nature structural & molecular biology

Molecular architecture of a eukaryotic DNA transposase

Alison B Hickman[1], Zhanita N Perez[1], Liqin Zhou[2], Primrose Musingarimi[1,4], Rodolfo Ghirlando[1], Jenny E Hinshaw[3], Nancy L Craig[2] & Fred Dyda[1]

Mobile elements and their inactive remnants account for large proportions of most eukaryotic genomes, where they have had central roles in genome evolution. Over 50 years ago, McClintock reported a form of stress-induced genome instability in maize in which discrete DNA segments move between chromosomal locations. Our current mechanistic understanding of enzymes catalyzing transposition is largely limited to prokaryotic transposases. The *Hermes* transposon from the housefly is part of the eukaryotic *hAT* superfamily that includes *hobo* from *Drosophila*, McClintock's maize *Activator* and *Tam3* from snapdragon. We report here the three-dimensional structure of a functionally active form of the transposase from *Hermes* at 2.1-Å resolution. The Hermes protein has some structural features of prokaryotic transposases, including a domain with a retroviral integrase fold. However, this domain is disrupted by the insertion of an additional domain. Finally, transposition is observed only when Hermes assembles into a hexamer.

Genome sequencing efforts have yielded many unanticipated results, among them the realization that large proportions of many eukaryotic genomes originated as mobile elements. These mobile elements are discrete pieces of DNA that can either move from one place to another within a genome or be copied into a new location. Among eukaryotic genomes, important classes of mobile elements include the long terminal repeat retrotransposons, non–long terminal repeat retrotransposons (comprising LINES and SINES, the long and short interspersed elements, respectively) and DNA transposons[1]. The relative proportions of elements from each class, and those that are currently active, vary from species to species. In humans, for example, ~35% of the genome is derived from non–long terminal repeat retrotransposons[2], a few of which remain active today[3]. DNA transposons are much less prevalent, contributing ~3% of the human genome, and are all believed to have been inactivated through mutation[2]. In contrast, in *Caenorhabditis elegans*, the largest class of mobile elements is the DNA transposons, and the movement of *Tc1/mariner* transposons is responsible for most spontaneous mutations in this organism[4].

Eukaryotic DNA transposons have been classified into superfamilies, one of which contains the *hAT* elements[5], named after *hobo* from *Drosophila*, McClintock's maize *Activator*[6] and *Tam3* from snapdragon. All *hAT* elements share several defining features, including short terminal inverted repeats at each end of the element, the generation of 8-base-pair (bp) target-site duplications upon transposition, and a gene encoding a transposase that catalyzes the DNA cleavage and target-joining steps of transposition. We have been particularly interested in the biochemistry and mechanism of the Hermes *hAT* transposase from the housefly *Musca domestica*[7]. *Hermes* is a 2,749-bp element that contains 17-bp terminal inverted repeats and encodes a 70-kDa transposase[7].

Recent biochemical work has shown that the transposition of *Hermes* uses a cut-and-paste mechanism and that the DNA flanking the excised element transiently forms hairpins before the gap is repaired[8]. Similar hairpins are also observed during recombination activating gene (RAG)-catalyzed V(D)J recombination[9,10], in which variable (V), joining (J) and diversity (D) gene segments are joined to generate immunoglobulins and T-cell receptors in vertebrates, suggesting that an ancient *hAT* element may have been the evolutionary predecessor of the V(D)J recombination system. Although initially thought to be unrelated in structure to any of the characterized prokaryotic transposases, the Hermes transposase contains three essential acidic amino acids reminiscent of the active site residues of members of the retroviral integrase superfamily[8]. To gain insight into the mechanism of *hAT* element transposition and to further investigate the connection between *hAT* elements and V(D)J recombination, we have determined the three-dimensional structure of a catalytically active portion of the Hermes transposase.

RESULTS

Hermes$_{79-612}$ is a three-domain protein

The full-length Hermes transposase (residues 1–612) is soluble, but it forms large aggregates in solution when expressed as an N-terminally histidine (His)-tagged fusion protein in *Escherichia coli*. However,

[1]Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. [2]Howard Hughes Medical Institute, Department of Molecular Biology and Genetics, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA. [3]Laboratory of Cell Biochemistry and Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. [4]Present address: London School of Economics and Political Science, Department of Social Policy, London WC2A 2AE, UK. Correspondence should be addressed to F.D. (Fred.Dyda@nih.gov).

