Transposition of *hAT* elements links transposable elements and V(D)J recombination

Liqin Zhou[1], Rupak Mitra[1], Peter W. Atkinson[2], Alison Burgess Hickman[3], Fred Dyda[3] & Nancy L. Craig[1]

[1]*Howard Hughes Medical Institute and Department of Molecular Biology & Genetics, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA*
[2]*Department of Entomology and Institute for Integrative Genome Biology, University of California, Riverside, California 92521, USA*
[3]*Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland 20892, USA*

Transposons are DNA sequences that encode functions that promote their movement to new locations in the genome. If unregulated, such movement could potentially insert additional DNA into genes, thereby disrupting gene expression and compromising an organism's viability. Transposable elements are classified by their transposition mechanisms and by the transposases that mediate their movement. The mechanism of movement of the eukaryotic *hAT* superfamily elements was previously unknown, but the divergent sequence of *hAT* transposases from other elements suggested that these elements might use a distinct mechanism. Here we have analysed transposition of the insect *hAT* element *Hermes in vitro*. Like other transposons, *Hermes* excises from DNA via double-strand breaks between the donor-site DNA and the transposon ends, and the newly exposed transposon ends join to the target DNA. Interestingly, the ends of the donor double-strand breaks form hairpin intermediates, as observed during V(D)J recombination, the process which underlies the combinatorial formation of antigen receptor genes. Significant similarities exist in the catalytic amino acids of *Hermes* transposase, the V(D)J recombinase RAG, and retroviral integrase superfamily transposases, thereby linking the movement of transposable elements and V(D)J recombination.

Transposons are discrete mobile DNA segments that have been found in virtually every genome examined. The major fraction of some genomes, including the human genome, is composed of transposable elements[1,2]. The movement of many transposable elements involves a DNA intermediate whose ends are specifically recognized by the element-encoded transposase. These ends are processed by transposase-mediated DNA breakage reactions that expose the element termini and join them to the target DNA[3,4]. Such elements include both DNA-only elements whose termini are short terminal inverted repeats (TIRs), represented by the prokaryotic IS50/IS10, transposon Tn7, and Mu elements as well as the eukaryotic P and Tc1/Mariner elements, and elements such as retroviruses that generate the integrating DNA version of their genome by reverse transcription of an RNA intermediate. Almost all of the elements with DNA intermediates whose transposition has been investigated biochemically belong to the retroviral integrase superfamily[3,5].

Another major family of transposons is the widespread eukaryotic *hAT* superfamily that includes active elements in fungi, plants and animals, including vertebrates[6–8]; these elements have not yet been studied at the molecular level. *hAT* elements[6,9] include *hobo* of *Drosophila* and the closely related *Hermes* element first identified in housefly[10], McClintock's *Ac* element of maize, and the snapdragon *Tam3* element. They share a number of amino-acid motifs distinct from those shared among the retroviral integrase superfamily[6]. Using purified *Hermes* transposase, we have determined the mechanism of DNA breakage and joining that underlies *hAT* element transposition and have identified amino acids critical for the catalytic steps of recombination.

Hermes transposase binds the transposon ends

The extreme ends of *Hermes* are imperfect TIRs 17 base pairs (bp) long. They are likely to be central to recombination, but other sequences in the ends that may be required for transposition remain to be defined[10,11]. DNA band shift assays reveal that purified *Hermes* transposase produced in *Escherichia coli* binds specifically to a 131-bp fragment containing 80 bp of the *Hermes* left end (*Hermes*-L) flanked by housefly genomic DNA[10] and to a 178-bp fragment containing 110 bp of the *Hermes* right end (*Hermes*-R); much less binding is observed with *Hermes*-R (Fig. 1). The many species observed with *Hermes*-L probably represent the binding of several transposase protomers because *Hermes* transposase binds

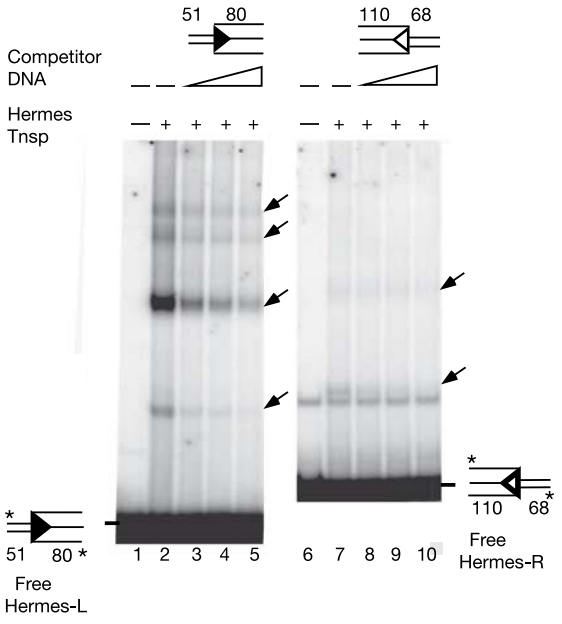

Figure 1 *Hermes* transposase binds specifically to the ends of *Hermes*. DNA binding assays using *Hermes* transposase (Tnsp) and radiolabelled *Hermes*-L (lanes 1–5) and *Hermes*-R (lanes 6–10) end fragments are shown; positions of radiolabelling are indicated by asterisks. Specific competitor DNA is unlabelled. Arrows indicate nucleoprotein complexes. The convention used throughout the figures is that the filled triangle indicates the TIR of *Hermes*-L and the open triangle the TIR of *Hermes*-R.

multiple positions on this DNA segment by DNA footprinting assays (J. Spencer, P. Kuduvalli, X. Li and N.L.C.; R.H. Hice, T.A. Laver and P.W.A., unpublished observations). Some of the *Hermes*-L complexes might also reflect pairing between end segments. The presence of multiple transposase binding sites in a single end and protein bridging of transposon ends has been observed in many other systems[4].

Hermes transposase makes hairpin ends

Genetic analysis suggests that *Hermes* transposes by excision from a donor site, followed by insertion into a target site[11]. To probe these steps, we incubated *Hermes* transposase with a 194-bp fragment containing 91 bp of *Hermes*-L flanked by housefly genomic DNA and a target plasmid, then examined the resulting transposition intermediates and products (Fig. 2). *Hermes*-mediated DNA-processing reactions require Mn^{2+} or Mg^{2+}; more recombination is often seen with Mn^{2+} (see below and Supplementary Information) and the sequence selectivity of some DNA cleavages is greater with Mn^{2+} than Mg^{2+} (Supplementary Fig. 1). Cleavage of the flanking donor-site DNA from *Hermes*-L via a double-strand break results in two products on a native gel (Fig. 2a). The slower-migrating flanking DNA fragment increases in amount as the incubation time is extended; however, little of the faster-migrating *Hermes*-L segment accumulates, because it joins to the target plasmid forming a product which is not resolved on this gel (see Fig. 3). Cleavage of a similar *Hermes*-R fragment was also observed but at a much lower level (data not shown).

