10286–10301 *Nucleic Acids Research, 2018, Vol. 46, No. 19* *Published online 20 September 2018*
doi: 10.1093/nar/gky838

Structural insights into the mechanism of double strand break formation by Hermes, a *hAT* family eukaryotic DNA transposase

Alison B. Hickman[1,†], Andrea Regier Voth[1,†], Hosam Ewis[2], Xianghong Li[2], Nancy L. Craig[2] and Fred Dyda[1,*]

[1]Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA and [2]Howard Hughes Medical Institute, Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

Received July 09, 2018; Revised September 05, 2018; Editorial Decision September 06, 2018; Accepted September 07, 2018

ABSTRACT

Some DNA transposons relocate from one genomic location to another using a mechanism that involves generating double-strand breaks at their transposon ends by forming hairpins on flanking DNA. The same double-strand break mode is employed by the V(D)J recombinase at signal-end/coding-end junctions during the generation of antibody diversity. How flanking hairpins are formed during DNA transposition has remained elusive. Here, we describe several co-crystal structures of the *Hermes* transposase bound to DNA that mimics the reaction step immediately prior to hairpin formation. Our results reveal a large DNA conformational change between the initial cleavage step and subsequent hairpin formation that changes which strand is acted upon by a single active site. We observed that two factors affect the conformational change: the complement of divalent metal ions bound by the catalytically essential DDE residues, and the identity of the –2 flanking base pair. Our data also provides a mechanistic link between the efficiency of hairpin formation (an A:T basepair is favored at the –2 position) and Hermes' strong target site preference. Furthermore, we have established that the histidine residue within a conserved C/DxxH motif present in many transposase families interacts directly with the scissile phosphate, suggesting a crucial role in catalysis.

INTRODUCTION

DNA transposition is the process by which a discrete segment of DNA is either moved or copied into a new genomic location. In both prokaryotes and eukaryotes, transposition is an important mechanism that contributes to genomic change and hence to the process of evolution (1–3). Several distinct types of enzymes catalyze DNA transposition (reviewed in 4,5), one of the largest groups of which are the so-called DDE/D enzymes, named for a cluster of essential acidic residues located at the active site, that are members of the RNase H superfamily. DDE/D enzymes exhibit surprisingly diverse *modi operandi* despite sharing a structurally similar catalytic core. For some, movement is coupled to replication; others move by 'cut-and-paste' transposition (Figure 1A) and among these there are several distinct pathways by which the needed double-strand breaks (DSBs) are generated at their transposon ends (6,7).

One common mechanism by which transposases generate DSBs involves forming a DNA hairpin on one side or the other of the break (Figure 1B). This mechanism is employed by the widespread IS*4* family of prokaryotic transposons and ISs, and by eukaryotic *piggyBac* (8), *hAT* (9), *Transib* (10), and *Mutator*-like transposons (11). Deliberate DNA hairpin formation accompanying DSB formation is otherwise rare in biology, shown to occur during V(D)J recombination (12), which is responsible for the generation of antibody diversity in higher vertebrates, and the formation of terminal hairpin telomeres in certain phages, bacteria, and plasmids (13). In the case of V(D)J recombination, it is clear that the RAG1/2 enzyme involved arose from an ancient transposon of the *Transib* family (10,14,15).

Whether the hairpin that accompanies DSB formation during transposition is formed on the transposon end (Fig-

*To whom correspondence should be addressed. Tel: +1 301 402 4496; Fax: +1 301 496 0201; Email: fred.dyda@nih.gov
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
Present addresses:
Andrea Regier Voth, Frederick National Laboratory for Cancer Research, Rockville, MD 20850, USA.
Hosam Ewis, Regenxbio Inc., Rockville, MD 20850, USA.
Xianghong Li, Poseida Therapeutics, Inc., San Diego, CA 92121, USA.

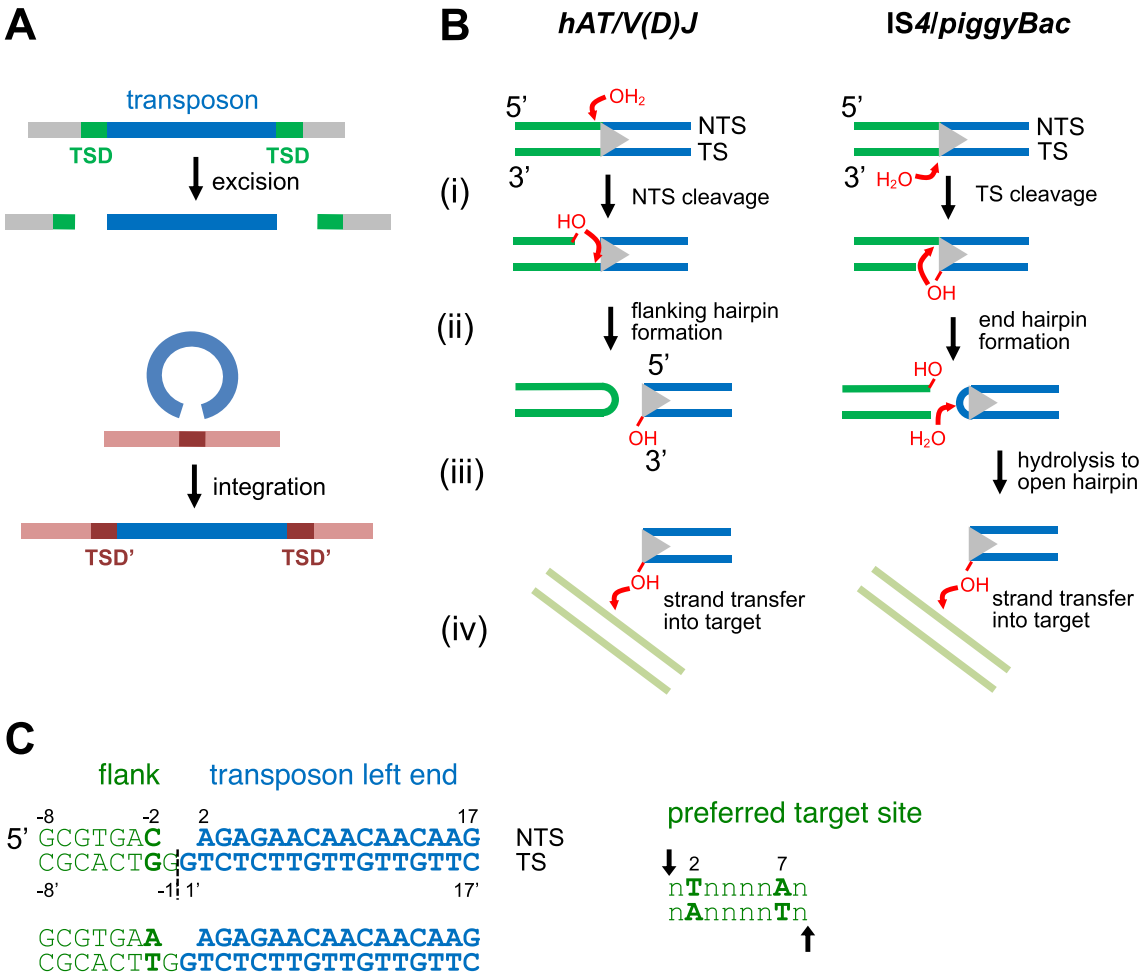

Figure 1. (**A**) Overview of cut-and-paste DNA transposition. Transposases catalyze the excision of a transposon (in blue) from one genomic location by introducing double-strand breaks at each end. Excised transposons are then integrated into a new genomic site, a process usually accompanied by the generation of short flanking target site duplications (TSDs). (**B**) Comparison of the steps comprising transposition when a hairpin is formed on flanking DNA (left, *hAT* transposons or during V(D)J recombination) or on the transposon end (right, IS*4* family of insertion sequences and eukaryotic *piggyBac* element). In step (i), *hAT* transposases cleave the non-transferred strand (NTS) one nt into the flank whereas the V(D)J recombinase cleaves precisely at the coding end-recombination signal sequence junction; TS: transferred strand. Throughout, flanking DNA is shown in green and transposon DNA in blue. (**C**) Oligonucleotides used for crystallography, 'DNA(C:G)' on top and 'DNA(A:T)' on the bottom. Also shown is the preferred target site for integration of *Hermes*.

ure 1B, right) or on flanking DNA (Figure 1B, left) has important mechanistic implications. In both cases, hairpin formation involves two sequential nucleophilic S_N2 reactions: water first attacks a DNA backbone phosphate group to cleave the DNA strand leaving a free 3′-OH group and a 5′-phosphate (step i, Figure 1B), and then the new 3′-OH group attacks a phosphate on the opposite strand (step ii). This second reaction seals the DNA end as a hairpin and in the process generates another 3′-OH group. When the hairpin is formed on flanking DNA, the new 3′-OH group is immediately available for use as the nucleophile in the subsequent 'paste' step of transposition (also known as strand transfer or integration, where the joined DNA strand is called the 'transferred strand', TS; step iv, left). However, when the hairpin is on the transposon end, a further hydrolysis reaction is needed to open up the hairpin and liberate the 3′-OH group needed for integration (steps iii and iv, right).

How transposases catalyze not just one but a series of coordinated reactions at their transposon ends has intrigued researchers ever since their discovery. When a hairpin is formed on the transposon end, all of the reactions can be catalyzed by a single active site (16), with the transferred strand - in theory - held steadfastly in the active site during the entire process. This is because the 3′-OH produced in the first nicking reaction is the nucleophile of the hairpin-forming transesterification reaction and is subsequently regenerated as the nucleophile for the integration reaction after hairpin opening. As all of the reactions comprising transposition occur at the same end of the transferred strand, in principle, once it has been bound in the active site, the transferred strand can remain there until the final integrated product is formed (17). Hairpin formation on transposon ends has been biochemically well-characterized (18), and important mechanistic details have also been pro-

vided by structures of the Tn*5* DNA transposase, an IS*4* family member, bound to DNA (19, reviewed in 20).

In contrast, how DSBs are generated when flanking hairpins are formed during transposition is less well understood, and the reaction is more conceptually problematic. In these cases, the first nick is on the non-transferred strand (NTS) and the liberated 3′-OH group on flanking DNA is the nucleophile for hairpin formation, but thereafter the hairpin has no further role in transposition (step ii, Figure 1B). Rather, the 3′-OH on the transferred strand is used for integration, and the transposase active site must therefore catalyze a 3′-OH attack from each strand, implying the need for significant rearrangement of the protein or the DNA, or the deployment of two active sites at each transposon end to process the opposing strands. Although structural work with the *Hermes* transposase (21) and, more recently, with the mechanistically-related RAG1/2 recombinase bound to its recombination signal sequences (22–24) has clearly shown that only one active site is involved, how hairpins are formed remains unclear. It is also not understood how the first nick on the NTS is directed: *hAT* transposases cleave the NTS one nt into the flank whereas the RAG1/2 recombinase cleaves precisely at the coding end-recombination signal sequence junction. Here, to understand the hairpin formation process, we have assembled protein-DNA complexes of the *Hermes* transposase, a *hAT* family member from the house fly *Musca domestica*, and used X-ray crystallography to visualize the conformation changes that occur. We find that the active site arrangement changes dramatically as a function of bound divalent metal ion cofactors and, surprisingly, as a function of the identity of a specific flanking base pair. The structures also reveal that a conserved histidine residue, part of a conserved C/DxxH motif present in several eukaryotic DNA transposon superfamilies (25), forms a crucial part of the enzyme active site by directly coordinating the scissile phosphate. Finally, the structure of Hermes bound to DNA and two catalytically essential metal ions suggests how active site back-and-forth usage is coordinated to catalyze the sequential steps that comprise DNA transposition.