Published online 24 July 2005; doi:10.1038/nsmb970

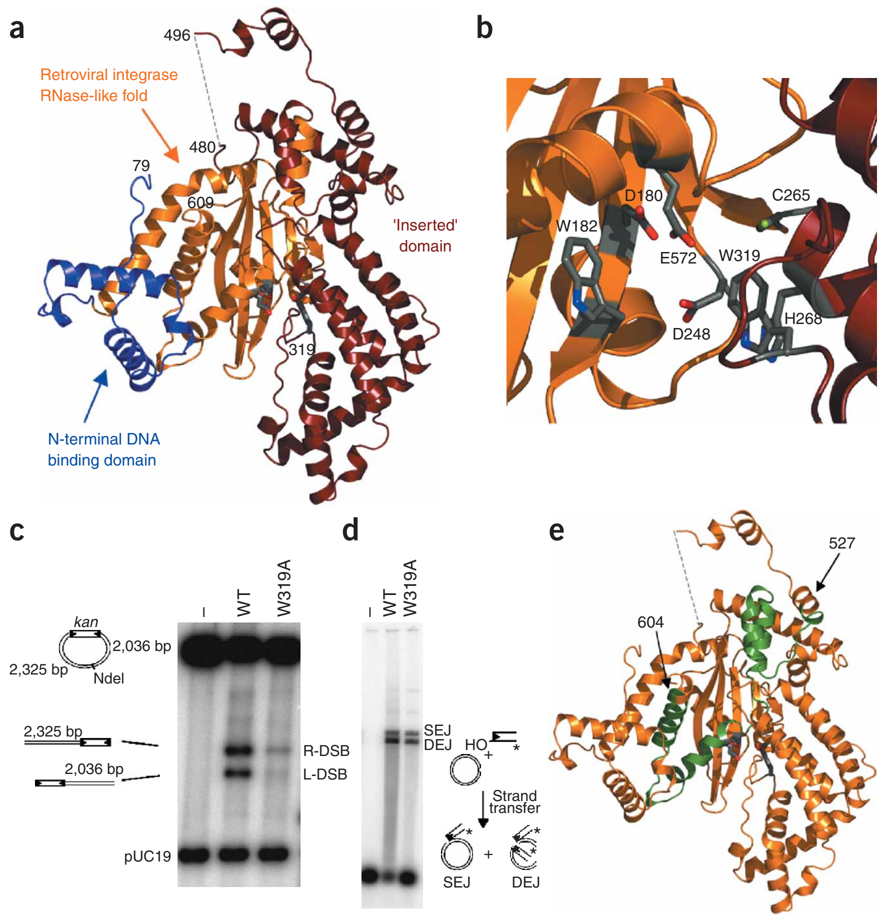

Figure 1 Structure of Hermes$_{79–612}$. (**a**) Three-domain organization of Hermes$_{79–612}$. Residues that comprise the DDE motif and Trp319 converge at the active site. The dashed line indicates a disordered region. (**b**) Close-up of the active site. (**c**) Plasmid cleavage assay of wild-type (WT) and W319A mutant Hermes$_{79–612}$, visualized on a 1% agarose gel. R-DSB, right-end double-stand break; L-DSB, right-end double-stand break. (**d**) Target-joining assay using 40-nucleotide, precleaved *Hermes* left end. SEJ, single-end join; DEJ, double-end join. (**e**) The *hAT* dimerization domain (green) winds through the inserted domain and forms the C-terminal portion of the retroviral integrase RNase-like domain.

removal of the N-terminal 78 residues results in a version of Hermes that readily crystallized[11]. The structure of Hermes$_{79–612}$ was solved using X-ray crystallography with multiwavelength anomalous dispersion on selenomethionyl-substituted protein (see Methods and **Supplementary Fig. 1** online). Hermes$_{79–612}$ consists of three domains: an N-terminal domain (residues 79–150); a catalytic domain with a retroviral integrase–like fold[12]; and a large, meandering, all-α-helical domain (residues 265–552) inserted into the catalytic domain after the final β-strand of its central β-sheet (**Fig. 1a**). Three catalytically essential residues[8]—Asp180, Asp248 and Glu572—are in close proximity and suitably arranged (**Fig. 1b**) to coordinate the catalytically required Mg^{2+} ions. The locations of these essential residues on the appropriate secondary structure elements place Hermes—and presumably all other *hAT* transposases—within the retroviral integrase, or DDE, transposase family[13].

The N-terminal domain is likely to be the site-specific DNA-binding domain responsible for recognizing transposon ends. An N-terminally truncated version of Hermes (residues 146–612) did not bind DNA. In contrast, Hermes$_{79–612}$ bound a 30-nucleotide fragment of the *Hermes* left end but did not bind nonspecific DNA (**Supplementary Fig. 2** online). The missing 78 residues did not seem to be crucial for DNA binding or catalysis, as Hermes$_{79–612}$ was as active as full-length Hermes in *in vitro* assays of hairpin formation and target joining (data not shown). Rather, residues 1–78 are important for nuclear localization[14] and have been proposed to contain a Zn-binding BED domain[15] that may contribute to nonspecific DNA binding.