On a denaturing gel, two reaction products are also evident (Fig. 2b): the faster-migrating species of about 90 nucleotides (nt) is the top strand of *Hermes*-L, resulting from a single-strand break that

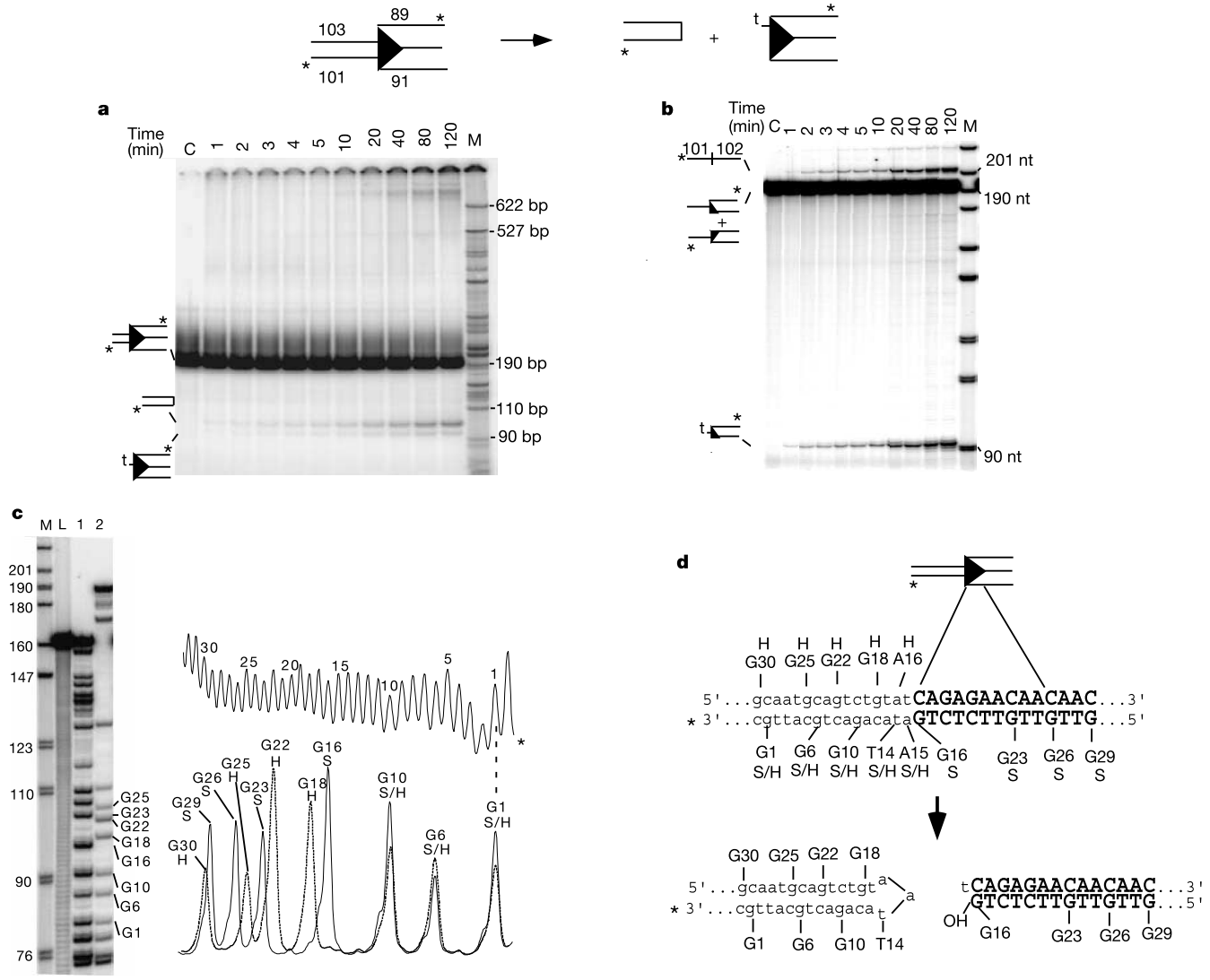

Figure 2 *Hermes* transposition involves hairpin formation on the flanking donor-site DNA Transposition reactions containing Mn^{2+} were performed with *Hermes* Tnsp and a DNA fragment containing *Hermes*-L flanked by housefly genomic DNA. The schematic diagram represents the *Hermes* double-strand break reaction. The 't' on the 5′ transposon end in **a** and **b** indicates the nucleotide of flanking donor-site DNA that remains attached to that end. **a**, Reaction products on a native polyacrylamide gel reveal that the double-strand break reaction separates the transposon end from the flanking donor-site DNA. Higher-molecular-mass strand-transfer products that are not resolved in this gel are trapped in the wells (see Fig. 3). **b**, Reaction products on a denaturing polyacrylamide gel reveal that the flanking donor-site DNA forms a hairpin structure with a lower mobility than the substrate. Lane C, buffer control; lane M, radiolabelled *Msp*I-digested pBR322 DNA. **c**, Analysis of the apparent higher-molecular-mass product by Maxam–Gilbert G sequencing establishes its hairpin nature. Lane L, substrate DNA treated with hydroxy radicals to provide length markers; lane 1, substrate DNA; lane 2, hairpin product. Positions in the substrate strand are designated 'S' and those in the hairpin 'H'. **d**, The sequences of the transposon end, flanking donor-site DNA, and the cleaved transposon end and hairpin on the released donor-site DNA are shown.

separates the top strand of the flanking donor-site DNA from the top strand of the transposon. The same *Hermes*-L fragment is seen when the substrate 192-bp duplex fragment is isolated from a native gel and then run on a denaturing gel, indicating that the donor cleavage reaction begins with a nick near the 5′ end of the transposon that precedes the double-strand break step (data not shown). The second product of the 'flank + *Hermes*-L' cleavage reaction migrates not at 101 nt, as would be expected from a simple double-strand break, but instead migrates more slowly than even the 194-nt substrate strands. Its apparent mobility of >200 nt is consistent with its being a DNA hairpin containing both strands of the flanking donor-site DNA (101 + 102 nt). When this slow-migrating species was isolated from a denaturing gel and rerun on a native gel, its mobility was at a position indicating a fragment of about 100 bp, consistent with a hairpin structure (data not shown).