MATERIALS AND METHODS

Protein production and purification

The HermesΔ construct was generated from Hermes 79–612 in pET-15b (26) by introducing a deletion spanning residues 497–516 and adding the point mutation C519S to increase solubility. His-tagged protein was expressed in E. coli RosettaBlue (DE3) cells grown at 37°C to $OD_{600} = 0.6$ and induced with the addition of 0.3 mM IPTG at 19°C overnight. Cells were harvested ∼16 h after induction and the pellet was resuspended in 25 mM Tris pH 7.5, 0.5 M NaCl, 5 mM Imidazole, 0.4 mM TCEP, and 1X Proteoblock (Fermentas), flash frozen in liquid nitrogen, and stored at −80°C.

Four liters of cell pellets were thawed at room temperature and lysed by sonication at 0°C. The crude lysate was clarified by centrifugation at 44 000 × g for 45 min and loaded onto a column of ∼15 ml Ni-NTA resin (Novagen/Qiagen). The column was washed extensively with the loading buffer and then with 35 mM Tris pH 8.0, 0.5 M NaCl, 42 mM Imidazole, 0.3 mM TCEP, and 1% glycerol. HermesΔ was eluted from the column with a gradient of 42–400 mM Imidazole and fractions containing HermesΔ (as seen by SDS-PAGE) were combined. To remove the polyhistidine tag, 300 units of thrombin (Sigma) were added to the combined fractions and the sample was dialyzed overnight at 4°C against a buffer of 25 mM Tris pH 7.5, 1 mM EDTA, 0.5 M NaCl, 0.3 mM TCEP and 10% glycerol. Thrombin was removed by passage over a 0.5 ml benzamidine sepharose 6B column (Amersham Biosciences) and the recovered protein was concentrated to <5 ml before size exclusion chromatography on two 300 mm Bio-Sep-SEC-S3000 columns (Phenomenex) in series in 25 mM HEPES pH 7.3, 1.0 M NaCl and 0.3 mM TCEP. The fractions were then analyzed by SDS-PAGE and those containing only full-length HermesΔ were combined and concentrated to >10 mg/ml.

Protein–DNA co-crystallization and data collection

Gel-purified oligonucleotides purchased from IDT were resuspended in TE and annealed together at a 1:1:1 ratio by heating to 95°C for 15 min and then slow-cooling. The annealed DNA substrate was added to ≥8 mg/ml HermesΔ protein at a 1:1.05 ratio and the complex was dialyzed at 4°C in two stages against 25 mM HEPES pH 7.5, 0.3 mM TCEP and 0.5 M NaCl (with or without 5 mM $MgCl_2$), followed by the same buffer with 250 mM NaCl. The dialyzed complex was spun at 15 000 × g before being placed in crystallization trays.

Initial crystals were obtained from the following conditions: The protein:DNA complex was 10 mg/ml HermesΔ + PAGE-purified DNA(C:G) substrate in complex buffer with 5 mM $MgCl_2$. The 1 ml well solution was 50 mM MES pH 5.9, 50 mM sodium acetate, 15% PEG MME 2000, 12.5 mM HEPES pH 7.5, 250 mM NaCl, 0.1 mM TCEP and 2.5 mM $MgCl_2$. Hanging drops were set up at 4°C with 6 μl complex, 1 μl 100 mM EDTA, and 4 μl well solution. Crystals grew at 4°C in 1 week.

Subsequent crystals were obtained by microseeding using the Seed Bead kit (Hampton Research) with a seed stabilizing solution of 15% PEG MME 2000, 20 mM MES pH 6.0, 20 mM sodium acetate, 20 mM HEPES pH 7.5, 230 mM NaCl, and 0.15 mM TCEP. Hanging drops were set up at 4°C with 4 μl of 4–8 mg/ml HermesΔ + DNA in complex buffer without $MgCl_2$ plus 3 μl seed solution. Well solutions were 1 ml of 50 mM MES pH 6.5, 50 mM sodium acetate, 5–15% PEG MME 2000, 12.5 mM HEPES pH 7.5, 250 mM NaCl and 0.1 mM TCEP. Crystals grew to usable size in 4–6 weeks at 4°C. For the crystal structures with metal added, 0.3 μl 200 mM $MnCl_2$ or $CaCl_2$ was added to the drop for a final concentration of 8 mM metal ∼48 h before the crystals were frozen. For the crystal structure of HermesΔ and DNA(A:T) prior to soaking with metal ions, $MgCl_2$ was omitted throughout, and crystals were obtained after more than three rounds of microseeding in the absence of $MgCl_2$.

Crystals were cryoprotected in a solution of 30 mM MES pH 6.8, 30 mM sodium acetate, 30 mM HEPES pH 7.5, 245 mM NaCl, 0.15 mM TCEP, 5–15% PEG MME 2000, 15% ethylene glycol, and 5 μM DNA substrate (with or without 8 mM $MnCl_2$ or $CaCl_2$) for 20–120 s and flash-frozen into

liquid nitrogen for data collection. Data was collected at the Southeast Regional Collective Access Team beamline ID22 of the Advanced Photon Source on a MAR300 CCD detector. All data was collected at 95 K and integrated and scaled using XDS.

Structure determination and refinement

The structures were solved with molecular replacement by AMoRe, initially using the search model of a monomer from the octamer structure (PDB ID: 4D1Q) divided into two domains—amino acids 1–266 and 427–609 as one domain and amino acids 267–426 as the second domain. Subsequent structures were solved with the first HermesΔ structure as the search model. For refinement, several rounds of Cartesian simulated annealing, positional and restrained *B*-factor refinement cycles in Phenix were alternated with manual model building in O.

Enzymatic assays

In vitro hairpin formation with oligonucleotides (Figures 5B, 6A, Supplementary Figure S5). Full-length Hermes was purified as previously described (21). Plasmids encoding mutant proteins H268A and H268Q were generated by the QuikChange method, and proteins were expressed and purified as for the wild-type. Assays were carried out in 25 mM Tris pH 7.5, 150 mM NaCl, 10 mM DTT, 0.1 mg/ml BSA, and either 10 mM $MgCl_2$ or 0.1 mM $MnCl_2$, by the addition of protein to a final concentration of 3.1 μM in the presence of 0.92 μM DNA substrate for the indicated times at the temperatures shown. Samples were then mixed 1:1 with formamide gel-loading buffer, and loaded onto a 20% acrylamide TBE–urea gel run in 1X TBE at 300 V. DNA was detected by SYBR Gold staining.

In vitro cleavage assays with pRX1-Her (Figure 6B). The pRX1-Her LE/RE30 mini-transposon plasmid was used as template for *Pfu* polymerase to amplify and introduce the –2 basepair (bp) mutations to target site duplication flanking the *Hermes* LE/RE30 mini-transposon. The resulting 1540 bp PCR product was gel-purified, digested with *Xho*I, and ligated into *Xho*I-digested and dephosphorylated pRX1 plasmid. Ligation reactions were transformed into *Escherichia coli* DH5α, and subsequently selected on LB/Kanamycin plates. The resulting plasmids containing the *Hermes* LE/RE30 mini-transposon cassette with flanking target site –2 bp mutations from positive clones were further subjected to *Nco*I digestion, that would result in a diagnostic band of 713 bp. This step was required in order to select the plasmids in which the Hermes LE/RE30 mini-transposon had identical clonal directionality to pRX1-HerLE/RE30 mini-transposon containing wild type flanking sequences. The sequences of the selected plasmids were subsequently verified using sequencing primers that annealed to the pRX1-*URA*3 gene.

Reactions were carried out as previously described (21) with slight modifications. Briefly, individual reactions containing plasmid at 100 ng/μl were carried out in 25 mM HEPES pH 7.5, 200 mM KCl, 10 mM DTT, 5 mM $MgCl_2$ and 0.1 mg/ml BSA. Full-length Hermes at a final concentration of 0.06 μg in 100 μl was added, and the reaction incubated for 60 min at 15°C. Reactions were stopped by adding SDS/EDTA to a final concentration of 0.5% SDS and 25 mM EDTA, and incubated at 65°C for 10 min. Reactions were then incubated for 60 min at 37°C with 200 μg/ml proteinase K. Plasmid DNA was subsequently extracted, resuspended in 20 μl of 1X Smart-Cut buffer (NEB), and 1 μl of *Pml*I enzyme was added to each sample. Samples were then incubated for 60 min at 37°C, run on a 0.7% TBE-agarose gel overnight at 50 V in 0.5X TBE buffer, and visualized using ethidium bromide.

In vitro strand cleavage reactions (Supplementary Figure S4). The assay was performed essentially as previously described (21). Briefly, cleavage was assessed using a 71 bp *Hermes*-LE fragment containing 60 bp of *Hermes*-LE + 11 bp flanking DNA. The DNA was radiolabelled at the 5′ end of one strand with γ-^{32}P-ATP. 3.5 pmole of *Hermes* transposase was incubated with 0.5 pmole of radiolabelled *Hermes* LE DNA in 25 mM MOPS, pH 7.6, 2% (v/v) glycerol, 100 mM NaCl, 2 mM DTT, 1 mM $MgCl_2$, 0.01% BSA in a final volume of 10 μl at 30°C for 2 h. Reactions were stopped by adding EDTA to 40 mM and incubated at 37°C for 2 h, then spin-dried. The samples were resuspended in STOP solution (USB Corp.): 8M urea (1.8:1), and subsequently run on a 16% denaturing acrylamide gel.

In vitro hairpin formation using pre-nicked oligonucleotides (Supplementary Figure S4). The assay was performed as previously described (21). Briefly, an 11 bp oligonucleotide corresponding to the top-flanking donor-site DNA was radiolabelled at its 5′ end with γ-^{32}P-ATP, and annealed with equimolar amounts of a 61 nt oligonucleotide corresponding to 60 nt of the top strand of the *Hermes* LE + 1 nt of top flanking DNA and a 71 nt oligonucleotide containing 11 nt of bottom-strand flanking DNA + 60 nt of bottom strand of *Hermes* LE. This was then used as the substrate in hairpin formation reactions as described above.

RESULTS

Structure determination

Hermes is a four-domain protein consisting of an N-terminal BED domain, a dimerization domain involved in binding terminal inverted repeats (TIRs) at the transposon ends, a catalytic DDE domain that has three essential carboxylates (D180, D248, E572), and a large α-helical 'insertion' domain that interrupts the fold of the catalytic domain (21,27). It assembles as an ∼600 kDa octamer that is required for transposition *in vivo* (21). Structural information is available for the N-terminally truncated $Hermes_{79\text{-}612}$ octamer complexed with its TIRs (21), and for the work here we combined the N-terminal truncation with the deletion of residues 497–516 ($Hermes_{79\text{-}612\Delta 497\text{-}516}$; hereafter 'HermesΔ'). This protein is exclusively dimeric, proved tractable to structural studies, and remains active for all the steps of transposition *in vitro* (21).