Neither the N-terminal domain (residues 79–150) nor the all-α-helical inserted domain has any known three-dimensional homologs[16]. The inserted domain disrupts the retroviral integrase fold in the same location as does the four-β-strand insertion in the *E. coli* Tn5 DDE transposase[17], yet the two insertions are structurally unrelated. A role for the Hermes inserted domain is suggested by the structure of the Tn5 transposase synapsed with two transposon ends, which shows that residues on the inserted β-strands are important for binding the DNA hairpin that forms on each *Tn5* transposon end[17]. A crucial feature of *Tn5* hairpin formation and resolution is a stacking interaction between Trp298 and a flipped-out base at the −2 position[18]. Notably, the inserted domain of Hermes projects a highly conserved tryptophan residue, Trp319, into the enzyme active site (**Fig. 1a,b**). *In vitro* assays showed that Trp319 is required for one or more steps before target joining, as the W319A mutant was defective for either cleavage or subsequent hairpin formation (**Fig. 1c**) but formed single- and double-ended joined products when provided with precleaved transposon ends (**Fig. 1d**). Taken together with the data for the *Tn5* transposase, our findings strongly suggest that Trp319 participates in binding and stabilizing the flanking-end DNA hairpin.

Only a handful of residues are strictly conserved among the transposases of active *hAT* elements[8]. From the Hermes structure, these seem to be either directly involved in catalysis or DNA binding, by virtue of their location near the active site, or important for protein folding. The most highly conserved region of *hAT* transposases is located at the C terminus[5,19] and is designated the *hAT* family dimerization domain (Conserved Domain Database pfam 05699)[20]. In Hermes, this region (residues 527–604) bridges the inserted domain and the C-terminal portion of the disrupted retroviral integrase fold and consists of amino acids from each (**Fig. 1e**). The structure indicates that '*hAT* domain' is a misnomer, as residues 527–604 do not form an independent folding unit. Instead, the conserved residues have a crucial role in knitting the protein domains together into a functional whole.

Structural relationship of Hermes and RAG1

It has been proposed that the RAG1 and RAG2 proteins, which catalyze the initial steps of V(D)J recombination, originated from transposons[21]. The genes encoding both RAG proteins are intronless, and the mechanism by which they introduce double-strand breaks during V(D)J recombination is chemically similar to that of