To determine the structure of the hairpin species, we performed Maxam–Gilbert G sequencing (Fig. 2c). The G sequence of the hairpin species was identical to that of the bottom strand of the substrate flanking DNA with respect to positions G1, G6 and G10. However, the next G was not at position G16, as in the substrate, but rather at position G18, followed by Gs at G22, G25 and G30, reflecting the top strand sequence of the flanking donor-site DNA (Fig. 2d). This pattern is consistent with a hairpin that includes the bottom strand of the flanking donor-site DNA through positions T14 and A15 joined to A16 of the top strand of flanking DNA, and including the rest of the top strand of the flanking donor-site DNA. Analysis using Maxam–Gilbert G + A sequencing was consistent with this hairpin structure (data not shown). Additional evidence for a hairpin is that the T in the hairpin, T14, was hyper-reactive to potassium permanganate treatment (data not shown).

We conclude that *Hermes* donor cleavage involves two distinct chemical steps. Donor cleavage initiates with a nick one nucleotide into the donor strand that flanks the 5′ end of the transposon, leaving one donor nucleotide attached to the 5′ end of the transposon. The 3′ OH of the top-flanking donor strand generated by this nick then joins to the bottom strand at the junction of the flanking donor-site DNA and the 3′ end of the transposon. Hairpin formation releases the 3′ OH transposon terminus; the 5′ transposon terminus is attached to one nucleotide of flanking donor-site DNA. This same mechanism is used in variable(diversity)joining recombination to promote the double-strand break reactions critical to the combinatorial assembly of variable(diversity)joining gene segments to produce diverse immunoglobulin and T-cell receptor genes in V(D)J recombination[12,13]. However, in V(D)J recombination the double-strand break releases flush signal ends (analogous to the transposon ends) lacking unpaired 5′ bases and generates a hairpin coding end (analogous to the donor site).

Hermes 3′ OH termini join to target DNA

Hermes-L ends liberated by the hairpin-mediated breakage of the donor-site DNA join to a target DNA in the presence of *Hermes* transposase. On a native gel, target joining results in two new species (Fig. 3a, b). The lower species results from the concerted joining of two *Hermes*-L segments to the target plasmid, generating an appropriately sized linear plasmid; the upper species represents the joining of one *Hermes*-L transposon segment to the target plasmid, generating a nicked plasmid. As seen with the flanking donor hairpin, the amount of product resulting from joining of *Hermes*-L to the target plasmid increases over time.

To determine the chemistry of end joining to the target DNA, we examined the ability of *Hermes* transposase to join a pre-cleaved, that is, with its 3′ OH terminus already exposed, *Hermes*-L fragment to target plasmids of two different sizes (Fig. 3c). On a native gel, species reflecting the joining of one and two *Hermes*-L fragments to the targets are observed (Fig. 3c, lanes 2, 3). On a denaturing gel, only a single unit-length product is evident with each target, consistent with the joining of one *Hermes*-L segment to a single target strand (Fig. 3c, lanes 5, 6).

Detection of these single-stranded products in reactions using a *Hermes*-L oligonucleotide labelled at the 5′ end of its bottom strand reveals the chemistry of transposon end joining to the target: the 3′ OH terminus of *Hermes*-L joins covalently to the target DNA. Interestingly, the 3′ OH dinucleotide at the position of strand joining, TG-OH, which is highly conserved among *hAT* superfamily

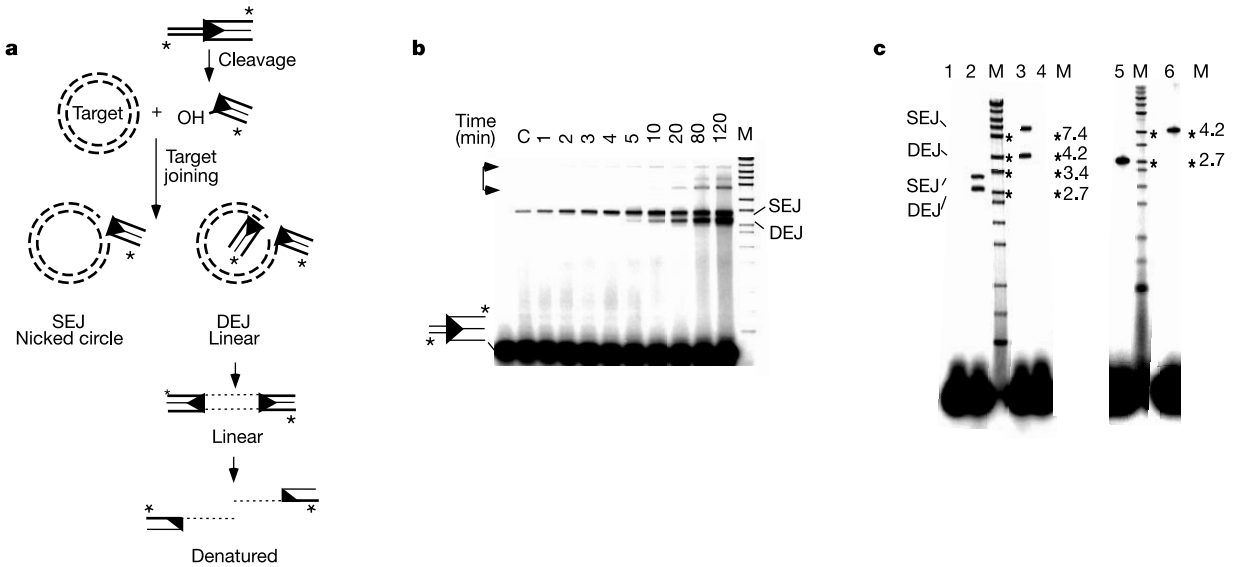

Figure 3 *Hermes* transposase joins the 3′ OH end of *Hermes*-L to a target DNA. **a**, A schematic representation of the double-strand break that releases the transposon end from the flanking donor-site DNA and the subsequent joining of the transposon end to the target DNA is shown. SEJ, single-end join; DEJ, double-end join. **b**, The target-joining products generated over time when *Hermes* Tnsp is incubated in Mn^{2+}-containing buffer with a *Hermes*-L-containing fragment labelled at its 5′ ends and a circular target plasmid (2.7 kb) as described in Fig. 2a are displayed on a native agarose gel. The 2.7-kb linear product results from the joining of two *Hermes*-L fragments to the target plasmid; the 3.4-kb nicked circular product results from the joining of one *Hermes*-L to the target plasmid. Products with sizes >3.4 kb result from the joining of *Hermes* ends to oligomeric forms of the target plasmid. **c**, The target joining products generated when *Hermes* Tnsp is incubated with pre-cleaved *Hermes*-L ends radiolabelled at the 5′ end of the bottom (transferred) strand and circular target plasmids of 2.7 and 4.2 kb are displayed on native (lanes 2, 3) and denaturing (lanes 5, 6) agarose gels. Lanes 1, 4: buffer control.

transposons[6], including *Hermes*, is also found at the position of DNA cleavage and joining in V(D)J recombination[12,14].