To obtain diffracting crystals, we formed complexes with two different oligonucleotides (Figure 1C) that place a 3′-OH on flanking DNA at the correct position to represent the first cleavage product; this 3′-OH is also the nucleophile that attacks the TS to form the flanking hairpin.

Both oligonucleotides represent the 17 base pair (bp) *Hermes* TIR joined to 8 bp of flanking DNA with a two nucleotide (nt) gap on the top strand (necessary to generate useful crystals) spanning the TIR/flank junction; they differ only in a single flanking bp.

In contrast to many transposons, *Hermes* integration displays a strong target site preference, corresponding to a T nt at the 2nd position and an A nt at the 7th position (28, 29; numbering corresponds to the plus strand sequence, Figure 1C). This results in flanking target site duplications (TSDs) that preferentially have A at the –2 and T at the –7 position on the left end (LE) plus strand (and A at the –2 and T at the –7 position on the right end (RE) minus strand as shown in Supplementary Figure S1; the first bp of each transposon end is designated 1 and the first bp of flanking DNA as –1). As we wondered if the –2 bp might influence hairpin formation, we solved the structures of HermesΔ bound to oligonucleotides containing either a C:G or A:T bp ('DNA(C:G)' and 'DNA(A:T)', respectively) at the –2 bp position, both before and after soaking crystals with divalent metal ions. Crystals of the four complexes described here diffracted to between 2.85 and 3.2 Å resolution in the various structures, and structures were solved by molecular replacement (Table 1; sample electron density is shown in Supplementary Figure S2).

Overall architecture of HermesΔ/pre-cleaved transposon end complexes

With the exception of the active site regions, the overall structures of the HermesΔ/DNA complexes are very similar. In all cases, the HermesΔ dimer binds two oligonucleotides in an approximately X-shape (Figure 2), with the two flanking segments (shown in light green; TIRs are in blue) bound in a large cavity. In the previously reported octamer/TIR complex structure, this cavity lined with basic residues was unoccupied by nucleic acid, exposed to solvent, and was predicted to be the target DNA binding site (21). The two active sites are located towards the peripheries of the cavity, buried beneath DNA.

Although each bound DNA oligonucleotide is almost straight with ends that are B-form DNA, the TIR and the flank are significantly twisted around the helical axis relative to each other in the direction of underwinding (Figure 2, inset). Thus, if the TIR were extended toward the flank as B-form DNA, the modeled extension would not overlap with the observed strands of the flank due to the twist. Curiously, in such a hypothetical ideal B-form extension of the TIR, the scissile phosphate at the site of the first cleavage step would face inappropriately outward into solvent, rather than inward toward the active site, suggesting the need to overcome a substantial structural hurdle before transposition can be initiated (24).

The extent of interaction between the TIR and the transposase differs in the region of the active site in the various structures, with the differences limited to bp 1–4. Beyond this, in all of the structures, the mode of TIR binding from bp 5–11 is essentially the same, and from then on to bp 17, the protein does not contact DNA. The more distal interactions are essentially identical to those previously seen in the octamer/TIR complex (21): the TIR is bound by interactions to the major groove dominated by R149 and by several contacts to the adjacent minor groove via residues contributed by the second half of the RNase H-like domain. When the structures are compared to that of the octamer bound to cleaved TIRs, the main difference is a slight inward twist of the insertion domains towards the flank-binding cavity (Supplementary Figure S3).

Interactions between the transposase and flanking DNA beyond bp –2 are relatively sparse, non-specific, and involve only backbone phosphates and the side chains of R218, R249, and K255, all part of the DDE catalytic domain. In the structure of the related RAG1/2 recombinase bound to DNA, the coding end is contacted far more intimately by RAG1 close to the cleavage site and also by RAG2 more distally, with interactions observed as far as 12–14 bp from the recombination signal sequences (RSS)-coding end boundary (22,23). As the Hermes structures suggest that the flanking DNA binding mode most likely also represents that of target DNA, the relative orientation of the two flanking DNA segments indicate that bound target DNA must be strongly bent. This would be consistent with the bent target DNA conformations observed in the structures of other transpososomes and intasomes (30–33).

Structure of HermesΔ complexed with DNA(C:G)

In the complex formed between HermesΔ and DNA(C:G) in the absence of metal ions (Figure 3A), the dimer occupies the crystallographic asymmetric unit and its monomers are related by a near-perfect two-fold non-crystallographic symmetry. Although the 3′-OH group of nucleotide C-2 (i.e. the product of NTS cleavage; shown as a red sphere) is in the enzyme active site, less than 4 Å away from the carboxylates of D180 and D248 of the DDE motif, the complex is clearly trapped in an inactive state for hairpin formation. Most notably, the scissile phosphate on the TS (highlighted in orange) and the nucleophilic 3′-OH group that will attack it are separated by over 13 Å and the path between them is blocked by the obstacle presented by the –2 C:G bp (highlighted in bright green) which retains Watson–Crick hydrogen bonding. Furthermore, as previously observed in the DNA-free $Hermes_{79\text{-}612}$ structure (27), the third catalytic residue of the DDE triad, E572, forms an ionic bond with R318 that in turn is stacked with the indole ring of W319. Finally, the α-helix that bears E572 is not engaged with the significantly widened minor groove of the DNA close to the transposon tip (Figure 3E, upper left). This is in contrast to synaptic complexes of the Tn*5* and *Mos1* transposases bound to their transposon ends, in which this helix is observed to dock snuggly into the minor groove of the transposon ends (19, 34; Figure 3E, bottom).

Structure of HermesΔ complexed with DNA(C:G) after the addition of divalent metal ions

To evaluate the effect of adding metal ions 'in-crystallo', we next soaked pre-formed crystals of HermesΔ/DNA(C:G) with different divalent metal ions over a range of soaking concentrations and times. In most cases, although there was difference electron density at the active sites of the resulting structures, the overall structures were essentially identical to that of the metal ion-free complex. However, in a few

Table 1. Crystallographic statistics

	(C:G) - M^{2+}	(C:G) + Ca^{2+}	(A:T) + Mn^{2+}	(A:T) - M^{2+}
PDB	6DWZ	6DWY	6DWW	6DX0
Crystal properties				
Space group	$P2_12_12_1$	$C222_1$	$C222_1$	$C222_1$
Unit cell: *a*, *b*, *c* (Å)	96.38	121.38	120.48	121.60
	121.11	135.35	136.46	135.50
	132.19	103.21	104.61	102.85
Unit cell: α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Data collection				
Wavelength (Å)	1.5498	1.5498	1.5498	1.00
Resolution range (Å)[a]	30–3.2 (3.24–3.15)	30–3.2 (3.27–3.18)	30–2.85 (2.92–2.85)	30–2.9 (2.98–2.90)
Total reflections	667 115	113 747	149 370	124 838
Unique reflections[a]	34 628 (2288)	15 708 (1113)	20 430 (1476)	17 367 (1395)
Multiplicity[a]	19.26 (19.77)	7.24 (7.22)	7.31 (7.42)	7.19 (7.57)
Completeness (%)[a]	99.1 (98.0)	99.7 (99.0)	99.7 (99.7)	90.6 (99.9%)
R-merge (%)[a]	12.4 (52.1)	10.8 (73.6)	6.6 (61.8)	7.7 (101.6)
R-meas (%)[a]	12.8 (53.5)	11.3 (79.3)	7.1 (66.4)	8.4 (109.1)
$I/\sigma(I)$[a]	18.6 (6.54)	13.6 (3.22)	20.4 (3.62)	14.2 (2.11)
Wilson *B*-factor	81.9	80.0	67.4	79.8
Refinement				
R-work	0.202	0.259	0.199	0.225
R-free	0.298	0.295	0.264	0.305
Number of non-hydrogen atoms	9578	4867	4943	4867
Protein residues	483	483	483	483
RMS (bonds)	0.017	0.005	0.004	0.008
RMS (angles)	1.734	0.747	0.644	1.001
Ramachandran favoured (%)	87.1	92.3	94.0	88.1
Ramachandran outliers (%)	1.4	0.4	1.2	2.1
Clashscore	28	12	6	15
Average *B*-factor	84	95	66	103

[a]Statistics for the highest-resolution shell are shown in parentheses.

cases we observed a striking conformational change in the proximity of the active site affecting both the transposase and bound DNA (Figure 3B; Movie 1). In the highest resolution structure (3.2 Å) obtained with crystals soaked in 8 mM $CaCl_2$ for 48 h at 4°C, the conformational change was accompanied by a change in the space group symmetry of the crystals ($P2_12_12_1 \rightarrow C222_1$) and the previously observed molecular non-crystallographic axis had become a crystallographic two-fold axis (Table 1).

Relative to the metal-free complex, a single Ca^{2+} ion (shown in green) is bound by two carboxylates of the DDE triad, and the scissile phosphate of the TS has moved ∼13 Å toward the active site, placing it close to the Ca^{2+} ion, with a non-bridging oxygen within chemical bonding distance of the Ca^{2+} (3.0 Å) (Figures 3B and 4A). The 3′-OH group of nucleotide C-2 is only 3.0 Å away from the phosphorus atom of the scissile phosphate, approximately in the correct position for an inline nucleophilic attack to form a hairpin.

Other substantial structural changes accompany the active site rearrangements relative to the metal-free complex (Figure 3B). To allow the movement of the scissile phosphate, the –2 bp has broken apart, with G'-2 of the flank rotated out of the helical axis and occupying a pocket defined by the reconfigured bottom strand and the first β-strand of the catalytic domain just downstream of D180 of the DDE motif. There is also a 5–6 Å shift of the TIR toward the active site DDE residues, mainly affecting the 6 bp of the TIR closest to the cleavage site. This motion brings the TS into the active site and results in the helix bearing E572 now engaging with the TIR, running parallel to its minor groove and forming a number of new interactions (similar to that shown in Figure 3E, top right).

As a consequence of the TIR shift, W319 now stacks against G'1, the first nt of the TS in the transposon that will ultimately supply the 3′-OH for the integration step; G'1 in turn stacks against T'2. R318, previously stacked against W318, moves significantly (with C_z of the guanadium group moving ∼8 Å), thereby breaking the salt bridge to E572. In its new position, R318 binds a phosphate group of the TS, 3 nt from the scissile phosphate. In contrast to the metal-free structure in which neither the transposon end nor transposase are appropriately positioned for the next step of hairpin formation, these changes and rearrangements recapitulate those for $Hermes_{79\text{-}612}$ bound to its cleaved transposon ends (21). Another notable observed change is that H268, part of the conserved CxxH motif of the *hAT* transposase family (35), is now within H-bonding distance (3.6 Å) of a non-bridging oxygen of the scissile phosphate.