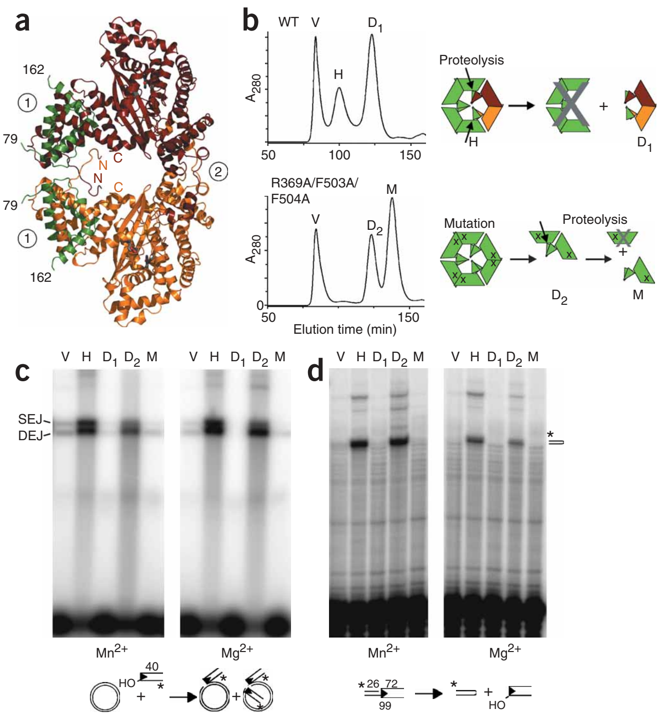

Figure 2 Hermes$_{79-61g}$ dimerizes through domain swapping. (**a**) The crystallographic asymmetric unit contains a heterodimer (two heterodimers are shown) consisting of Hermes$_{79-612}$ (red/orange) bound to Hermes$_{79-162}$ (green). The N-terminal domains dimerize through interface 1 (indicated by circled '1'), and the heterodimers in turn dimerize into heterotetramers through domain swapping at interface 2 (circled '2'). (**b**) Size-exclusion chromatograms for wild-type (WT) Hermes$_{79-612}$ (top) and a triple mutant (bottom; marked with 'X' in the schematic drawings to the right) in which interface 2 is disrupted. The first peak (V) represents material in the void volume, H corresponds to a hexamer, and D_1 is the heterotetramer. Although D_1 and D_2 have similar elution times, differences in molecular mass were confirmed by sedimentation equilibrium measurements (data not shown). Proteolytic cleavage sites (arrows in the schematic drawings) were identified by N-terminal sequencing, and the presence or absence of Hermes$_{79-162}$ was established by SDS-PAGE. Species crossed out in the schematics were not detected, presumably because of instability or insolubility during protein expression or purification. (**c**) Hermes$_{79-612}$ is active as a hexamer and as a dimer when interface 2 is perturbed by mutation. Shown are target-joining assays using a 40-nucleotide, precleaved *Hermes* left end. Different forms of Hermes$_{79-612}$ (V, H, D_1, D_2 and M) correspond to protein species separated by size-exclusion chromatography. The activity shown by the H species is comparable to that of wild-type full-length Hermes (data not shown). Single-end join (SEJ) and double-end join (DEJ) products were visualized on a native 1% agarose gel. $MnCl_2$ or $MgCl_2$ (1 mM) was included as indicated. (**d**) DNA hairpin formation using a prenicked flank and a *Hermes* left-end substrate. Products were visualized on a 5% urea acrylamide gel.

transposases[22]. Secondary-structure prediction[23] for RAG1, together with the identification of three catalytically essential acidic residues[24,25], suggest that it—like Hermes—contains a retroviral integrase–like scaffold into which an all-α-helical domain is inserted[8]. Notably, the predicted all-α-helical domain of RAG1 seems to be inserted into the catalytic domain at a site that corresponds to that of Hermes and Tn5. In the primary sequence of RAG1, Trp893 seems to be positioned appropriately in the predicted inserted domain to have a role in hairpin formation[8,26], and it seems likely that Trp893 will similarly project into the RAG1 active site.

The RAG1 core (residues 384–1008) consists of two topologically independent structural domains[27] that can be broadly mapped onto the Hermes structure (data not shown). The RAG1 central domain (residues 528–760) corresponds to the retroviral integrase–like fold of Hermes, which is preceded by an α-helix (Hermes residues 144–171) and followed by the first two α-helices of the Hermes inserted domain (Hermes residues 265–300). The RAG1 C-terminal domain (residues 761–979) corresponds to the rest of the inserted domain followed by the α-helix bearing the catalytically essential glutamate residue that continues the retroviral integrase fold. The Hermes structure therefore provides an explanation for the reported organization of the RAG1 core: one domain is essentially the retroviral integrase fold and the second corresponds to the inserted domain.

The asymmetric unit contains a heterodimer

In the crystallographic asymmetric unit, Hermes$_{79-612}$ formed a heterodimer (**Fig. 2a**) in which one molecule (shown in red) was bound to an N-terminal domain fragment (residues 79–162; in green) that was presumably generated by the inadvertent proteolytic cleavage of a Hermes multimer during protein expression. Attempts to separate the two chains using a variety of biochemical approaches (short of denaturing the protein) were unsuccessful. The structure of Hermes$_{79-612}$ provides an explanation for this observation, as the two DNA-binding domains form a highly intertwined, all-helical dimer with a tightly packed and entirely hydrophobic core. Measurement with a 1.8-Å radius probe revealed that 5,450 Å^2 is buried in this dimer interface (interface 1, **Fig. 2a**), whereas the accessible surface of an isolated Hermes$_{79-150}$ dimer is only 5,840 Å^2. A number of intertwined dimers are known[28], but to our knowledge, such a close relationship between the sizes of the buried and accessible dimer surface areas has not been previously observed. The importance of an N-terminal region for multimerization has been demonstrated by yeast two-hybrid studies, which showed that multimerization of N-terminally truncated Hermes (residues 253–612) is abolished by C-terminal point mutations, yet the same mutations have no effect on the self-association of full-length Hermes[29].

An explanation for the multimerization of Hermes$_{253-612}$ is provided by the presence of a second interface (interface 2) through which heterodimers form heterotetramers (**Fig. 2a**). This interface arises by domain swapping of two helices between residues 497 and 516 that project away from each Hermes$_{79-612}$ molecule and fit into a predominantly hydrophobic socket of the adjacent molecule. The two swapped helices from each molecule bury only ~1,160 Å^2 of accessible surface on the adjacent molecule. Flexible linkers (residues 481–496 and 517–520) join the two swapped helices to the rest of the molecule. The first linker is disordered and represents the only portion of Hermes for which we did not detect measurable electron density.