Transposition by *Hermes* is accurate

We asked whether *Hermes* transposase could accurately join a DNA segment bounded by the *Hermes* termini to a target DNA. We generated a mini-*Hermes* element with pre-cleaved *Hermes*-L ends flanking a kanamycin resistance gene and used this DNA as a substrate for *in vitro* transposition into a target plasmid. Transposition products were recovered by transformation of *E. coli* with selection for kanamycin resistance. Junctions between the plasmid and the newly inserted transposons were sequenced. All 43 insertions analysed were accompanied by 8-bp target site duplication (data not shown) as occurs with *Hermes* and other *hAT* transposons *in vivo*[15]. Analysis of a mini-*Hermes* element containing *Hermes*-L and -R ends gave similar results (data not shown). Thus, our *Hermes in vitro* transposition system reflects *hAT* transposition *in vivo*.

Hermes transposition compared to other elements

Our experiments have revealed the mechanism of breakage and joining events underlying the translocation of a *hAT* superfamily transposon: the transposon is excised from the donor site by double-strand breaks mediated by hairpin formation on the flanking donor-site DNA, and the 3′ OH transposon ends exposed by this cleavage join to the target DNA (Fig. 4). Element excision via double-strand breaks involving hairpin formation on the donor-site DNA has not been observed with any other transposon family. We note that although excision of Tn5 and Tn10 (members of the retroviral integrase superfamily) from their donor sites also involves hairpin formation[16,17], the hairpins in these systems are on the transposon ends, rather than on the flanking DNA as in *Hermes* transposition and V(D)J recombination. The strategy of joining of a 3′ OH end to a target DNA is, however, highly conserved among mobile elements and can occur with the V(D)J recombinase RAG[18,19] and members of the retroviral integrase superfamily[3]. It will be interesting to determine whether *Hermes*, like members of the retroviral integrase superfamily and RAG recombinases, uses direct one-step transesterification mechanisms both to promote hairpin formation[20,21] and to join the transposon end to the target DNA[21,22]. The experiments described below reveal that the same active site in *Hermes* can mediate the distinct chemical steps of nicking, hairpin formation and target joining, as has been found with other transposases and integrases[3,5].

It will be especially interesting to determine the organization of the transposase complex that executes recombination: do single or multiple transposase protomers interact with each transposon end and, if there are multiple protomers, how are the chemical steps of recombination distributed among them? *Hermes* transposase can multimerize[23] (A.B.H. and F.D., unpublished work). Studies of Mu[24,25] and Tn10[21] transposition indicate that all the chemical steps in the movement of these elements are executed by a single transposase monomer at each transposon end, but the different order of the chemical steps and spatial positions of DNA substrates in *Hermes* and V(D)J recombination could well require multiple transposase active sites at each terminus.

After transposon excision, the broken donor chromosome must be repaired. With *hAT* elements, this occurs by rejoining of the exposed flanking DNA ends. The patterns of these joints led to the suggestion that hairpin intermediates were involved in *hAT* transposition[26,27]; we have now verified that prediction biochemically. For repair to occur, these hairpins must be opened. Our experiments suggest that *Hermes* transposase does not perform this step, because the amount of hairpin formed increases over time (Fig. 2b) and no cleaved hairpin products have been detected. Therefore a host function like Artemis, which cleaves hairpins, is likely to be involved in *Hermes* transposition[28]; an open reading frame (ORF) encoding an Artemis-like protein is present in the *Drosophila* and *Anopheles* genomes (Blast data not shown). Subsequent joining of the opened hairpins is probably performed by the host-encoded non-homologous end-joining DNA repair machinery[27,29]. A potential active role for *Hermes* transposase in repair may be to maintain pairing of the broken DNAs to facilitate joining, a role suggested for the RAG recombinase in V(D)J recombination[12,30,31].

Unanticipated structure of the *Hermes* active site

Sequence comparisons among *hAT* elements have focused on genome-wide inspections and have identified amino-acid motifs that are often, but not always, present in *hAT* family members[6,8]. To identify amino acids essential to *hAT* transposase activity, we

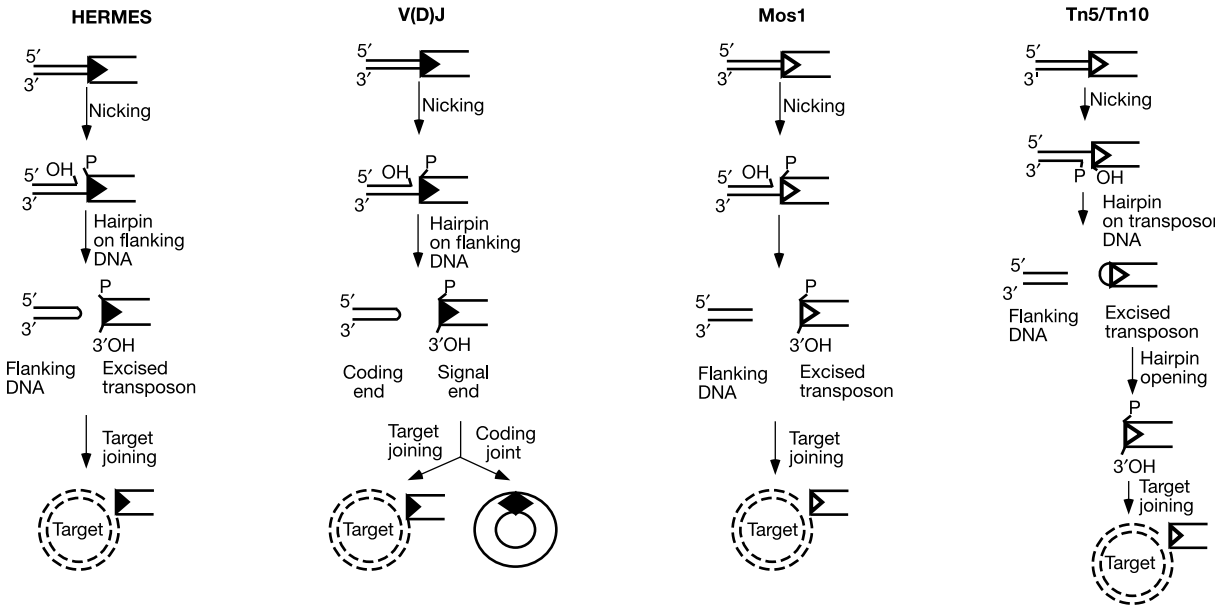

Figure 4 Comparison of recombinase-mediated cleavage and target-joining mechanisms. The positions of DNA cleavage reactions on various mobile DNAs are shown. Although the exact positions of the 5′ end cleavage vary, all elements are shown here with flush ends.