The new active site arrangement induced by calcium ion binding likely represents a close approximation to an active complex for hairpin formation as suggested by the proximity and relative orientation of the nucleophilic 3′-OH on flanking DNA and the scissile phosphate. At the current crystallographic resolution, we note that two slightly different models are consistent with the electron density: one in

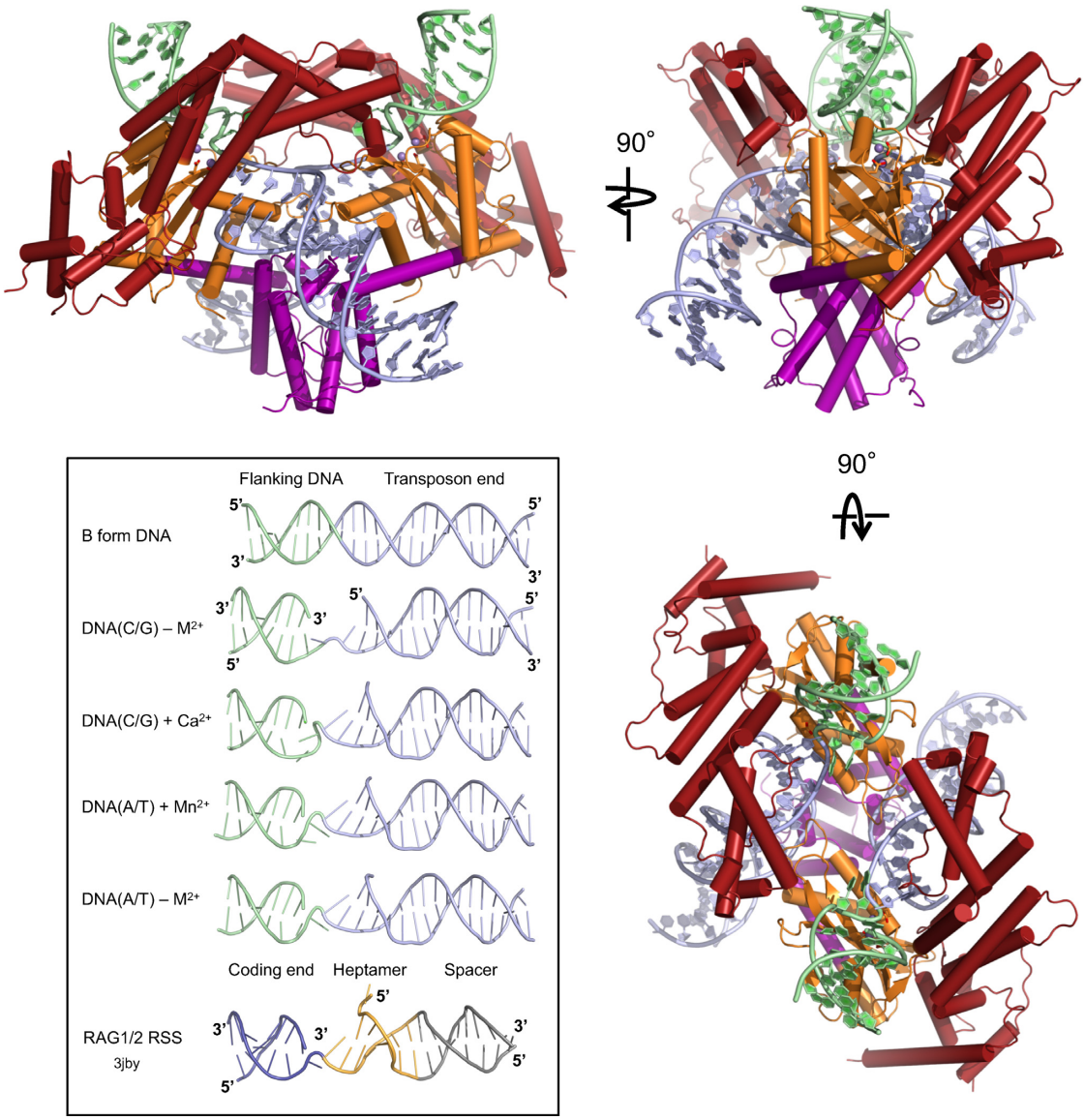

Figure 2. Structure of Hermes/DNA complex, here the DNA(A:T) complex bound to Mn^{2+}. Hermes coloring corresponds to its domains: dimerization domain in purple, insertion domain in red, and the RNase H-like catalytic domain in orange. Inset: Comparison of DNA conformations observed in four Hermes-DNA structures prior to hairpin formation to (top) ideal B-form DNA of the same bp length and (bottom) the nicked paired-end complex of RAG1/2 (22). In all cases, the trajectory of the free 5′-end of the transposon end is away from the active site.

which the 3′-OH group is poised for nucleophilic attack as shown in Figure 3B, and another in which hairpin formation has occurred and the chirality of the phosphate group is inverted. Although the two possible configurations are crystallographically indistinguishable, it seems unlikely that the hairpin has formed under the Ca^{2+} conditions in the crystals. Ca^{2+} generally does not support the chemistry of RNase H-like active sites (36)—although exceptions have been noted (37)—and does not catalyze Hermes hairpin formation *in vitro* (data not shown). In addition to the shortcomings of Ca^{2+} as an chemical substitute for Mg^{2+} in two-metal ion nucleases, the observation that the Hermes active site binds only one Ca^{2+} ion may also contribute to its inactivity.

Taken collectively, the observed conformational changes result in a net gain of stacking between W319 and G'1 and in the stabilizing electrostatic interaction between the metal ion and the scissile phosphate. There is also a new ionic interaction between R318 and a TS phosphate and there are new base-specific interactions between R573 and S576 and the TIR. At the same time, the C-2:G'-2 bp has been lost, as has the R318:E572 salt bridge and the stacking between R318 and W319. Nevertheless, the gains appear to outweigh the losses, and suggests that the energy difference between an inactive active site arrangement and one that appears close to optimal is slight, and that metal ion binding is one factor that can tip the balance toward an active state. *In vivo*, where metal ions are always present, it seems unlikely that the pre-conformational change state would exist for long.

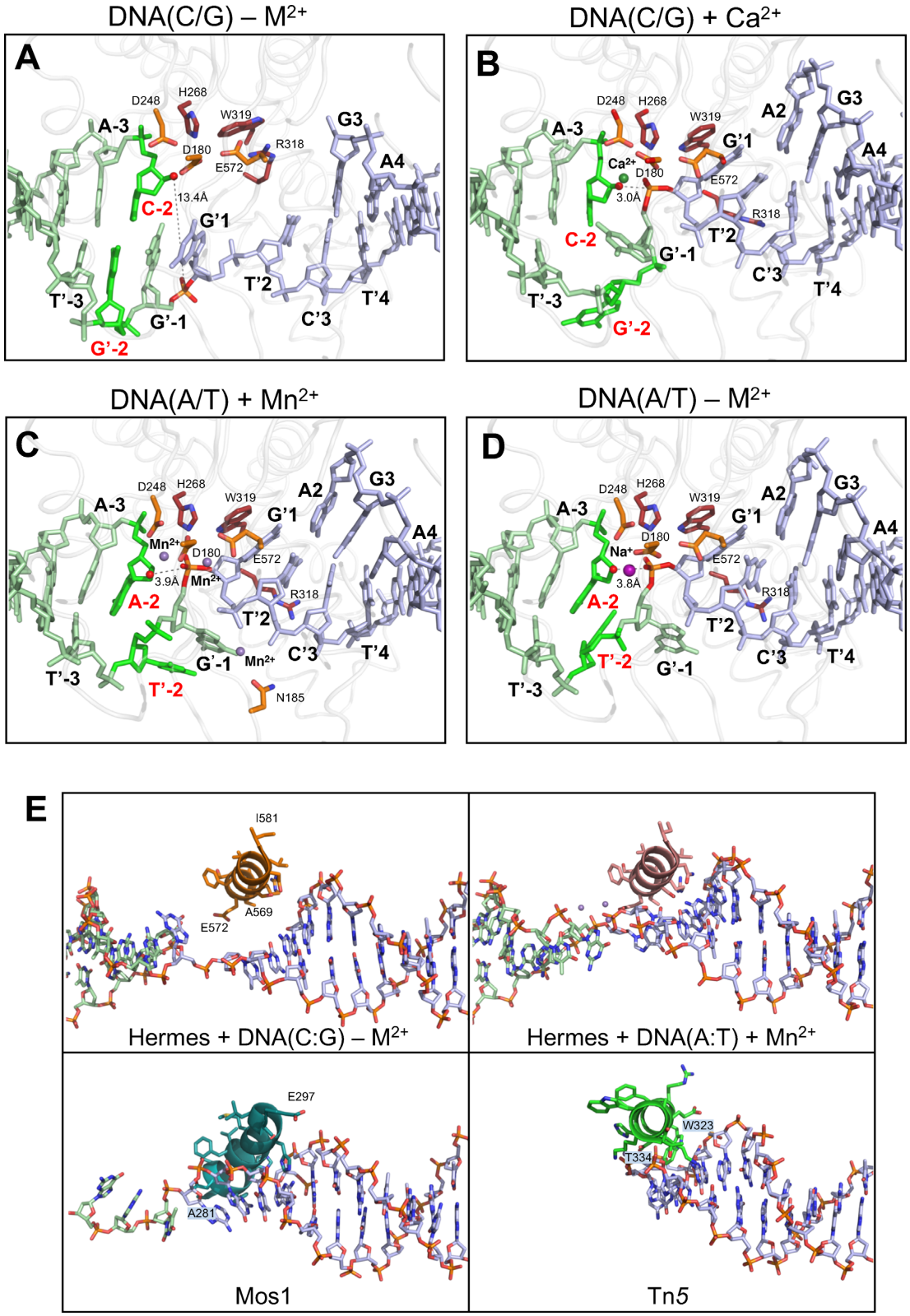

Figure 3. (**A–D**) Comparison of active site regions of four Hermes-DNA structures prior to hairpin formation. Stick colors correspond to the protein domain and DNA colors in Figure 1. The variable bp at –2 is colored bright green, and the scissile phosphate is indicated in orange. The distance between the nucleophilic 3′-OH group and the scissile phosphate is marked. Note that in the complex of Hermes with DNA(C:G) in the absence of metal ions (A), the second nt on the NTS (i.e. A2) is disordered and not visible in the electron density. (**E**) The interaction between the α-helix bearing the third catalytic acidic residue and the transposon tip. In the DNA(C:G) complex structure in the absence of metal ions (upper left), the catalytic helix (residues 569–581) does not interact with the DNA. Once the DNA conformational change has occurred, the helix is inserted into the minor groove close to the transposon tip (as shown here for the DNA(A:T) structure). The same insertion is observed for Mos1 (4U7B; residues 281–297 are shown) and Tn5 (1MUH; residues 323–334 are shown) transposases bound to their transposon ends.

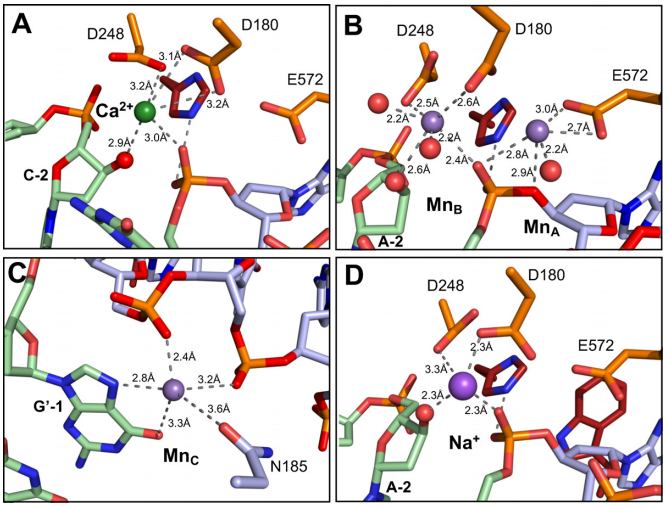

Figure 4. Metal ion binding sites. (**A**) Complex between Hermes and DNA(C:G) after soaking in Ca^{2+}. (**B**) Active site of the complex between Hermes and DNA(A:T) after soaking in Mn^{2+}. (**C**) Third Mn^{2+} ion binding site in the complex between Hermes and DNA(A:T) after soaking in Mn^{2+}. (**D**) Active site of the complex between Hermes and DNA(A:T) without metal ion soaking. Water molecules are shown as red spheres.