Hermes forms hexamers

In our size-exclusion chromatography experiments, recombinantly expressed Hermes$_{79-612}$ formed two oligomeric species (**Fig. 2b**, top), one that eluted at a position consistent with a hexamer (H) and a smaller species with a molecular weight consistent with a heterotetramer (D_1). These purified species were not interconvertible or in equilibrium, nor did their elution profiles vary with protein concentration, consistent with the interpretation that the smaller form is a degradation product of the larger one. Sedimentation equilibrium measurements (**Supplementary Fig. 3** online) confirmed that the larger species corresponds to a monodisperse hexamer and that the

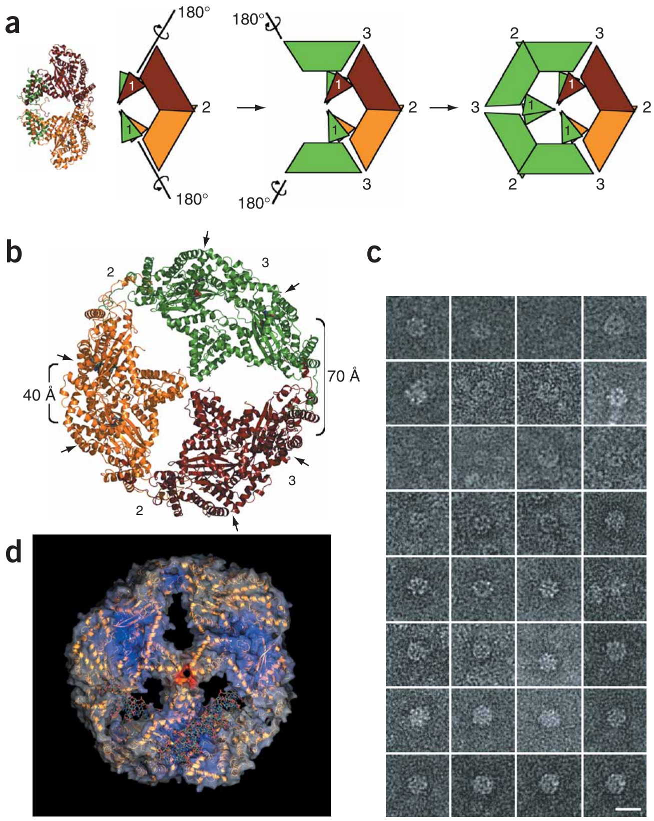

Figure 3 The hexameric form of Hermes$_{79–612}$. (**a**) A model for the hexamer was generated using the symmetry elements of the heterotetramer. A noncrystallographic two-fold axis relates the N-terminal domains, and a crystallographic two-fold axis relates the Hermes$_{79–612}$ monomers. The numbers refer to three interfaces, where interfaces 1 and 2 are observed and interface 3 is modeled. (**b**) Modeled hexamer showing alternating interfaces 2 and 3. Six arrows indicate active sites where Asp180, Asp248 and Glu572 converge (carboxylate oxygen atoms are shown in red). (**c**) Gallery of electron micrographs of negatively stained active Hermes$_{79–612}$. Scale bar, 20 nm. (**d**) Surface representation of Hermes$_{79–612}$ showing one possible mode of DNA binding (DNA is shown as ball-and-stick model and Hermes$_{79–162}$ as ribbons).

smaller species has an experimental molecular mass (∼131 kDa) consistent with the crystallographically observed heterotetramer.

In vitro activity assays revealed that, of the two separable oligomers of Hermes$_{79–612}$, only the hexamer was active (**Fig. 2c,d**, lanes H and D$_1$), although we cannot rule out the possibility that it forms assemblies other than hexamers under assay conditions. We used the two-fold symmetry axes of the heterotetramer to generate a structural model of the hexamer (**Fig. 3a**). Strict application of the symmetry operators resulted in a spiral of six Hermes$_{79–612}$ monomers, with dangling N-terminal Hermes$_{79–162}$ domains on the terminal monomers. Because the structure suggests that the N-terminal domains are obligate dimers, we brought the two unpaired N-terminal domains together by applying an ∼10° rotation to one monomer across each crystallographically observed interface 2, thereby flattening and sealing the ring. This seemed justifiable because interface 2 is bordered by flexible linkers. The resulting hexamer was ∼15 nm in diameter and had alternating interfaces (2, 3, 2, 3, 2, 3; **Fig. 3b**), where interface 2 was observed in the structure and interface 3 was modeled. This trimer-of-dimers arrangement had three-fold, rather than six-fold, symmetry. Direct electron microscopic images of the biochemically active Hermes$_{79–612}$ species (**Fig. 3c**) showed round assemblies ∼16 nm in diameter, consistent with the model. This hexameric assembly is not unique to truncated Hermes$_{79–612}$, as full-length Hermes expressed in Sf9 insect cells also appears to form hexamers, as suggested by size-exclusion chromatography (data not shown).