aligned the known active members of the *hAT* family using Clustal W[32] (Supplementary Fig. 2) and identified several invariant acidic amino acids, in *Hermes* D180, D248 and E572, that could provide carboxylates to interact with the metal ion. The acidic DDE motif has previously been shown to be involved in the catalytic activity of members of the retroviral integrase superfamily and RAG recombinase[3,12]. To probe the function of these amino acids in *Hermes* transposase, we purified mutant proteins with alanine substitutions at these positions and assayed them for DNA binding and recombination using several DNA substrates: (1) a substrate in which the transposon end was flanked by donor-site DNA, to probe coupled donor-site cleavage and strand transfer; (2) a 'pre-nicked' substrate with a nick in the top strand of the flanking donor-site DNA adjacent to the 5′ end of the transposon, as is generated at the first step of *Hermes* transposition, to probe hairpin formation; and (3) a 'pre-cleaved' substrate in which the *Hermes* 3′ OH ends were already exposed, to probe strand transfer.

Although still capable of binding specifically to *Hermes*-L (Supplementary Fig. 3), the D180A, D248A and E572A transposase mutants were inactive or had greatly reduced activity in almost all DNA breakage and joining steps (Fig. 5). These conserved acidic amino acids therefore play critical roles in catalysis. The fact that a single mutation can block multiple catalytic steps also suggests that a single active site mediates all the chemical steps of recombination. The mutants were greatly impaired at the first step in recombination, that is, single-strand nicking at the 5′ end of the transposon; unlike in the wild type, no cleaved transposon top strand was seen after incubation of the mutants with a DNA substrate in which donor-site DNA flanked the transposon end with either Mn^{2+} or Mg^{2+} (Fig. 5a). Even when supplied with a 'pre-nicked' substrate, no DNA hairpin formation was observed with any mutant in Mg^{2+}, although D248A was able to promote some hairpin formation in presence of Mn^{2+} (Fig. 5b). The alanine substitution mutants were also defective in strand transfer, because no joining of a 'pre-cleaved' transposon to a target DNA was detectable in Mg^{2+}; some strand transfer with D248A was observed in the presence of Mn^{2+} (Fig. 5c).

We also examined the effects of cysteine substitutions at these DDE positions: specific suppression of a Cys mutant by Mn^{2+} has been associated with a direct interaction between the metal and amino acid[33–35]. The presence of Mn^{2+} dramatically suppressed the D180C defect in DNA hairpin formation and strand transfer, suggesting that this amino acid does interact with metal ion in the active site (Fig. 5b, c). We also observed some modest suppression by Mn^{2+} of the D248C defect in hairpin formation and target joining (Fig. 5b, c). The failure to observe Mn^{2+} suppression of E572C in any step of recombination or of D180C and D248C in a coupled cleavage and strand transfer assay (Fig. 5; Supplementary Fig. 4) probably reflects the fact that each metal binding site has a slightly different role in recombination[36].

Structural analysis of the HIV integrase and Mu and Tn5 transposases has previously demonstrated that the highly conserved catalytic DDE acidic amino acids lie on a common structure, that is, an RNaseH-like fold, and that they are closely juxtaposed, as appropriate for formation of a single active site[5,37]. We performed secondary structure predictions with the *hAT* transposases from *Hermes*, *hobo*, *Ac* and *Tam3*[38]. The results suggest that the conserved *hAT* catalytic acidic amino acids are also arranged in an RHaseH-like fold, suggesting that the active sites of *hAT* elements and retroviral integrase elements are actually similar (Supplementary Fig. 5). This finding is unexpected because no sequence homology between *hAT* transposases and retroviral integrase transposases has been detected[6]. The RAG1 subunit of the V(D)J recombinase also contains catalytic acidic amino acids that may also be organized in an RNaseH-like fold[39–41]. Interestingly, with both *hAT* transposases and RAG1, and in contrast to other already known members of the retroviral integrase superfamily[3], the second D and the E are widely separated in the sequence, D248–E572 in *Hermes* and D708–E962 in RAG1. It will be interesting to determine whether this wider spacing is functionally related to the fact that these enzymes make hairpins on the flanking DNA rather than on the transposon end as occurs with retroviral integrase elements Tn5 and Tn10.

Our identification of a retroviral-integrase-superfamily-like active site in what had been thought to be a distinct superfamily[6] extends and broadens the types of mobile DNA elements that are likely to use the same mechanisms of DNA breakage and joining to transpose. Two other transposon superfamilies are *MuDR/Mu* of plants[42] and the more widespread *piggyBac*[43]. We have also analysed the secondary structures of members of these superfamilies (data not shown). The positions of conserved acidic amino acids suggests that at least part of the active sites of these transposases is related to

