drosophila*-expressed Hermes and a LE30 probe. Lane 1: probe alone. Lane 2: EGFP control nuclear extract. Lane 3: Hermes nuclear extract. Lanes 4 and 5: Hermes nuclear extract with the indicated excess of LE30-specific competitor.

(F) Results of EMSA competition assay with single base mutant probes. Dark bars indicate 20× competitor levels and lighter bars indicate 200× levels. E1 is the nonspecific competitor at 200×. If the mutation had no effect, values similar to that for the specific competitor (Spec) are expected. The double-headed arrow indicates the region of strongest interaction, which includes the 5′-GTGGC repeat (boxed).

See also Figure S7 and Table S4.

feature of *hAT* transposons which is an abundance of short DNA repeats scattered throughout transposon ends. The structure therefore suggests a model for end binding in which the transposon ends provide multiple internal binding sites to tether the transposase, and this directs weakly bound TIRs into two active sites within the same dimer (Figure 7B). As we only observe TSDs of 8 bp (Table S5), productive transposition must only occur when the two ends are bound by the same dimer.

Multiple site-specific DNA binding domains could also provide increased avidity to regions of DNA containing many BED domain binding sites, enhancing binding specificity to the transposon against a background of nonspecific DNA binding. The lack of in vivo activity by Hermes dimers, although an effect of physiological ionic strength, might also be a manifestation of not having enough BED domains to stably bind transposon ends. Another consequence of octamerization are four nonspecific DNA binding sites around the rim of the ring that could also have an important avidity effect in capturing target DNA.

The proposal that an octamer provides multiple BED domains to recognize multiple subterminal repeats also explains the observation that for all *hAT* transposons studied to date, activity in vivo depends on long transposon ends and the bipartite transposon tips are not sufficient. For example, *Tol2* requires the terminal 200 bp of its LE and 150 bp of its RE for excision and transposition (Urasaki et al., 2006). For *Ac*, 200 bp at both ends are needed for wild-type levels of transposition (Coupland et al., 1989), and for *Tag1*, ~100 bp are required at each end (Liu et al., 2001). Although short transposon ends can be used for in vitro assays with Hermes, perhaps because of the high concentration of specific DNA involved, 30 bp of each end are not sufficient for in vivo transposition in *Drosophila* (Table S4).

The structure provides insight into another aspect of *hAT* transposition: where it has been examined, in vivo transposition does not occur with artificial transposons consisting of either two LEs or two REs (Coupland et al., 1989; Urasaki et al., 2006). The DNA binding asymmetry we observe within each dimer—one TIR juts out from the plane of the ring whereas the second is essentially in plane—suggests that if multiple interactions between BED domains and repeat sequences are needed, these will be structurally asymmetric as well. Consistent with this possibility, the subterminal repeats we identified here are arranged differently on the two ends. This in turn dictates that, to model LE and RE binding to the Hermes octamer, the two ends must adopt different conformations. Thus, if asymmetry is an inherent feature of the transpososome, it seems unlikely that symmetric ends could be structurally accommodated. Interestingly, asymmetric recognition sites are also a property of RAG1/2-mediated recombination which only rearranges gene segments with two different spacer lengths in their recombination signal sequences (RSSs), the so-called 12/23 rule (for review, see Schatz and Swanson, 2011).

It is not yet known whether all *hAT* transposases are active as octamers. It is possible that the basic building block of *hAT* transposases is a dimer and that some have evolved different assembly properties to exploit DNA binding to subterminal repeats. Among *hAT* transposases, only hobo and Homer have three phenylalanine residues that align with Hermes F502/F503/F504 of the small interface helix (Arensburger et al., 2011) and would