Two specific structural features of the modeled hexamer suggested why only the hexameric form of Hermes$_{79–612}$ was active. When *hAT* elements integrate into their target sites, the two transposon ends are inserted 8 bp apart; this implies that the two active sites that catalyze transesterification should be less than ∼30 Å apart. In the heterotetramer, the active sites are ∼70 Å apart (**Fig. 3b**), whereas in the modeled hexamer, two other active sites—represented by the two molecules across interface 3—are separated by only ∼40 Å (measured as a straight line). This suggests that the latter pair of active sites carries out the target-site reactions, perhaps accompanied by a conformational change upon target binding to reconcile the ∼10 Å discrepancy. Conformational flexibility may be a general feature of transposases, as two Tn5 transposase active sites that catalyze target-site insertions spaced 9 bp apart are observed to be separated by ∼41 Å before target binding[17]. In addition, ∼30 Å separates the 3′-OH ends of the MuA transpososome[30], which inserts with a target spacing of 5 bp.

The second observation that suggests why the hexamer was active is that only the hexamer showed continuous positive surface potential connecting the N-terminal DNA-binding domains to the active sites (**Fig. 3d**). This was consistent with our observation that a 17-bp oligonucleotide representing the *Hermes* left end bound specifically to the hexamer but not to the heterotetramer (data not shown). The hexamer model generated three channels (∼22 Å in diameter in the center) lined with basic residues that penetrated the ring and passed

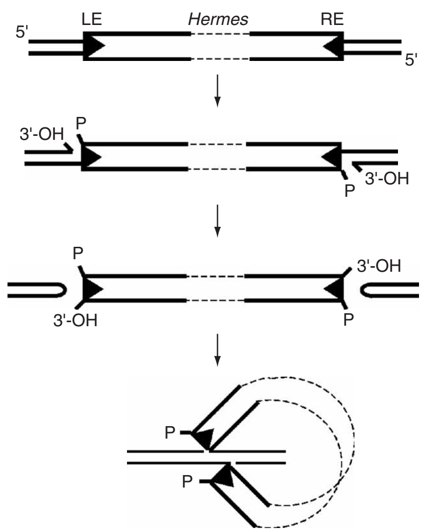

Figure 4 Schematic of the mechanism of *Hermes* transposition. Initial cleavage at the left ends (LE) and right ends (RE) of the *Hermes* element occurs one nucleotide into the flanking strand at the 5′ ends of the transposon, generating a flanking 3′-OH group. Subsequent nucleophilic attack by this 3′-OH group on the opposite strand results in a flanking hairpin and a 3′-OH group at the 3′ end of the transposon. The two new 3′-OH groups become the nucleophiles for coordinated attack on target DNA, in which the two insertion events occur on opposite strands and are separated by 8 bp.

over the active sites. We speculate that DNA might bind within these channels (**Fig. 3d**), requiring interface 2 to transiently open and close; this seems possible given the small size of the interface.

To show that interface 2 is present in hexamers in solution, we introduced three point mutations (R369A, F503A and F504A) designed to disrupt interface 2. Size-exclusion chromatography of the triple-mutant protein revealed only dimers and a smaller species corresponding to a $Hermes_{79–612}$ monomer bound to a $Hermes_{79–162}$ fragment (**Fig. 2b**, bottom). Thus, interface 2 is crucial for hexamer formation. Notably, the region corresponding to Hermes residues 460–510, which includes important residues within interface 2, is one of the least conserved regions of *hAT* transposases. It is possible that the role of residues in this region is to ensure that a given *hAT* transposase can only form higher-order multimers with itself, a notion supported by the report that Hermes does not bind Activator in a yeast two-hybrid assay[29].

The $Hermes_{79–612}$ triple mutant in which interface 2 was eliminated showed robust *in vitro* target joining and hairpin formation (**Fig. 2c,d**, lane D_2). However, in a *Saccharomyces cerevisiae in vivo* assay, the $Hermes_{1–612}$ triple mutant was less active than wild-type $Hermes_{1–612}$ by a factor of at least 20 (P. Bafuma, J. Fain-Thornton and N.L.C., unpublished data). This suggests that although two Hermes monomers joined by interface 3 represent the basic catalytic unit, the hexamer is important for functions that are not well recapitulated in *in vitro* assays. One inherent difference between the assays is that transposon ends are within chromosomes in cells, whereas they are encountered as oligonucleotides *in vitro*; it is possible that the hexamer is required to latch onto chromosomal DNA whereas a dimer is sufficient to capture oligonucleotides. Another possibility is that the hexamer might slide along chromosomal DNA searching for transposon end sequences, an unnecessary function in *in vitro* assays with purified and plentiful oligonucleotide substrates.