Structure of HermesΔ complexed with DNA(A:T) in the presence of divalent metal ions

Intrigued by the observation that *Hermes* integration displays a strong preference for an A:T base pair at the –2 position in the Hermes-generated TSDs (28,29), we next asked whether the nature of this specific bp affects the propensity for the observed conformational change in crystals under our experimental conditions. When we crystallized complexes of HermesΔ bound to DNA(A:T) and solved their X-ray structures after soaking with various metal ions, in contrast to the difficulties we experienced in inducing the conformational change with the DNA(C:G) version, we routinely observed a DNA configuration with the scissile phosphate in the active site. Upon soaking crystals with 8 mM Mn^{2+} for 48 h at 4°C, we obtained the highest resolution data set in this study, which diffracted to 2.85 Å.

The active site organization in the presence of Mn^{2+} (Figures 3C and 4B) is slightly different from that of the DNA(C:G) complex in the presence of Ca^{2+}. Most importantly, there are now two bound Mn^{2+} metal ions at the active site, 4.0 Å apart from each other in a configuration very similar to other RNase H-like active sites where the scissile phosphate is also present (38). Unexpectedly, this configuration also creates a third Mn^{2+} binding site away from the active site, formed by the constellation of G'-1, the phosphate groups of T'2 and C'3, and N185 (Figure 4C, Supplementary Figure S2); it is unlikely to be catalytically relevant. Another structural change relative to the DNA(C:G) structure in the presence of Ca^{2+} is that the unpaired T'-2 base has completely turned out of the helical axis and is now stacked against its neighbor, the unpaired G'-1 of the bottom strand.

At the active site (Figure 4B), one metal ion is coordinated by D180 and D248 of the DDE motif, a non-bridging oxygen of the scissile phosphate, the 3′-OH group of A-2 that will attack the phosphate to form the hairpin, and two water molecules. The arrangement exhibits slightly distorted octahedral coordination and the bond lengths are relatively short. The coordination sphere of the second metal ion includes D180 and E572 of the DDE motif, a non-bridging oxygen of the scissile phosphate, and a water molecule. By convention based on the active site ligands (38,39), we denote these metal ion B (Mn_B) and metal ion A (Mn_A), respectively.

In this structure, the 3′-OH group of nt A-2 on the flanking DNA that will attack the scissile phosphate to form the hairpin is 3.9 Å from the phosphorus atom and in a position turned away from that of in-line nucleophilic attack. This is a consequence of the C2′-endo sugar pucker of A-2. At this resolution, the alternate C3′-endo sugar pucker is also consistent with the electron density; this would decreases the P–O distance to 3.6 Å and orient the 3′-OH towards the scissile phosphate. As the electron density is consistent with either pucker, we cannot establish crystallographically which (or what proportion of each) is present in the crystal. However, difference electron density gave no indication that hairpins have formed. As HermesΔ is active for hairpin formation in the presence of Mn^{2+} even at 4°C (Figure 6A and data not shown), it is possible that there is an equilibrium between the two pucker forms in solution but crystallization favors the (apparently inactive) C2′-endo pucker. It may also be that the equilibrium for hairpin formation is driven forward by steps not possible in crystals, such as hairpin release.

His268 of the conserved CxxH motif is a crucial component of the active site. A notable feature of the active site arrangement observed in the Hermes complex with DNA(A:T) in the presence of Mn^{2+} is the direct interaction between the scissile phosphate and H268, which is part of the conserved CxxH motif. Direct interaction between any protein atom and the scissile phosphate has not been observed before for an RNase H-like domain, with the exception of a similarly placed His in RAG1/2-RSS complexes (22,23). The structure provides some hints as to the functional role of the CxxH motif (Figure 5A). NE2 of H268 is only 2.8 Å from a nonbridging oxygen of the scissile phosphate, indicating a likely hydrogen bond and an important role in constraining the position of the reactive phosphate group. ND1 of H268 is 3.1 Å away from the sulphur atom of C265, which is located at the positive end of the helix dipole of an α-helix that spans C265-T281 and in contact with the main chain amides of S267, H268, and S568. This extensive network of interactions suggests that the structural role of the conserved Cys is to correctly align the conserved His.

Where it has been examined for *hAT* superfamily members, the conserved histidine residue of the CxxH motif has been found to be crucial. Mutation of H268 in Hermes severely reduces catalytic activity (9). Furthermore, a domesticated *hAT* transposase-like element involved in mating type switching in *Kluyveromyces lactis*, Kat1 (40), has retained the CxxH motif of its ancestor; introducing the double mutation C402A/H405A does not affect Kat1 binding to DNA nor the initial DNA cleavage, but the mutant protein is unable to form hairpins (41). In the *Mutator* su-

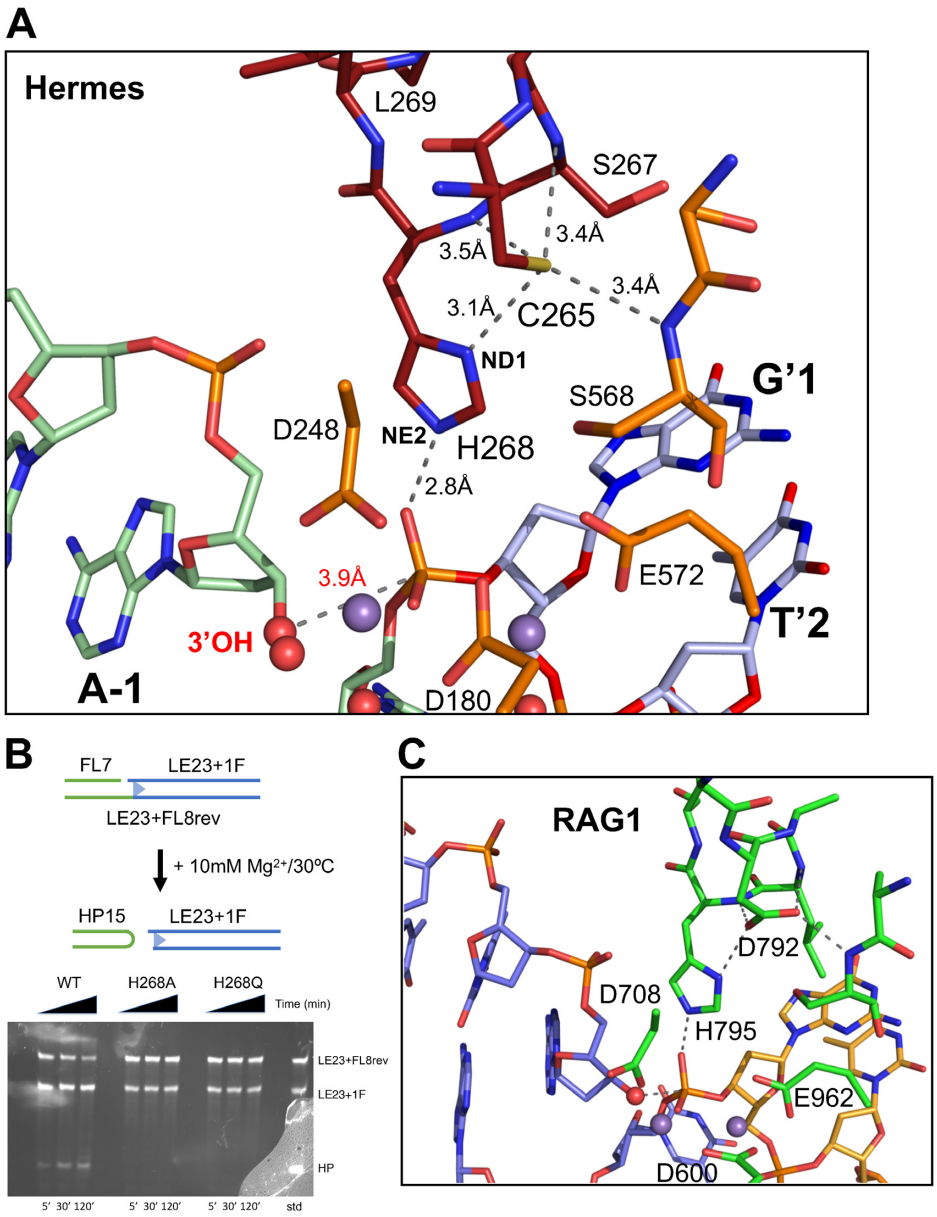

Figure 5. (**A**) A network of interactions link C265, H268, and the scissile phosphate prior to hairpin formation. Shown is HermesΔ complexed with DNA(A:T) in the presence of Mn^{2+}. (**B**) *In vitro* hairpin formation requires H268 of the conserved CxxH motif. The hairpin assay using full-length protein and the oligonucleotide shown on the left was performed as a function of time under conditions as described in Materials and Methods. Reaction products were run on a 20% acrylamide TBE-urea gel and detection was by silver staining. (**C**) The active site of RAG1 prior to hairpin formation, shown for mouse RAG1/2 (PDB ID: 5ZE1).

perfamily, the conserved histidine is part of a DxxH motif, and introduction of a H307A mutation to *Mutal*, an insect *Mutator*-like transposable element, abolishes *in vivo* excision (42).

To experimentally investigate the importance of H268, we evaluated the effect of mutating it in various *in vitro* assays using oligonucleotides. As shown in Figure 5B, both the H268A mutation and the conservative substitution H268Q abolish hairpin formation. Mutation of H268 to either Phe, Trp, or Tyr (Supplementary Figure S4) prevents not only hairpin formation but also the initial cleavage step on the top strand. Thus, consistent with its observed position in the structures, the conserved His residue is crucial for multiple steps along the transposition pathway.

Structure of HermesΔ complexed with DNA(A:T) without divalent metal ions

We also collected several data sets on the complex formed between HermesΔ and DNA(A:T) prior to soaking with metal ions. To our surprise, in all cases the scissile phosphate was located in the active site, and difference electron density indicated that an ion we tentatively identify as Na^+ was now located close to the site of the previously seen Mn_B whereas the Mn_A and Mn_C binding sites were empty (Figure 3D).

There was no stacking between T'-2 and G'-1 as was seen in the presence of soaked Mn^{2+} ions, and T'-2 adopted an intermediate configuration between that of base-pairing to A-2 and being fully turned away from the helical axis.

To verify that the observed difference density at metal site B is indeed due to Na^{+} and not contaminating divalent metal ion, we collected diffraction data at X-ray energies 12.4, 8.0 and 7.6 keV in order to maximize the anomalous signal, and calculated anomalous difference Fourier maps based on these data sets (data not shown). None of these indicated the presence of either Mn^{2+} or Ca^{2+}. Although we cannot rule out the possibility of a single Mg^{2+} ion at the active site, neither the protein nor the complex was exposed to buffer containing Mg^{2+} and there was no indication of an octahedral coordination environment, as commonly observed for Mg^{2+} (43). Thus, it appears that an A:T bp at the flanking –2 position is sufficient to bring the scissile phosphate to the active site, even in the absence of divalent metal ions.