be predicted to exhibit the same potential for multimerization through a hydrophobic helix motif. Hermes, hobo, and Homer are phylogenetically each other's closest relatives and are further distinguished from other *hAT* transposases by a CCHC-type BED motif. Perhaps these three transposases have a particular transpososome architecture that takes advantage of a large preassembled multimer to bind DNA. *Ac* and certain other characterized transposons have many more repeats within their transposon ends than the eight we have been able to identify for Hermes. For example, the minimal *Ac* transposase binding site is repeated 25 times at its 5' end and 20 times at its 3' end (Becker and Kunze, 1997). In *Tol2*, the 5-bp sequences 5'-AAGTA and 5'-GAGTA occur 33 times in the transposon ends (Urasaki et al., 2006), and *Tam3* possesses 40 copies of a 5-bp motif transposase binding site (Hashida et al., 2006). This difference might reflect an architecturally distinct solution for binding subterminal repeats during transposition.

Although the catalytic unit of Hermes appears to be a dimer, the monomer arrangement and trajectories of TIR ends within the synaptic complex differ from those of the structurally characterized dimeric transpososomes, of Tn5 and *Mos1* (Davies et al., 2000; Richardson et al., 2009). On the other hand, there is an uncanny resemblance between the Hermes dimer and the "anchor-shaped" organization of RAG1/2 bound to RSS DNA deduced by EM (Grundy et al., 2009). The contours of the assemblies are similar, and the proposed paths of DNA through RAG1/2 echo those of the TIRs within a Hermes dimer. In addition to their shared mechanism of generating double-strand breaks via a flanking DNA hairpin, other parallels between *hAT* transposases and RAG1 have been noted including a predicted α -helical insertion domain into the RNaseH-like fold of RAG1 (Lu et al., 2006) and the bipartite organization of their DNA binding partners (Kunze and Weil, 2002).

Transposons are versatile genetic tools and several such as Tn5, *Tol2*, *piggyBac*, and *Sleeping Beauty* are widely used for genomic manipulation experiments (VandenDriessche et al., 2009). The structure of Hermes determined here opens up avenues for developing improved versions of *hAT* transposons for nonviral gene delivery or other genome applications. For example, it might be possible to insert a specific DNA binding domain close to the clefts at the periphery of the ring where it would contact target DNA leading in or out of the active sites; this might direct insertion to a specific chromosomal location. Alternatively, our experiments implicate sloppy RE processing in the generation of aberrant transposition events, a result that evokes the effect of mutating one end of retroviral DNA so that it is a poor substrate for integrase, which in turn leads to abnormal proviruses and genomic rearrangements (Oh et al., 2006). It will be interesting to see if changing the RE subterminal repeat sequences might improve insertion fidelity, and the question of how BED domains interact with subterminal repeats to coordinate transpososome function is an area of active investigation.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification

For *E. coli* expression, the coding regions for full-length Hermes and Hermes79–612 were cloned into pBAD/Myc-His (Invitrogen) such that the

resulting proteins were untagged at either terminus. After screening Cys mutations for improved solubility properties, all structural work was performed using Hermes79–612 C519S. Proteins were expressed at 19°C in Top10 cells and soluble protein purified using Heparin Sepharose and size exclusion. Purified untagged full-length Hermes was used for SEC-MALS, SAXS, and negative stain EM analysis. The deletion dimer was similarly purified. Hermes W319 and W182 point mutants were purified as described in Zhou et al. (2004).

Crystallization and Structure Determination

After size exclusion, Hermes79–612 was concentrated and LE16-1T DNA (IDT) added to a final molar ratio of 1:1.2 protein:DNA. The complex was formed by dialysis against 0.18 M KCl, 20 mM HEPES pH 7.5, and 5 mM DTT, and crystals were grown at 19°C by the hanging drop method. To screen for derivatives, crystals in stabilizing buffer were soaked in 0.2–5 mM Ta₆Br₁₂²⁺ (Jena Bioscience) for 3 to 24 hr prior to cryoprotection. Crystallographic details can be found in the [Extended Experimental Procedures](#).