DISCUSSION

The formation of hexameric assemblies by Hermes results in more active sites than substrate DNA ends. This is not unprecedented in transposition. For example, the MuA transposase of bacteriophage Mu is active as a tetramer, and although all four monomers contribute to DNA binding, only two active sites are needed for catalysis[31,32]. There has been no previous suggestion of hexamers in transposition systems; however, Hermes, representing the *hAT* superfamily, and the RAG1 protein of V(D)J recombination are mechanistically distinct from other recombinases. The well-studied bacterial transposases and retroviral integrases use the same 3′-OH group that is generated upon end cleavage as the attacking nucleophile during target joining; thus, it is reasonable that the same active site could be used for both reactions[33].

In contrast, the observation that *Hermes* transposition involves a hairpin intermediate on the flanking DNA[8] implies that Hermes must act on the two DNA strands during different steps of the transposition pathway. After initial cleavage on the top strand at the transposon's 5′ end, the liberated 3′-OH on the flanking DNA forms a hairpin by attacking the bottom strand[8] (**Fig. 4**). This potential nucleophile is therefore no longer available to participate in further reactions. However, hairpin formation generates a new 3′-OH on the bottom strand of the transposon end, and this group becomes activated for a nucleophilic attack on target DNA (**Fig. 4**). Thus, sequential transesterification reactions occur diagonally across what becomes the double-strand break in the chromosome, and either a single active site must switch from one DNA strand to the other or two closely spaced active sites must be involved. Our model of the hexamer structure supports only the former possibility, as all the adjacent pairs of active sites are separated by substantial distances (**Fig. 3b**). An appealing aspect of the model in which DNA binds within the channels is that this might prevent the escape of DNA while a single active site switches strands.

Mobile elements have had enormous roles in genome evolution[1]. Approximately 50% of the sequence of the human genome originates from mobile elements[2]. Although retrotransposons are the most prevalent, DNA transposons contribute ~3% of the human genome; of these, the *hAT* transposons are the most highly represented[2]. It is believed that all the human *hAT* transposons are currently inactive[2,34], a suggestion worth revisiting now that the Hermes structure has provided us with knowledge of its crucial catalytic residues. Rather than being mutated into silence, several *hAT* transposases have been co-opted into the human proteome: at least 26 human genes have been identified as having originated from *hAT* elements[2]. The Hermes structure provides a foundation from which to begin evaluating their functions and may provide other avenues of investigation into the processes of genome evolution.

METHODS

Proteins. $Hermes_{79–612}$, $Hermes_{146–612}$ and the $Hermes_{79–612}$ R369A/F503A/F504A triple mutant were cloned into pET-15b, expressed in *E. coli* and purified as His-tagged proteins[11]. Selenomethionyl-substituted $Hermes_{79–612}$ (with a single point mutation of S163G) was obtained by transformation into *E. coli* B834(DE3) grown in modified minimal medium supplemented with selenomethionine. Crystals of the S163G mutant and the native version of $Hermes_{79–612}$ were grown under conditions identical to those previously described[11]. The full-length protein and W319A point mutant were expressed and purified as described[8].

Crystallization and data collection. X-ray diffraction data were collected at the Southeast Regional Collective Access Team beamline ID22 of the Advanced Photon Source. One selenomethionyl-substituted crystal was used at three different X-ray energies around the Se-K absorption edge using a MAR225 mosaic charge-coupled device detector in 0.5° oscillation frames (**Table 1**). To minimize systematic errors from radiation damage, data were collected in 15° wedges at the inflection point of the absorption curve and the remote energy and in an 'inverse beam' setting at the absorption curve peak. Data were integrated and scaled internally using the HKL suite[35]. Diffraction data used in refinement were collected in 0.5° oscillation frames on a native crystal using a RU200 rotating anode source equipped with multilayer focusing optics and a RAXIS IV imaging plate detector. All diffraction data were collected at 95 K.