The flanking –2 bp influences hairpin formation *in vitro*

To further probe the role of the –2 bp in flanking DNA, we first used an *in vitro* hairpin formation assay using oligonucleotides that mimic the DNA used for the structural studies with each of the possible 4 bp combinations at position –2. When we measured hairpin formation as function of time (Figure 6A), HP formation occurred most efficiently with A:T, exhibited an intermediate level with T:A or G:C, and was barely detectable with C:G. This order of effectiveness correlates precisely to the observed frequency at the –2 bp for target site selection *in vivo* (29), suggesting a direct link between the efficiency of hairpin formation and target site preference. To further probe this observation, we repeated the assay with substrates incorporating mismatches at the –2 position (Supplementary Figure S5) but did not observe any difference in hairpin formation efficiency when A:T was replaced by A:A, A:G or A:C. This suggests that the stacking interaction we observed in the DNA(A:T) + Mn^{2+} structure between T'-2 and G'-1 is unlikely to be a source of stabilization that leads to an A:T bp preference.

The preference for an A:T bp at the –2 position was confirmed in an *in vitro* plasmid cleavage assay (21). The 7.9 kb pRX1-Her plasmid containing a mini-transposon with Hermes 30-mer left ends (LE) and right ends (RE) (Figure 6B) was mutated to generate LE and RE variants with all the possible bp combinations at the –2 bp. If a DSB is formed at either end through the sequential steps of NTS nicking and then hairpin generation, distinct sets of cleavage products are observed: LE cleavage produces 4965 + 2972 bp products (following PmlI digestion) whereas RE cleavage results in 4447 bp +3490 bp fragments. Cleavage on both ends can be readily detected when an A:T is present at the –2 position on the LE and T:A at the RE (Figure 6B, lane 1). Varying the LE –2 bp in the presence of the RE T:A (lanes 2–4) leads to decreased LE cleavage activity. Similarly, varying the RE –2 bp while the LE A:T remains unchanged (lanes 5–7) leads to decreased cleavage on the RE. Thus, optimal activity is seen on the LE when A:T is the –2 bp and on the RE when the –2 bp is T:A.

DISCUSSION

In this study of flanking hairpin formation by a *hAT* transposase, we have captured several views of the Hermes–DNA complex that vary as a function of the metal ions present and the identity of the flanking –2 base pair. Notably, there is a major conformational change at the transposon end that brings the reacting groups into the active site in preparation for the hairpin-generating reaction. Once this conformational change has occurred, the active site arrangement appears compatible both with hairpin formation and the next step along the transposition pathway, strand transfer into target.

An A:T base pair at –2 on the flank is preferred for hairpin formation

It has been observed that, *in vivo*, an A:T bp is prevalent at the –2 position, reflecting the *Hermes* transposase preferred target site sequence (28,29). Our data indicate that, relative to C:G, A:T at the –2 position dramatically favors an active site arrangement in which the scissile phosphate involved in hairpin formation is in an appropriate position for nucleophilic attack by the flanking 3′-OH group. When the –2 bp is A:T, the HermesΔ/DNA complex adopts the aligned conformation whether or not the catalytic metal ions are bound at the active site: we observed the same conformation of reacting partners when two Mn^{2+} ions are bound and when the active site is occupied only by Na^{+}. However, when the –2 bp is C:G, the conformational change did not occur until divalent metal ion was added, nor did it occur consistently, suggesting a finely-tuned balance between the various interactions at the active site. The importance of flanking sequences near the cleavage site has also been observed during V(D)J recombination, where hairpin formation is enhanced by A:T-rich or disrupted base-pairing in the three terminal coding bp (44,45).

In vitro, Hermes cleavage and hairpin formation is most efficient when there is an A:T bp at the –2 position (Figures 5 and 6), although the rationale for this remains unclear. It is not as straight-forward as the ease with which an A:T or T:A bp can be disrupted relative to C:G or G:C, as this does not reflect the observed order of efficiency. Furthermore, there are no stabilizing interactions between bases displaced from the helical axis and the protein that might provide an explanation for a base preference at the –2 position. This contrasts, for example, with *mariner* transposases in which biochemical (46,47) and structural work (34) has shown that the observed preference for a flanking TA for cleavage is established by interactions between two conserved protein motifs that form a base-specific hydrophobic pocket that accommodates only a T base. The stacking interaction we observed in the DNA(A:T) + Mn^{2+} structure between T'-2 and G'-1 does not appear to be of importance either (Supplementary Figure S5), leaving the question unresolved.

However, it is probably not a coincidence that Hermes' preferred target with T:A at target bp 2 and A:T at bp 7 (Figure 1C) would generate TSD sequences with an A:T bp at the –2 position (Supplementary Figure S1). That Hermes preferentially selects a target site that facilitates subsequent mobilization makes sense from the point of view

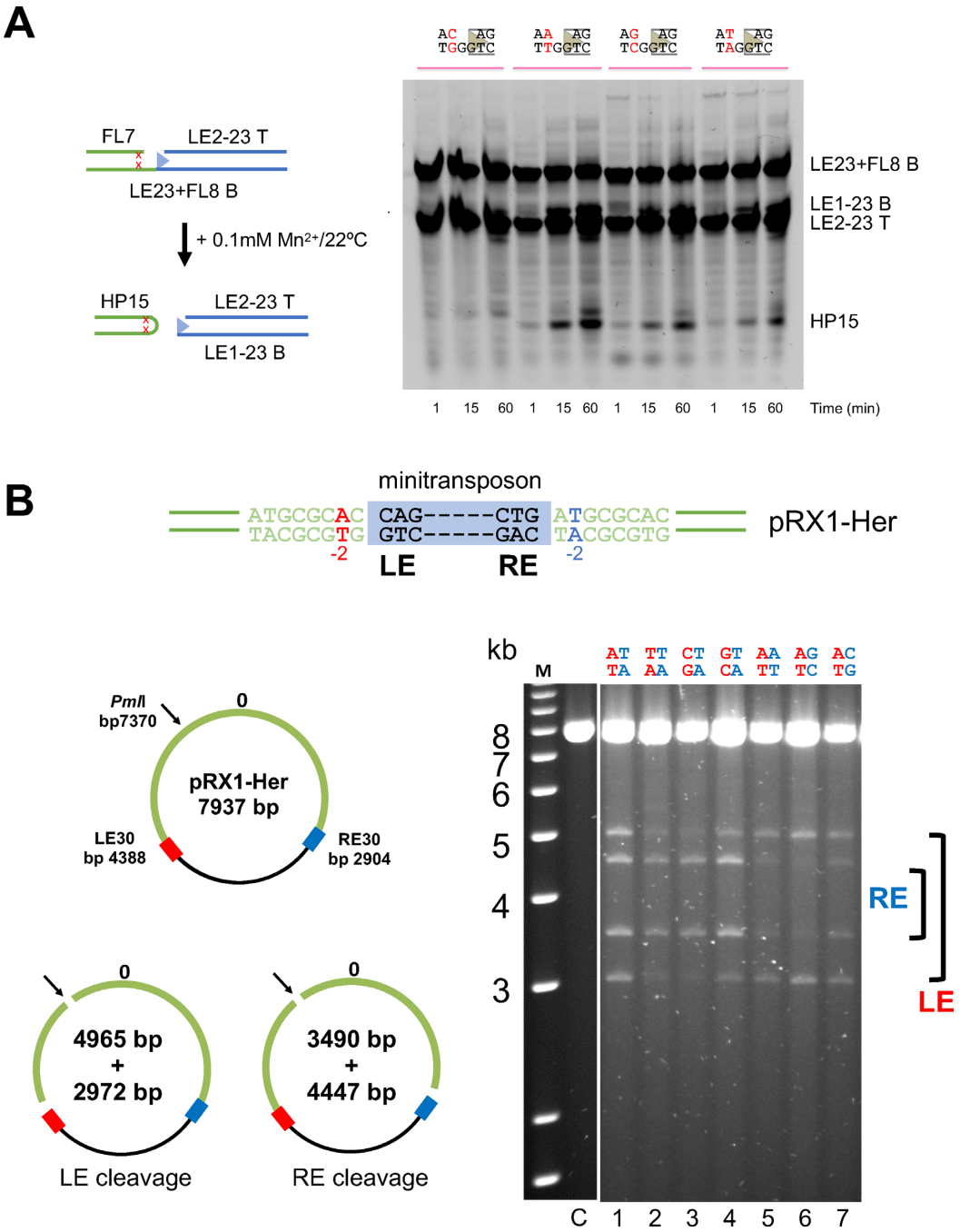

Figure 6. (**A**) *In vitro* hairpin formation as a function of the –2 bp. The reaction was run as a function of time using HermesΔ and the oligonucleotide shown on the left. Reaction products were run on a 20% acrylamide TBE-urea gel and detection was by silver staining. (**B**) *In vitro* end cleavage as a function of changing the –2 bp on the left end (LE) and right end (RE) of a Hermes minitransposon in the plasmid pRX1. In each lane, the –2 bp on the LE of the *Hermes* minitransposon is indicated at the top in red and that on the RE is indicated in blue. The assay was performed under conditions as described in Materials and Methods.

of the transposon lifestyle, as integrating into a sequence that could become a trap would hinder further mobilization. What we still do not know is what mechanism links the two observations.

One possible explanation for the observed Hermes target preference may relate to the conformation of the target DNA at the integration step. Although our attempts to obtain a co-crystal structure with target DNA have failed thus far, flanking DNA in our structures presumably occupies two ends of the target binding site with the 8 bp located in-between, and it seems likely that target DNA must bend (as modeled in 21). Quite possibly, DNA bending is generally facilitated by T:A at target position 2 and A:T at position 7, as other transposases have been reported to have a similar target site preference. For instance, the P-element transposase, which similarly generates 8 bp TSDs, exhibits

the same preference for T:A at target position 2 and A:T at position 7 (48,49). Furthermore, although *Sleeping Beauty*, like other Tc1/*mariner* elements, integrates specifically at TA dinucleotides, its consensus sequence for target site selection extends over 8 bp (50,51) and, within this, T:A is the preferred bp at position 2 and A:T at position 7 (–2 and +2, relative to the central TA). In the absence of a structure of a Hermes/target DNA complex, we cannot exclude the possibility that Hermes specifically recognizes the bp at positions 2 and 7 by some other direct mechanism.

Curiously, a 5′-nTnnnnAn-3′ target site preference is not a general feature of the *hAT* transposon superfamily, which is composed of two distinct families, the *Ac* family (of which Hermes is a member) and the *Buster* family (52). *Ac* family members show a target site preference for 5′-nTnnnnAn-3′ whereas the target site preference for *Buster* transposons is 5′-nnnTAnnn-3′ (52,53). Analysis of *hAT* transposon integration sites suggests that preferred target sequences have structural features corresponding to DNA flexibility (54), a correlation that has been reported for many other transposons (55–58) and appears broadly observed (59).