In Vitro Assays

To assess the ability of Hermes to carry out discrete steps of the transposition reaction, several assays were used. To assess cleavage, purified Hermes was incubated with plasmids containing either the full Hermes LE and RE (pHL2577 and variants) or the final 30 bp of each end (pRX1-Her). When mutated ends were used, changes were introduced using the QuikChange method (Agilent). After incubation, DNA was isolated and subjected to restriction digest to generate DNA fragments that could be identified after separation on an agarose gel. Cleavage assays were also carried out using oligonucleotides containing 60 bp of *Hermes* LE and 11 bp of flanking DNA.

To assess the ability of Hermes to generate hairpins when supplied with a prenicked LE, protein was incubated with a 71 bp oligonucleotide containing an intact bottom strand with a nicked top strand. For strand transfer, Hermes was incubated with oligonucleotides representing various lengths of pre-cleaved LE sequence and a pUC19 target plasmid.

For in vitro transposition into pGDV1, nuclear extracts of a *Drosophila* S2 cell line stably expressing Hermes were incubated with a *Hermes* donor plasmid and pGDV1. Recovered plasmid DNA was transformed into *E. coli* and the transposition events characterized.

EMSA and Competition Assay

Nuclear extracts of S2 cell lines stably expressing either Hermes or EGFP as a negative control were prepared, and extracts and probes were incubated for 20 min at room temperature prior to being run on a 5% TBE polyacrylamide gel. When competitors were used, nuclear extracts and competitors were incubated for 15 min at room temperature, then labeled probe was added and incubated for an additional 20 min.

In Vivo Transposition

To compare the activities of WT and HermesTM in *Drosophila* embryos, a five-plasmid assay was carried out essentially as described by Sarkar et al. (1997). The transposition frequency was calculated as the number of independent events divided by the donor titer. In *Drosophila* cells stably expressing Hermes, a two-plasmid assay was carried out in which one plasmid supplied Hermes transposon ends of varied length and pGDV was the target.

Model for Transposon End Binding

To model LE binding to the Hermes octamer, the “make-na” server at <http://structure.usc.edu/make-na> was used to generate three bent DNA models that were overlapped to produce a 81-mer.

Details of experimental procedures can be found in the [Extended Experimental Procedures](#). See also [Table S6](#).

ACCESSION NUMBERS

The Protein Data Bank accession number for the coordinates of the Hermes-DNA structure reported in this paper is 4D1Q.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.05.037>.

AUTHOR CONTRIBUTIONS

A.B.H., F.D., N.L.C., and P.W.A. designed the biochemical experiments. A.B.H. and F.D. performed the X-ray crystallography, determined and analyzed the structure. A.B.H., H.E.E., and X.L. performed in vitro biochemistry experiments. X.L. performed mammalian in vivo experiments. J.A.K., T.L., A.D., and P.W.A. were responsible for in vivo *Drosophila* experiments. The EM analysis was performed by G.T. and A.C.S. The SAXS analysis was performed by A.G. and A.B. A.B.H., P.W.A., N.L.C., and F.D. wrote the manuscript.

ACKNOWLEDGMENTS

This work was partially funded by the NIH Intramural Programs of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) (F.D. and A.B.) and the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) (A.C.S.). P.W.A. and N.L.C. were funded by Public Health Service (PHS) award A1045741. Crystallographic data were collected at the SER-CAT 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory (ANL). For the SAXS experiments, we gratefully acknowledge use of the shared scattering beamline 12-IDC resource allocated under the PUP-77 agreement between the National Cancer Institute and ANL and thank Soenke Seifert (ANL) for his support. Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract W-31-109-Eng-38. We thank Sriram Subramanian and members of his laboratory for their preliminary EM efforts. We also thank Nadine Samara for critical reading of the manuscript, Primrose Musingarimi for help with Sf9-expressed Hermes, Robert Hice for assistance with EMSA and insect cell culture experiments, and Susan Chacko of the Helix systems group, CIT, NIH for advice and help with fuzznuc.