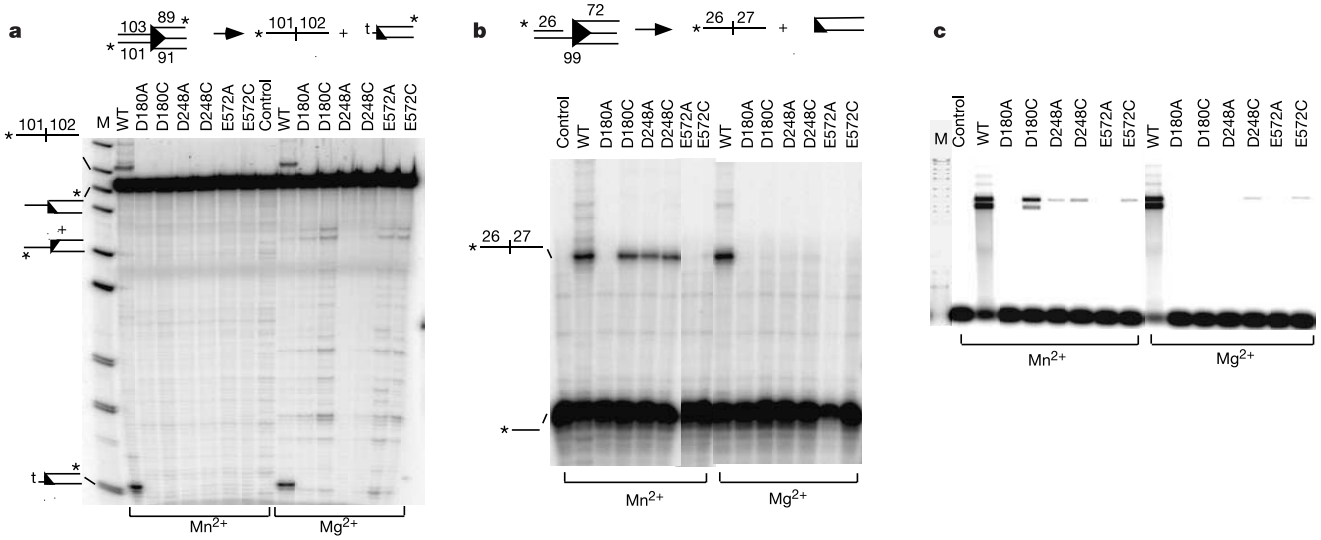

Figure 5 Mutation of *Hermes* DDE residues block DNA cleavage and strand transfer. **a**, **b**, The products of incubation of wild-type, alanine and cysteine substitutions at the DDE (conserved Asp-Asp-Glu) amino-acid motif of *Hermes* Tnsp in the presence of either Mn^{2+} or Mg^{2+} with various *Hermes*-L substrates are displayed along with schematic representations of each reaction. **a**, DNA nicking and hairpin formation using a flank + *Hermes*-L substrate is visualized on a denaturing agarose gel. **b**, DNA hairpin formation using a pre-nicked flank + *Hermes*-L substrate is visualized on a denaturing acrylamide gel. **c**, DNA target joining using a pre-cleaved *Hermes*-L substrate is visualized on a native agarose gel.

those of the RNaseH-fold of the retroviral integrase and *hAT* superfamilies. Thus, very many superfamilies of elements all appear to contain similar active sites for DNA cleavage and strand transfer.

Other shared regions of *Hermes* and RAG1

In addition to the conserved DDE amino acids, there are several other interesting similarities between *Hermes* and RAG1. The formation of a DNA hairpin involves considerable DNA distortion. Such distortions can be facilitated by 'base flipping' during which a particular base becomes extrahelical, allowing distortion of the backbone[44]. Base flipping can be stabilized by interactions between the base and aromatic amino acids (for example, tryptophan). A highly conserved region of *hAT* transposases[6] contains a tryptophan, which is W319 in *Hermes* (Supplementary Fig. 2). Strikingly, there is a short region of amino-acid similarity between this region of *Hermes* sequence and the region of Tn5 transposase that contains a tryptophan involved in the base-flipping event that is essential to formation of the hairpin during Tn5 transposition[45,46] (Supplementary Fig. 5). A highly conserved basic residue, R318 of *Hermes*, is closely juxtaposed to this tryptophan. We speculate that W319 of *Hermes* and related sequences in other *hAT* transposases will play key parts in DNA hairpin formation and base flipping. Mutational analysis of RAG1 has identified a number of basic amino acids whose alteration affects the hairpin formation step of RAG recombination[47]; two of these, K889 and R894, flank a conserved tryptophan W893, and we speculate that W893 will also play a key role in hairpin formation, probably by promoting base flipping (Fig. 5). A similar role for this tryptophan has also been suggested on the basis of work on telomere resolvases[48].

Another very conserved feature of *hAT* transposases[6] (Supplementary Fig. 2) is a CxxH motif (C265 and H268 in *Hermes*) downstream of the second catalytic D residue (D248 in *Hermes*); see Supplementary Figs 2 and 5. Mutation in H268 resulted in a severe defect in transposase catalytic activity (data not shown). RAG1 contains a C_2H_2 zinc-finger motif beginning at C720 just downstream of the second D, D708 (ref. 49).

A significant difference between *Hermes* transposition and RAG recombination is the fate of the DNA bound by the recognition and sites of action of the recombinases, that is, the TIRs of *Hermes* and the recombination signal sequences of V(D)J recombination. Recognition of these sites results in the DNA breakage that leaves hairpins of the flanking DNA and excises the segment bound by the TIRs or the recombination signal sequences. Whereas *Hermes* transposase often promotes insertion of the TIR-bounded fragment (the transposon) into a new target site after excision, the recombination-signal-sequence ends of a DNA fragment that excises from between the immunoglobulin and T-cell receptor genes to juxtapose novel gene coding segments almost always join to each other to form a 'signal sequence joint' and hence a circular species that is simply lost from the cell[12]. Elucidation of how *Hermes* and RAG1 promote these different outcomes will lead to greater understanding of their mechanisms of action.

Conclusions

We have directly established the pathway of DNA breakage and joining reactions that underlie movement of a *hAT* family transposon and have also begun to dissect the structure and function of a *hAT* transposase. Our finding of an RNaseH-like fold with conserved acidic amino acids is surprising because the *hAT* family has long been considered to be quite distinct from the retroviral integrase superfamily. It had been suggested that the V(D)J recombination system may have evolved from an ancient transposable DNA element[12,50]. Our findings here of such a close mechanistic relationship between *hAT* transposition and V(D)J recombination—that is, a double-strand break via hairpin formation on flanking DNA and 3′ OH joining to the target DNA—and the related active sites of *hAT* transposases and RAG1 provides strong support for the view that V(D)J recombination evolved from transposable elements. Dissection of the structure and function of *hAT* transposases will probably continue to provide important contributions to understanding V(D)J recombination. □

Methods

Hermes transposase expression and purification

The *Hermes* transposase (Tnsp) ORF (613 amino acids) was amplified by polymerase chain reaction (PCR) from plasmid pBCHSHH1.9 (ref. 11) and cloned between the *Nco*I and *Pvu*II sites of plasmid pBAD-Myc-HisB (Invitrogen) to generate a *Hermes*–Myc–His fusion construct. *Escherichia coli* strain Top10 (Invitrogen) transformed with the *Hermes*–Myc–His plasmid was grown overnight with shaking at 30 °C in LB medium containing 100 mg ml^{-1} carbenicillin. The following day the overnight culture was diluted 1:100 with fresh LB+carbenicillin and cells were then grown to an absorbance at 600 nm of 0.6 at 30 °C. The culture was then shifted to 16 °C and induced with 0.1% L-arabinose for 16 h. After induction, cells were washed by centrifugation at 4 °C with TSG (20 mM Tris-HCl, pH 7.9, 500 mM NaCl, 10% v/v glycerol), and frozen in liquid nitrogen; all subsequent steps were performed at 4 °C. Frozen cells were resuspended in 10 ml TSG and lysed by sonication. The cleared lysate was loaded onto a pre-equilibrated Ni^{2+} Sepharose column (Amersham) and washed with ten column volumes of TSG, six column volumes of TSG+50 mM imidazole and six column volumes of TSG+100 mM imidazole. *Hermes*–Myc–His fusion protein was eluted with six column volumes of TSG+200 mM imidazole, dialysed against TSG, and stored at −80 °C.