The *hAT* active site arrangement is DDHE, not just DDE

Given the conservation of the CxxH motif in the *hAT* family, the direct contact between H268 and the scissile phosphate observed here, and mutational data indicating the requirement for histidine at position 268, it is reasonable to define the *hAT* active site arrangement as DDHE, rather than DDE. The importance of a DDHE RNase H-like active site is highlighted by its conservation in other eukaryotic DNA transposases. A conserved C/DxxH motif is found in other members of the superfamily cluster that includes the *hAT* transposases, as well as the P-element, *Mutator*, *Rehavkus*, and *Kolobok* superfamilies (25). A variation, a [M/L]H motif, was identified in the superfamily cluster that comprises the *CACTA*, *Mirage*, *Chapaev* and *Transib* elements (and may be part of a longer motif, DxMH in the *CACTA* superfamily and CxLH in the *Transib* elements (10)). Furthermore, a C/DxxH motif is also conserved in several families of prokaryotic insertion sequences including IS*256* (60), IS*66* (61) and IS*L3* (A.B.H., data not shown), correlating precisely with those predicted to contain an all-α-helical insertion between the second Asp and the Glu of the catalytic triad (62). Our structures here, together with those of the structurally similar RAG1 (an enzyme that arose from an ancient *Transib* transposon), suggest that an active site His is likely playing an important functional role in all cases. Precisely defining that role and why it is conserved among a specific subset of DNA transposases but not others, is a particularly interesting avenue for further investigation.

At the current diffraction limit, it is difficult to unequivocably establish the role of the Hermes CxxH motif, in part because it is not possible to assign protonation states to C265 and H268. It seems likely that His268 helps to pull the scissile phosphate at the hairpin formation step into the active site, and that this occurs prior to binding of the two catalytically required metal ions as binding of two divalent metal ions has not been observed in crystal structures involving RNase H-like enzymes in the absence of the scissile phosphate. It is also possible that H268, with its imidazole ring located approximately midway between the two bound metal ions, may play a role in proton-transfer pathways that are required for hydrolysis and transesterification reactions.

The RAG1/2 recombinase active site is very similar to that of Hermes

The V(D)J recombinase, RAG1/2, is mechanistically related to *hAT* transposases as it generates DSBs at junctions between signal ends and coding ends of V, D, and J segments by forming hairpins on the coding ends (12). Although the overall sequence identity between RAG1 and Hermes is insignificant, they share an RNase H-like catalytic domain and a structure-based, 3D alignment can be generated (63). Remarkably, the active site arrangement, including the substrate DNA configuration just before the hairpin-generating transesterification step (22,23), is essentially identical (Figure 5C). There is also a structural equivalent of Hermes H268 within hydrogen-bonding distance of the scissile phosphate (H817 in zebrafish RAG1, H795 in mouse RAG1). Its importance has been demonstrated, as mutation of mouse RAG1 H795 leads to a defect in both nicking and hairpin formation (64), and mutation of mouse D792 (the equivalent of Hermes C265) leads to a severe defect in cleavage and impaired hairpin formation (65).

Hairpin formation in *Hermes* transposition and V(D)J recombination differ in that Hermes nicks the NTS 1 bp into the flank whereas RAG1/2 nicks precisely at the signal/coding junction. This difference implies a different final form of the hairpin: that formed by Hermes likely forms a trinucleotide loop whereas the coding hairpin formed by RAG1/2 has been shown to adopt two alternative conformations with four unpaired nucleotides in the loop (23). The likely conformation of the trinucleotide loop structure formed by Hermes—at least when still bound to protein—is suggested by the three structures here that have undergone the NTS conformational change (Figure 3B–D). In all cases, nt –2 of the top strand is retained within the hairpin stem axis whereas the other two bases of what will become the trinucleotide loop are flipped out.

The Hermes active site is compatible with alternating 'forward' and 'reverse' usage

The steps of *hAT* transposition are shown schematically in Figure 7. Details provided by the structures here, combined with information provided by the structures of the apoenzyme (27) and $Hermes_{79\text{-}612}$ bound to its cleaved TIRs (21), yield a framework for understanding how they are accomplished.

The model of alternating 'forward' and 'reverse' active site usage in transposition invokes the alternating use of metal A and metal B in activating the nucleophile for either hydrolysis or transesterification (16,17,66). The highest resolution HermesΔ complex in the presence of Mn^{2+} provides a view of the catalytically relevant arrangement poised for hairpin formation (Figure 4D; step iv in Figure 7) in which metal B participates in activating the nucleophilic 3′-OH. A backward count through the phosphoryl transfer reactions therefore suggests that—if the model holds true—metal A

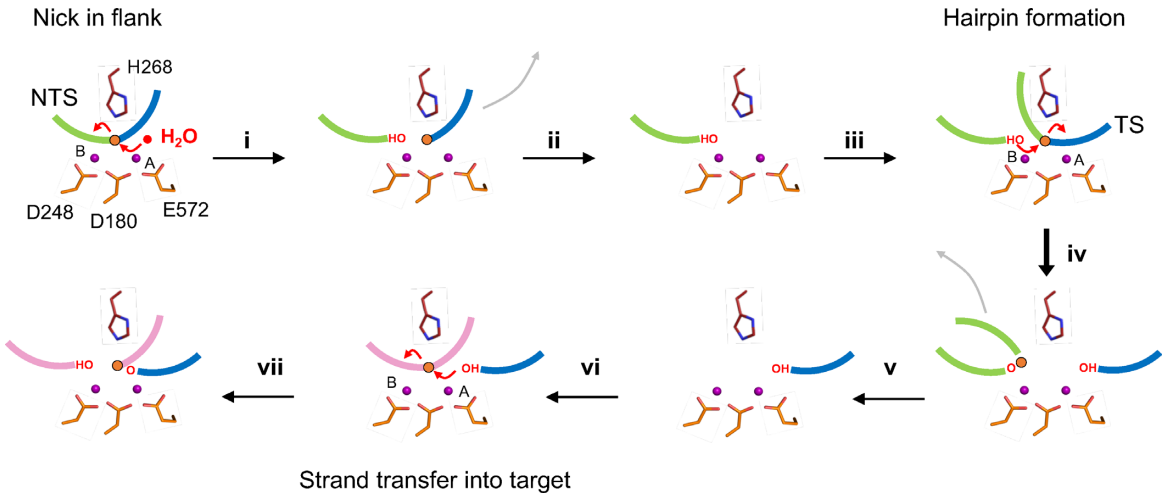

Figure 7. Proposed 'ping-pong' mechanism of Hermes transposition. The active site is used first in one configuration and then in the reverse direction, with the attacking nucleophile and the leaving group alternating sides of a centrally bound scissile phosphate group. The scissile phosphate, presumed to undergo stereochemical inversion with each step, is marked by the orange circle.

activates a water nucleophile for the first step of the reaction, which is cleavage on the top strand with the concommitant release of a 3′-OH group on flanking DNA (step i, Figure 7). In the final step (vii, Figure 7), strand transfer completes the transposition reaction, most likely with metal A activating the 3′-OH on the transposon end for nucleophilic attack on bent target DNA.

How DNA moves during *Hermes* transposition

Although the structures indicate that the same active site can be used throughout the *Hermes* transposition reaction, the various reacting partners must come and go. We do not yet know how transposition is initiated, but once the nick on the NTS has occurred, the structures are consistent with DNA relaxation as it becomes more stretched-out and underwound, a distortion that would prevent reversal or resealing of the nick. The NTS must then leave the active site to make way for the TS (ii, Figure 7), but whether the metal ions are also released is an open question. The structures suggest that most likely 2 nt on the top strand (the flanking -1 nt and the first nt on the transposon NTS) must move out of the active site to allow TS access. That non-participating bases must shift out of the way at transposon ends as the reaction continues appears to be a common theme among transposases and transposases-like enzymes: bases can be flipped out of the way, ends frayed, or the NTS actively excluded from the active site (22,67,68).

After NTS cleavage, a large conformational change of the transferred strand brings the scissile phosphate for hairpin formation to the active site (iii, Figure 7). This is accompanied by a TIR shift towards the active site, made possible by the sparseness of specific protein/DNA interactions involving the TIR end. Once the active site has been properly aligned and the hairpin has formed, it seems likely that it readily diffuses away (v, Figure 7) given the very limited number of interactions observed between flanking DNA and the transposase and consistent with the need to use the same surface cavity for target binding. The structural contrast to how extensively hairpins are bound by the RAG1/2 recombinase might reflect a need during V(D)J recombination to suppress the dangerous possibility of integration by preventing hairpin release and hence access to a possible target binding site. As we have observed the same TIR interactions for the pre-hairpin formation step as for the octamer/cleaved TIR complex in the absence of metal ions (21), the 3′-OH group on the transferred strand most likely remains tightly bound in the active site at this stage, whether or not metal ions stay bound or are released.

Finally, target binding (vi, Figure 7) to Hermes can be readily modeled (21). The observed organization of the active site after the conformational change is not only appropriate for hairpin formation (nucleophilic attack by the flanking 3′-OH group to generate an -OH leaving group on the 3′ end of the TS) but also for strand transfer (nucleophilic attack by the 3′-OH group of the TS on the scissile phosphate of a target strand to generate a strand transfer product), as shown in Supplementary Figure S3. Thus, it appears that once the hairpin formation step of the reaction has been reached, there is no need for any further major protein movements. It has been proposed that the strain on bent DNA is released upon strand transfer, thereby preventing reaction reversal (30,69).

Our structural characterization of HermesΔ with nicked DNA mimicking the first cleavage step in transposon excision has helped clarify the link between the organization of the transposase DDHE active site and its preferred integration sequence, and the ping-pong mechanism by which it coordinates the excision and integration reactions required for transposon mobilization by exchanging binding of the initial flanking DNA for a new target.

DATA AVAILABILITY

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession numbers 6DWZ (Hermes/DNA(C:G) – M^{2+}), 6DWY (Hermes/DNA(C:G)

+ Ca^{2+}), 6DWW (Hermes/DNA(A:T) + Mn^{2+}), and 6DX0 (Hermes/DNA(A:T) - M^{2+}).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Rachel Sorensen for helpful comments on the manuscript, and anonymous reviewers for their insights. N.L.C. is an investigator of the Howard Hughes Medical Institute.

FUNDING

Intramural Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health (in part). Funding for open access charge: Intramural Program of the National Institutes of Health.
Conflict of interest statement. None declared.