Received: February 10, 2014

Revised: April 10, 2014

Accepted: May 12, 2014

Published: July 17, 2014

REFERENCES

- Aravind, L. (2000). The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. *Trends Biochem. Sci.* 25, 421–423.
- Arensburger, P., Hice, R.H., Zhou, L., Smith, R.C., Tom, A.C., Wright, J.A., Knapp, J., O'Brochta, D.A., Craig, N.L., and Atkinson, P.W. (2011). Phylogenetic and functional characterization of the *hAT* transposon superfamily. *Genetics* 188, 45–57.
- Atkinson, P.W., Warren, W.D., and O'Brochta, D.A. (1993). The *hobo* transposable element of *Drosophila* can be cross-mobilized in houseflies and excises like the *Ac* element of maize. *Proc. Natl. Acad. Sci. USA* 90, 9693–9697.
- Aziz, R.K., Breitbart, M., and Edwards, R.A. (2010). Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res.* 38, 4207–4217.
- Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* 98, 10037–10041.
- Becker, H.-A., and Kunze, R. (1997). Maize *Activator* transposase has a bipartite DNA binding domain that recognizes subterminal sequences and the terminal inverted repeats. *Mol. Genet.* 254, 219–230.
- Biémont, C. (2010). A brief history of the status of transposable elements: from junk DNA to major players in evolution. *Genetics* 186, 1085–1093.