DNA binding assay

DNA binding reactions were performed using *Hermes* end fragments generated by PCR from pBS*Hermes*[10,11]. The 131-bp *Hermes*-L end fragment contained 81 bp *Hermes*-L+50 bp flanking housefly genomic DNA; the 178-bp *Hermes*-R fragment contained 110 bp *Hermes*-R+68 bp flanking housefly genomic DNA. DNAs were 5′ end radiolabelled on both strands with γ-P^{32}-ATP. 140 nM *Hermes* Tnsp and 1.5 nM radiolabelled DNA were incubated in 25 mM HEPES, 3.5% (v/v) glycerol, 0.01% bovine serum albumin (BSA) and 4 mM dithiothreitol (DTT) in the presence of a 500-fold excess of sheared herring sperm DNA at room temperature for 20 min. Unlabelled 131-bp *Hermes*-L and 178-bp *Hermes*-R DNA was added in 50, 100 and 200 molar excess as indicated. The products were run on a 5% Tris-boric acid acrylamide gel at 4 °C. All gels were dried and exposed to phosphorimager plates and analysed by Imagequant software (Molecular Dynamics).

Analysis of coupled cleavage and strand transfer

Cleavage and strand transfer reactions were performed with a 193-bp *Hermes*-L end fragment containing 91 bp of *Hermes*-L+103-bp flanking housefly genomic DNA. The 194-bp *Hermes*-L fragment was radiolabelled on the 3′ end of both strands with α-P^{32}-dATP. 140-nM *Hermes* transposase was incubated with 1.5 nM radiolabelled *Hermes*-L DNA in 25 mM HEPES, pH 7.6, 2% (v/v) glycerol, 50 mM NaCl, 2 mM DTT, 1 mM $MnCl_2$, 100 μg ml^{-1} BSA and 9.55 nM pUC19 plasmid as target DNA in a final volume of 20 μl at 30 °C for the indicated times. Reactions were stopped by adding SDS and EDTA to 1% SDS and 20 mM EDTA and incubated for 1 h at 37 °C. The products were displayed on a 5% native acrylamide and on 1% agarose gels to look for double-strand break and target-joining products, respectively. For analysis of DNA nicking and hairpin formation, the reaction products were extracted with phenol/chloroform, precipitated with ethanol and run on a 5% urea acrylamide gel.

Analysis of hairpin formation using a pre-nicked *Hermes*-L end

A 26-bp oligonucleotide corresponding to the top-flanking donor-site DNA was radiolabelled at its 5′ end with γ-P^{32}-ATP. After purification on a Sephadex G10 column, the end-labelled oligonucleotide was mixed with equimolar amounts of a 72-nt oligonucleotide corresponding to top strand of *Hermes*-L, and a 99-nt oligonucleotide containing 27 nt of bottom-strand flanking DNA+72 nt of bottom strand of *Hermes*-L (see Fig. 5b); the mixture was allowed to anneal and then used as substrate in coupled cleavage and strand-transfer reactions.

Strand-transfer reaction with pre-cleaved *Hermes*-L ends

Pre-cleaved *Hermes*-L for strand-transfer reactions was made by annealing oligonucleotides 5′-CAGAGAACAACAACAAGTGGCTTATTTTGATACTTATGCG-3′ (top) and 5′-CGCATAAGTATCAAAATAAGCCACTTGTTGTTGTTCTCTG-3′ (bottom) radiolabelled at its 5′ end with γ-P^{32}-dATP and used directly as a substrate for strand-transfer reactions with pUC19 and pBR322 target DNAs. Reactions were incubated for 2 h at 30 °C. The reactions were stopped by addition of SDS and EDTA to 1% SDS and 20 mM EDTA, incubated at 37 °C for 1 h, DNA was extracted with phenol/chloroform, precipitated with ethanol and loaded onto 1% TAE agarose and 1% NaOH agarose gels.

Recovery of transposon insertions

Primers NLC 1267 5′-CAGAGAACAACAACAAGTGGCTTATTTTGAGCTTGCCTCGTCCCC-3′ and NLC 1268 5′-CAGAGAACAACAACAAGTGGCTTATTTTGACACTGGATGGCGGCG-3′ were used to generate a kanamycin fragment flanked by 30 bp of *Hermes*-L at each end. These kanamycin fragments were used as substrate for a target-joining reaction with pUC19 as target DNA in the presence of *Hermes* Tnsp. The target-joining products were transformed into *E. coli*, kanamycin resistant colonies were selected, and the insertion site was sequenced.

Received 15 August; accepted 3 November 2004; doi:10.1038/nature03157.

1. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
2. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291,** 1304–1351 (2001).
3. Haren, L., Ton-Hoang, B. & Chandler, M. Integrating DNA: transposases and retroviral integrases. *Annu. Rev. Microbiol.* **53,** 245–281 (1999).
4. Craig, N., Craigie, R., Gellert, M. & Lambowitz, A. *Mobile DNA II* (ASM, Washington DC, 2002).
5. Rice, P. A. & Baker, T. A. Comparative architecture of transposase and integrase complexes. *Nature Struct. Biol.* **8,** 302–307 (2001).
6. Rubin, E., Lithwick, G. & Levy, A. A. Structure and evolution of the *hAT* transposon superfamily. *Genetics* **158,** 949–957 (2001).
7. Kunze, R. W. & Weil, C. F. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A.) 565–610 (ASM, Washington DC, 2002).
8. Robertson, H. M. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A.) 1093–1110 (ASM, Washington DC, 2002).
9. Calvi, B. R., Hong, T. J., Findley, S. D. & Gelbart, W. M. Evidence for a common evolutionary origin of inverted repeat transposon in *Drosophila* and plants: *hobo*, *Activator*, and *Tam3*. *Cell* **66,** 465–471 (1991).
10. Warren, W. D., Atkinson, P. W. & O'Brochta, D. A. 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 (1994).
11. O'Brochta, D. A., Warren, W. D., Saville, K. J. & Atkinson, P. W. Hermes, a functional non-Drosophilid insect gene vector from *Musca domestica*. *Genetics* **142,** 907–914 (1996).
12. Gellert, M. V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu. Rev. Biochem.* **71,** 101–132 (2002).
13. Roth, D. B., Menetski, J. P., Nakajima, P. B., Bosma, M. J. & Gellert, M. V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in SCID mouse thymoctes. *Cell* **70,** 983–991 (1992).
14. Okuda, M., Ikeda, K., Namiki, F., Nishi, K. & Tsuge, T. Tfo1: an *Ac*-like transposon from the plant pathogenic fungus *Fusarium oxysporum*. *Mol. Gen. Genet.* **258,** 599–607 (1998).
15. Sarkar, A. *et al.* The Hermes element from *Musca domestica* can transpose in four families of cyclorrhaphan flies. *Genetics* **99,** 15–29 (1997).
16. Kennedy, A., Guhathakurta, A., Kleckner, N. & Haniford, D. B. Tn10 transposition via a DNA hairpin intermediate. *Cell* **95,** 125–134 (1998).
17. Bhasin, A., Goryshin, I. Y. & Reznikoff, W. S. Hairpin formation in Tn5 transposition. *J. Biol. Chem.* **274,** 37021–37029 (1999).
18. Hiom, K., Melek, M. & Gellert, M. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* **94,** 463–470 (1998).
19. Agrawal, A., Eastman, O. M. & Schatz, D. G. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394,** 744–751 (1998).
20. van Gent, D. C., Mizuuchi, K. & Gellert, M. Similarities between initiation of V(D)J recombination and retroviral integration. *Science* **271,** 1592–1594 (1996).
21. Kennedy, A. K., Haniford, D. B. & Mizuuchi, K. Single active site catalysis of the successive phosphoryl transfer steps by DNA transposases: insights from phosphorothioate stereoselectivity. *Cell* **101,** 295–305 (2000).
22. Mizuuchi, K. & Adzuma, K. Inversion of the phosphate chirality at the target site of the Mu DNA strand transfer: evidence for a one-step transesterification mechanism. *Cell* **66,** 129–140 (1991).
23. Michel, K., O'Brochta, D. A. & Atkinson, P. W. The C-terminus of the Hermes transposase contains a protein multimerization domain. *Insect Biochem. Mol. Biol.* **33,** 959–970 (2003).
24. Namgoong, S. & Harshey, R. The same two monomers within a MuA tetramer provide the DDE domains for the strand cleavage and strand transfer steps of transposition. *EMBO J.* **17,** 3775–3785 (1998).
25. Williams, T. L., Jackson, E. L., Carritte, A. & Baker, T. A. Organization and dynamics of the Mu transpososome: recombination by communication between two active sites. *Genes Dev.* **13,** 2725–2737 (1999).
26. Coen, E. S., Carpenter, R. & Martin, C. Transposable elements generate novel spatial patterns of gene expression in *Antirrhinum majus*. *Cell* **47,** 285–296 (1986).
27. Weil, C. F. & Kunze, R. Transposition of maize *Ac*/*Ds* transposable elements in the yeast *Saccharomyces cerevisiae*. *Nature Genet.* **26,** 187–190 (2000).
28. Ma, Y., Pannicke, U., Schwarz, K. & Lieber, M. R. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108,** 781–794 (2002).
29. Yu, J., Marshall, K., Yamaguchi, M., Haber, J. E. & Weil, C. F. Microhomology-dependent end joining and repair of transposon-induced DNA hairpins by host factors in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **24,** 1351–1364 (2004).
30. Qiu, J. X., Kale, S. B., Yarnell, S. H. & Roth, D. B. Separation-of-function mutants reveal critical roles for RAG2 in both the cleavage and joining steps of V(D)J recombination. *Mol. Cell* **7,** 77–87 (2001).
31. Lee, G., Neiditch, M., Salus, S. & Roth, D. RAG proteins shepherd double-strand breaks to a specific pathway, suppressing error-prone repair, but RAG nicking initiates homologous recombination. *Cell* **117,** 171–184 (2004).
32. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680 (1994).
33. Sarnovsky, R., May, E. W. & Craig, N. L. The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *EMBO J.* **15,** 6348–6361 (1996).
34. Piccirilli, J. A., Vyle, J. S., Caruthers, M. H. & Cech, T. R. Metal ion catalysis in the *Tetrahymena* ribozyme reaction. *Nature* **361,** 85–88 (1993).
35. Allingham, J., Pribil, P. & Haniford, D. All three residues of the Tn10 transposase DDE catalytic triad function in divalent metal ion binding. *J. Mol. Biol.* **289,** 1195–1206 (1999).
36. Brautigam, C. & Steitz, T. Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* **8,** 54–63 (1998).
37. Dyda, F. *et al.* Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266,** 1981–1986 (1994).
38. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16,** 404–405 (2000).
39. Kim, D. R., Dai, Y., Mundy, C. L., Yang, W. & Oettinger, M. A. Mutations of acidic residues in RAG1 define the active site of the V(D)J recombinase. *Genes Dev.* **13,** 3070–3080 (1999).
40. Fugmann, S. D., Villey, I. J., Ptaszek, L. M. & Schatz, D. G. Identification of two catalytic residues in RAG1 that define a single active site within the RAG1/RAG2 protein complex. *Mol. Cell* **5,** 97–107 (2000).
41. Landree, M. A., Wibbenmeyer, J. A. & Roth, D. B. Mutational analysis of RAG1 and RAG2 identifies three catalytic amino acids in RAG1 critical for both cleavage steps of V(D)J recombination. *Genes Dev.* **13,** 3059–3069 (1999).
42. Walbot, V. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A.) 533–563 (ASM, Washington DC, 2002).
43. Sarkar, A. *et al.* Molecular evolutionary analysis of the widespread piggyBac transposon family and related "domesticated" sequences. *Mol. Genet. Genom.* **270,** 173–180 (2003).
44. Roberts, R. & Cheng, X. Base flipping. *Annu. Rev. Biochem.* **67,** 181–198 (1998).
45. Davies, D. R., Braam, L. M., Reznikoff, W. S. & Rayment, I. The three-dimensional structure of a Tn5 transposase-related protein determined to 2.9 Å resolution. *J. Biol. Chem.* **274,** 11904–11913 (1999).
46. Ason, B. & Reznikoff, W. S. Mutational analysis of the base flipping event found in Tn5 transposition. *J. Biol. Chem.* **277,** 11284–11291 (2002).
47. Huye, L. E., Purugganan, M. M., Jiang, M. M. & Roth, D. B. 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 (2002).
48. Bankhead, T. & Chaconas, G. Mixing active-site components: a recipe for the unique enzymatic activity of a telomere resolvase. *Proc. Natl Acad. Sci. USA* **101,** 13768–13773 (2004).
49. De, P. & Rodgers, K. Putting the pieces together: identification and characterization of structural domains in the V(D)J recombination protein RAG1. *Immunol. Rev.* **200,** 70–82 (2004).
50. Lewis, S. M. & Wu, G. E. The origins of V(D)J recombination. *Cell* **88,** 159–162 (1997).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We are grateful to G. Chaconas for sharing his thoughts about hairpin binding pockets before publication. N.L.C. is an Investigator of the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.L.C. (ncraig@jhmi.edu).