REFERENCES

1. Rankin,D.J., Rocha,E.P.C. and Brown,S.P. (2011) What traits are carried on mobile genetic elements, and why? *Heredity*, **106**, 1–10.
2. Gillings,M.R., Paulsen,I.T. and Tetu,S.G. (2017) Genomics and the evolution of antibiotic resistance. *Ann. N. Y. Acad. Sci.*, **1388**, 92–107.
3. Chuong,E.B., Elde,N.C. and Feschotte,C. (2017) Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.*, **18**, 71–86.
4. Siguier,P., Gourbeyre,E. and Chandler,M. (2014) Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol. Rev.*, **38**, 865–891.
5. Hickman,A.B. and Dyda,F. (2016) DNA transposition at work. *Chem. Rev.*, **116**, 12758–12784.
6. Turlan,C. and Chandler,M. (2000) Playing second fiddle: second-strand processing and liberation of transposable elements from donor DNA. *Trends Microbiol.*, **8**, 268–274.
7. Curcio,M.J. and Derbyshire,K.M. (2003) The outs and ins of transposition: From Mu to kangaroo. *Nat. Rev. Mol. Cell. Biol.*, **4**, 865–877.
8. Mitra,R., Fain-Thornton,J. and Craig,N.L. (2008) *piggyBac* can bypass DNA synthesis during cut and paste transposition. *EMBO J.*, **27**, 1097–1109.
9. 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.
10. Hencken,C.G., Li,X. and Craig,N.L. (2012) Functional characterization of an active Rag-like transposase. *Nat. Struct. Mol. Biol.*, **19**, 834–836.
11. Liu,K. and Wessler,S.R. (2017) Transposition of *Mutator*-like transposable elements (MULEs) resembles *hAT* and *Transib* elements and V(D)J recombination. *Nucleic Acids Res.*, **45**, 6644–6655.
12. Roth,D.B., Menetski,J.P., Nakajima,P.B., Bosma,M.J. and Gellert,M. (1992) V(D)J recombination: Broken DNA molecules with covalently sealed (hairpin) coding ends in scid mouse thymocytes. *Cell*, **70**, 983–991.
13. Kobryn,K. and Chaconas,G. (2014) Hairpin telomere resolvases. *Microbiol. Spectrum*, **2**, doi:10.1128/microbiolspec.MDNA3-0023-2014.
14. 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.
15. Kapitonov,V.V. and Koonin,E.V. (2015) Evolution of the RAG1-RAG2 locus: both proteins came from the same transposon. *Biol. Direct*, **10**, 20.
16. 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.
17. Kennedy,A.K., Haniford,D.B. and Mizuuchi,K. (2000) Single active site catalysis of the successive phosphoryl transfer steps by DNA transposases: insights from phosphorothioate stereoselectivity. *Cell*, **101**, 295–305.
18. Kennedy,A.K., Guhathakurta,A., Kleckner,N. and Haniford,D.B. (1998) Tn10 transposition via a DNA hairpin intermediate. *Cell*, **95**, 125–134.
19. 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.
20. Reznikoff,W.S. (2008) Transposon Tn5. *Annu. Rev. Genet.*, **42**, 269–286.
21. Hickman,A.B., Ewis,H.E., Li,X., Knapp,J.A., Laver,T., Doss,A.-L., Tolun,G., Steven,A.C., Grishaev,A., Bax,A. *et al.* (2014) Structural basis of *hAT* transposon end recognition by Hermes, an octameric DNA transposase from *Musca domestica*. *Cell*, **158**, 353–367.
22. Ru,H., Chambers,M.G., Fu,T.M., Tong,A.B., Liao,M. and Wu,H. (2015) Molecular mechanism of V(D)J recombination from synaptic RAG1-RAG2 complex structures. *Cell*, **163**, 1138–1152.
23. Kim,M.S., Chuenchor,W., Chen,X., Cui,Y., Zhang,X., Zhou,Z.H., Gellert,M. and Yang,W. (2018) Cracking the DNA code for V(D)J recombination. *Mol. Cell*, **70**, 358–370.
24. Ru,H., Mi,W., Zhang,P., Alt,F.W., Schatz,D.G., Liao,M. and Wu,H. (2018) DNA melting initiates the RAG catalytic pathway. *Nat. Struct. Mol. Biol.*, **25**, 732–742.
25. Yuan,Y.-W. and Wessler,S.R. (2011) The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 7884–7889.
26. Perez,Z.N., Musingarimi,P., Craig,N.L., Dyda,F. and Hickman,A.B. (2005) Purification, crystallization and preliminary crystallographic analysis of the *Hermes* transposase. *Acta Crystallogr.*, **F61**, 587–590.
27. 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.
28. 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. U.S.A.*, **107**, 21966–21972.
29. Guo,Y., Park,J.M., Cui,B., Humes,E., Gangadharan,S., Hung,S., FitzGerald,P.C., Hoe,K.L., Grewal,S.I.S., Craig,N.L. *et al.* (2013) Integration profiling of gene function with dense maps of transposon integration. *Genetics*, **195**, 599–609.
30. Montaño,S.P., Pigli,Y.Z. and Rice,P.A. (2012) The Mu transpososome structure sheds light on DDE recombinase evolution. *Nature*, **491**, 413–417.
31. 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.
32. Arias-Palomo,E. and Berger,J.M. (2015) An atypical AAA+ ATPase assembly controls efficient transposition through DNA remodeling and transposase recruitment. *Cell*, **162**, 860–871.
33. Morris,E.R., Grey,H., McKenzie,G., Jones,A.C. and Richardson,J.M. (2016) A bend, flip and trap mechanism for transposon integration. *eLife*, **5**, e15537.
34. Dornan,J., Grey,H. and Richardson,J.M. (2015) Structural role of the flanking DNA in *mariner* transposon excision. *Nucleic Acids Res.*, **43**, 2424–2432.
35. Rubin,E., Lithwick,G. and Levy,A.A. (2001) Structure and evolution of the hAT transposon superfamily. *Genetics*, **158**, 949–957.
36. Palermo,G., Cavalli,A., Klein,M.L., Alfonso-Prieto,M., Dal Peraro,M. and De Vivo,M. (2015) Catalytic metal ions and enzymatic processing of DNA and RNA. *Acc. Chem. Res.*, **48**, 220–228.
37. 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.
38. Nowotny,M. (2009) Retroviral integrase superfamily: the structural perspective. *EMBO Rep.*, **10**, 144–151.
39. Steitz,T.A. and Steitz,J.A. (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 6498–6502.
40. Rajaei,N., Chiruvella,K.K., Lin,F. and Åström,S.U. (2014) Domesticated transposase Kat1 and its fossil imprints induce sexual differentiation in yeast. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 15491–15496.

41. Chiruvella,K.K., Rajaei,N., Jonna,V.R., Hofer,A. and Åström,S.U. (2016) Biochemical characterization of Kat1: a domesticated *hAT*-transposase that induces DNA hairpin formation and *MAT*-switching. *Sci. Rep.*, **6**, 21671.
42. Liu,K. and Wessler,S.R. (2017) Functional characterization of the active *Mutator*-like transposable element, *Mutal* from the mosquito *Aedes aegypti*. *Mobile DNA*, **8**, 1.
43. Harding,M.M., Nowicki,M.W. and Walkinshaw,M.D. (2010) Metals in protein structures: a review of their principal features. *Crystall. Rev.*, **16**, 247–302.
44. Sadofsky,M.J., Hesse,J.E., van Gent,D.C. and Gellert,M. (1995) RAG-1 mutations that affect the target specificity of V(D)J recombination: a possible direct role of RAG-1 in site recognition. *Genes Dev.*, **9**, 2193–2199.
45. 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.
46. Claeys Bouuaert,C. and Chalmers,R. (2010) Transposition of the human *Hsmar* transposon: rate-limiting steps and the importance of the flanking TA dinucleotide in second strand cleavage. *Nucleic Acids Res.*, **38**, 190–202.
47. Claeys Bouuaert,C., Walker,N., Liu,D. and Chalmers,R. (2014) Crosstalk between transposase subunits during cleavage of the *mariner* transposon. *Nucleic Acids Res.*, **42**, 5799–5808.
48. Liao,G.C., Rehm,E.J. and Rubin,G.M. (2000) Insertion site preferences of the P transposable element in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 3347–3351.
49. Linheiro,R.S. and Bergman,C.M. (2008) Testing the palindromic target site model for DNA transposon insertion using the *Drosophila melanogaster* P-element. *Nucleic Acids Res.*, **36**, 6199–6208.
50. Vigdal,T.J., Kaufman,C.D., Izsvák,Z., Voytas,D.F. and Ivics,Z. (2002) Common physical properties of DNA affecting target site selection of *Sleeping Beauty* and other Tc1/*mariner* transposable elements. *J. Mol. Biol.*, **323**, 441–452.
51. Guo,Y., Zhang,Y. and Hu,K. (2018) Sleeping Beauty transposon integrates into non-TA dinucleotides. *Mobile DNA*, **9**, 8.
52. 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.
53. Li,X., Ewis,H., Hice,R.H., Malani,N., Parker,N., Zhou,L., Feschotte,C., Bushman,F.D., Atkinson,P.W. and Craig,N.L. (2013) A resurrected mammalian *hAT* transposable element and a closely related insect element are highly active in human cell culture. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, E478.
54. Vrljicak,P., Tao,S., Varshney,G.K., Quach,H.N.B., Joshi,A., LaFave,M.C., Burgess,S.M. and Sampath,K. (2016) Genome-wide analysis of transposon and retroviral insertions reveals preferential integrations in regions of DNA flexibility. *Genes Genomes Genetics*, **6**, 805–817.
55. Kuduvalli,P.N., Rao,J.E. and Craig,N.L. (2001) Target DNA structure plays a critical role in Tn*7* transposition. *EMBO J.*, **20**, 924–932.
56. Pribil,P.A. and Haniford,D.B. (2003) Target DNA bending is an important specificity determinant in target site selection in Tn*10* transposition. *J. Mol. Biol.*, **330**, 247–259.
57. Hackett,C.S., Geurts,A.M. and Hackett,P.B. (2007) Predicting preferential DNA vector insertion sites: implications for functional genomics and gene therapy. *Genome Biol.*, **8**, S12.
58. Pflieger,A., Jaillet,J., Petit,A., Augé-Gouillou,C. and Renault,S. (2014) Target capture during Mos1 transposition. *J. Biol. Chem.*, **289**, 100–111.
59. Linheiro,R.S. and Bergman,C.M. (2012) Whole genome resequencing reveals natural target site preferences of transposable elements in *Drosophila melanogaster*. *PLoS ONE*, **7**, e30008.
60. Eisen,J.A., Benito,M.-I. and Walbot,V. (1994) Sequence similarity of putative transposases links the maize *Mutator* autonomous element and a group of bacterial insertion sequences. *Nucleic Acids Res.*, **22**, 2634–2636.
61. Gourbeyre,E., Siguier,P. and Chandler,M. (2010) Route 66: investigations into the organisation and distribution of the IS*66* family of prokaryotic insertion sequences. *Res. Microbiol.*, **161**, 136–143.
62. 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.
63. Kim,M.S., Lapkouski,M., Yang,W. and Gellert,M. (2015) Crystal structure of the V(D)J recombinase RAG1-RAG2. *Nature*, **518**, 507–511.
64. Huye,L.E., Purugganan,M.M., Jiang,M.M. and Roth,D.B. (2002) Mutational analysis of all conserved basic amino acids in RAG-1 reveals catalytic, step arrest, and joining-deficient mutants in the V(D)J recombinase. *Mol. Cell. Biol.*, **22**, 3460–3473.
65. Landree,M.A., Wibbenmeyer,J.A. and Roth,D.B. (1999) Mutational analysis of RAG1 and RAG2 identified three catalytic amino acids in RAG1 critical for both cleavage steps of V(D)J recombination. *Genes Dev.*, **13**, 3059–3069.
66. Nowotny,M., Gaidamakov,S.A., Crouch,R.J. and Yang,W. (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: Substrate specificity and metal-dependent catalysis. *Cell*, **121**, 1005–1016.
67. Hare,S., Maertens,G.N. and Cherepanov,P. (2012) 3′-Processing and strand transfer catalysed by retroviral integrase *in crystallo*. *EMBO J.*, **31**, 3020–3028.
68. Katz,R.A., Merkel,G., Andrake,M.D., Roder,H. and Skalka,A.M. (2011) Retroviral integrases promote fraying of viral DNA ends. *J. Biol. Chem.*, **286**, 25710–25718.
69. Fuller,J.R. and Rice,P.A. (2017) Target DNA bending by the Mu transpososome promotes careful transposition and prevents its reversal. *eLife*, **6**, e21777.