- Blackman, R.K., Koehler, M.M.D., Grimaila, R., and Gelbart, W.M. (1989). Identification of a fully-functional *hobo* transposable element and its use for germ-line transformation of *Drosophila*. *EMBO J.* **8**, 211–217.
- Cherepanov, P., Maertens, G.N., and Hare, S. (2011). Structural insights into the retroviral DNA integration apparatus. *Curr. Opin. Struct. Biol.* **21**, 249–256.
- Coupland, G., Plum, C., Chatterjee, S., Post, A., and Starlinger, P. (1989). Sequences near the termini are required for transposition of the maize transposon *Ac* in transgenic tobacco plants. *Proc. Natl. Acad. Sci. USA* **86**, 9385–9388.
- Davies, D.R., Goryshin, I.Y., Reznikoff, W.S., and Rayment, I. (2000). Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **289**, 77–85.
- Dawson, A., and Finnegan, D.J. (2003). Excision of the *Drosophila* mariner transposon *Mos1*. Comparison with bacterial transposition and V(D)J recombination. *Mol. Cell* **11**, 225–235.
- Gangadharan, S., Mularoni, L., Fain-Thornton, J., Wheelan, S.J., and Craig, N.L. (2010). DNA transposon *Hermes* inserts into DNA in nucleosome-free regions in vivo. *Proc. Natl. Acad. Sci. USA* **107**, 21966–21972.
- Gao, J., and Scott, J.G. (2006). Use of quantitative real-time polymerase chain reaction to estimate the size of the house-fly *Musca domestica* genome. *Insect Mol. Biol.* **15**, 835–837.
- Grishaev, A., Guo, L., Irving, T., and Bax, A. (2010). Improved fitting of solution X-ray scattering data to macromolecular structures and structural ensembles by explicit water modeling. *J. Am. Chem. Soc.* **132**, 15484–15486.
- Grundy, G.J., Ramón-Maiques, S., Dimitriadis, E.K., Kotova, S., Biertümpfel, C., Heymann, J.B., Steven, A.C., Gellert, M., and Yang, W. (2009). Initial stages of V(D)J recombination: the organization of RAG1/2 and RSS DNA in the post-cleavage complex. *Mol. Cell* **35**, 217–227.
- Hallet, B., Rezsőházy, R., Mahillon, J., and Delcour, J. (1994). IS231A insertion specificity: consensus sequence and DNA bending at the target site. *Mol. Microbiol.* **14**, 131–139.
- Hashida, S.-N., Uchiyama, T., Martin, C., Kishima, Y., Sano, Y., and Mikami, T. (2006). The temperature-dependent change in methylation of the *Antirrhinum* transposon *Tam3* is controlled by the activity of its transposase. *Plant Cell* **18**, 104–118.
- Hehl, R., Nacken, W.K.F., Krause, A., Saedler, H., and Sommer, H. (1991). Structural analysis of *Tam3*, a transposable element from *Antirrhinum majus*, reveals homologies to the *Ac* element from maize. *Plant Mol. Biol.* **16**, 369–371.
- Hickman, A.B., Perez, Z.N., Zhou, L., Musingarimi, P., Ghirlando, R., Hinshaw, J.E., Craig, N.L., and Dyda, F. (2005). Molecular architecture of a eukaryotic DNA transposase. *Nat. Struct. Mol. Biol.* **12**, 715–721.
- Hickman, A.B., Chandler, M., and Dyda, F. (2010). Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. *Crit. Rev. Biochem. Mol. Biol.* **45**, 50–69.
- Honeybee Genome Sequencing Consortium (2006). Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**, 931–949.
- Huang, C.R.L., Burns, K.H., and Boeke, J.D. (2012). Active transposition in genomes. *Annu. Rev. Genet.* **46**, 651–675.
- Kahlon, A.S., Hice, R.H., O'Brochta, D.A., and Atkinson, P.W. (2011). DNA binding activities of the *Herves* transposase from the mosquito *Anopheles gambiae*. *Mob. DNA* **2**, 9.
- Kapitonov, V.V., and Jurka, J. (2005). RAG1 core and V(D)J recombination signal sequences were derived from *Transib* transposons. *PLoS Biol.* **3**, e181.
- Kim, Y.J., Hice, R.H., O'Brochta, D.A., and Atkinson, P.W. (2011). DNA sequence requirements for *hobo* transposable element transposition in *Drosophila melanogaster*. *Genetica* **139**, 985–997.
- Kornberg, R.D. (1977). Structure of chromatin. *Annu. Rev. Biochem.* **46**, 931–954.
- Kunze, R., and Starlinger, P. (1989). The putative transposase of transposable element *Ac* from *Zea mays* L. interacts with subterminal sequences of *Ac*. *EMBO J.* **8**, 3177–3185.
- Kunze, R., and Weil, C.F. (2002). The *hAT* and CACTA superfamilies of plant transposons. In *Mobile DNA II* (Washington, DC: ASM Press), pp. 565–610.
- Lavoie, B.D., Chan, B.S., Allison, R.G., and Chaconas, G. (1991). Structural aspects of a higher order nucleoprotein complex: induction of an altered DNA structure at the Mu-host junction of the Mu type 1 transpososome. *EMBO J.* **10**, 3051–3059.
- Li, M., Mizuuchi, M., Burke, T.R., Jr., and Craigie, R. (2006). Retroviral DNA integration: reaction pathway and critical intermediates. *EMBO J.* **25**, 1295–1304.
- Liu, D., and Crawford, N.M. (1998). Characterization of the putative transposase mRNA of *Tag1*, which is ubiquitously expressed in Arabidopsis and can be induced by Agrobacterium-mediated transformation with *dTag1* DNA. *Genetics* **149**, 693–701.
- Liu, D., Mack, A., Wang, R., Galli, M., Belk, J., Ketpura, N.I., and Crawford, N.M. (2001). Functional dissection of the *cis*-acting sequences of the Arabidopsis transposable element *Tag1* reveals dissimilar subterminal sequence and minimal spacing requirements for transposition. *Genetics* **157**, 817–830.
- Lu, C.P., Sandoval, H., Brandt, V.L., Rice, P.A., and Roth, D.B. (2006). Amino acid residues in Rag1 crucial for DNA hairpin formation. *Nat. Struct. Mol. Biol.* **13**, 1010–1015.
- Mack, A.M., and Crawford, N.M. (2001). The Arabidopsis TAG1 transposase has an N-terminal zinc finger DNA binding domain that recognizes distinct subterminal motifs. *Plant Cell* **13**, 2319–2331.
- Maertens, G.N., Hare, S., and Cherepanov, P. (2010). The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* **468**, 326–329.
- McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA* **36**, 344–355.
- Montaño, S.P., Pigli, Y.Z., and Rice, P.A. (2012). The μ transpososome structure sheds light on DDE recombinase evolution. *Nature* **491**, 413–417.
- Oh, J., Chang, K.W., and Hughes, S.H. (2006). Mutations in the U5 sequences adjacent to the primer binding site do not affect tRNA cleavage by Rous sarcoma virus RNase H but do cause aberrant integrations in vivo. *J. Virol.* **80**, 451–459.
- Pribil, P.A., and Haniford, D.B. (2003). Target DNA bending is an important specificity determinant in target site selection in Tn70 transposition. *J. Mol. Biol.* **330**, 247–259.
- Rambo, R.P., and Tainer, J.A. (2013a). Super-resolution in solution X-ray scattering and its applications to structural systems biology. *Annu. Rev. Biophys.* **42**, 415–441.
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277.
- Richardson, J.M., Colloms, S.D., Finnegan, D.J., and Walkinshaw, M.D. (2009). Molecular architecture of the *Mos1* paired-end complex: the structural basis of DNA transposition in a eukaryote. *Cell* **138**, 1096–1108.
- Rubin, E., Lithwick, G., and Levy, A.A. (2001). Structure and evolution of the *hAT* transposon superfamily. *Genetics* **158**, 949–957.
- Sarkar, A., Coates, C.J., Whyard, S., Willhoeft, U., Atkinson, P.W., and O'Brochta, D.A. (1997). The *Hermes* element from *Musca domestica* can transpose in four families of Cyclorhaphan flies. *Genetica* **99**, 15–29.
- Schatz, D.G., and Swanson, P.C. (2011). V(D)J recombination: mechanisms of initiation. *Annu. Rev. Genet.* **45**, 167–202.
- Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112–1115.
- Streck, R.D., Macgaffey, J.E., and Beckendorf, S.K. (1986). The structure of *hobo* transposable elements and their insertion sites. *EMBO J.* **5**, 3615–3623.
- Urasaki, A., Morvan, G., and Kawakami, K. (2006). Functional dissection of the *To12* transposable element identified the minimal *cis*-sequence and a highly repetitive sequence in the subterminal region essential for transposition. *Genetics* **174**, 639–649.

- VandenDriessche, T., Ivics, Z., Izsák, Z., and Chuah, M.K.L. (2009). Emerging potential of transposons for gene therapy and generation of induced pluripotent stem cells. *Blood* 114, 1461–1468.
- Warren, W.D., Atkinson, P.W., and O'Brochta, D.A. (1994). The *Hermes* transposable element from the house fly, *Musca domestica*, is a short inverted repeat-type element of the *hobo*, *Ac*, and *Tam3 (hAT)* element family. *Genet. Res.* 64, 87–97.
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhou, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982.
- Yin, F.F., Bailey, S., Innis, C.A., Ciubotaru, M., Kamtekar, S., Steitz, T.A., and Schatz, D.G. (2009). Structure of the RAG1 nonamer binding domain with DNA reveals a dimer that mediates DNA synapsis. *Nat. Struct. Mol. Biol.* 16, 499–508.
- Yuan, Y.-W., and Wessler, S.R. (2011). The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proc. Natl. Acad. Sci. USA* 108, 7884–7889.
- Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9, 40.
- Zhou, L., Mitra, R., Atkinson, P.W., Hickman, A.B., Dyda, F., and Craig, N.L. (2004). Transposition of *hAT* elements links transposable elements and V(D)J recombination. *Nature* 432, 995–1001.