Structure determination and refinement. The Se positions (ten per asymmetric unit) were located with SHELXD[36] using Fa coefficients computed by XPREP (Bruker-AXS), and their positions and thermal parameters were optimized with phase-integrating least-squares as implemented in PHASES[37]. The solvent-flattened experimental electron density map calculated at 2.5-Å resolution was manually interpreted using O[38]. The resulting structure was refined with several rounds of simulated annealing, energy minimization and restrained individual B-factor refinement at 2.1-Å resolution (**Table 1**) using the OpenMP version of CNX2002[39]. At the end of this process, 416 water molecules were also included. The most recent model contained all residues between Gln79 and Lys609 except for a disordered region between Ser481 and Lys496. The Ramachandran plot of the final model showed 93.1% of all residues in the most favored region and none in the disallowed region. The figures were either prepared with SPOCK[40] and ray-traced with Povray[41] (http://www.povray.org) or prepared with Pymol[42] (http://pymol.sourceforge.net).

Size-exclusion chromatography. Preparative-scale size-exclusion chromatography was carried out at 4 °C on a TSK-Gel G3000SW column (TosoHaas) equilibrated in 20 mM Tris (pH 7.5), 1 mM EDTA, 0.5 M NaCl, 5 mM DTT and 10% (w/v) glycerol. Samples were typically loaded at a protein

Table 1 Data collection, phasing and refinement statistics

	Native	Crystal 1		
Data collection				
Space group	*C*2	*C*2		
Cell dimensions				
a, *b*, *c* (Å)	116.3, 84.9, 73.8	116.4, 84.9, 73.9		
α, β, γ (°)	90, 93.8, 90	90, 93.7, 90		
		Peak	Inflection	Remote
Wavelength (nm)	0.154	0.097942	0.097949	0.096863
Resolution (Å)	2.1	2.0	2.0	2.0
R_{sym}[a]	0.060 (0.271)	0.085 (0.244)	0.084 (0.174)	0.076 (0.171)
$I / \sigma I$[a]	11.6 (4.29)	12.5 (4.2)	12.4 (6.1)	14.5 (7.0)
Completeness (%)[a]	99.8 (99.9)	100.0	100.0	100.0
Redundancy[a]	3.71 (3.53)	7.18 (3.80)	3.85 (3.82)	3.85 (3.79)
Refinement				
Resolution (Å)	30.0–2.1			
No. reflections	41,879			
R_{work}/R_{free}	19.7/23.2			
No. atoms	5,174			
Protein	4,758			
Water	416			
B-factors				
Protein	36.7			
Water	40.6			
R.m.s. deviations				
Bond lengths (Å)	0.006			
Bond angles (°)	1.48			

[a]Highest-resolution shell is shown in parentheses (2.05–2.00 Å for selenium energies and 2.16–2.10 Å for the native data set).

concentration of 8–10 mg ml^{-1}. To estimate the molecular weight of the eluted species, samples were reinjected onto a Superose 6 or Superdex 200 column (Amersham Biosciences) calibrated using protein standards (BSA, 66 kDa; amylase, 200 kDa; urease, 272 kDa; apoferritin, 443 kDa; and thyroglobulin, 669 kDa).

DNA binding assays. PAGE-purified oligonucleotides (L30, 5′-CAGAGAACAACAACAAGTGGCTTATTTTGA-3′ (top); random 29-mer, 5′-CCTCTCTGCGCGCTCGCTCGCTCACTGAG-3′ (top)) were obtained from Integrated DNA Technologies. The oligonucleotides were annealed and added to either the hexameric form of Hermes$_{79–612}$ (at 23 μM) or Hermes$_{146–612}$ (at 7.6 μM) at a 1:1 molar ratio of protein to DNA in buffer containing 0.5 M NaCl. The solution was then dialyzed into 20 mM Tris (pH 7.5), 0.2 M NaCl and 5 mM DTT, and 50 μl was applied to a Superdex 200 column equilibrated at 4 °C in the same buffer and run at 50 μl min^{-1}. Binding was assessed by the ability to form a complex stable enough to persist under size-exclusion chromatography conditions.

Transposition assays. Target-joining and hairpin-formation assays were performed as described[8]. In all cases, 140 nM Hermes was incubated with the appropriate substrates for 2 h at 30 °C before quenching. For the plasmid cleavage assay (**Fig. 1c**), a donor plasmid was constructed by inserting a kanamycin resistance gene, flanked by 30 bp of the *Hermes* left terminal inverted repeat and 30 bp of the right one, into pBR322. This donor plasmid was then incubated with pUC19 and full-length Hermes or the W319A point mutant for 2 h at 30 °C. Target and donor DNA was purified by phenol extraction and then linearized with NdeI. DNA was visualized by agarose gel electrophoresis and Southern blotting with a kanamycin-specific probe.

Sedimentation equilibrium. Sedimentation equilibrium experiments were conducted at 4 °C on a Beckman Optima XL-A analytical ultracentrifuge. Samples (loading volume of 125 μl) in 20 mM Tris (pH 7.5), 0.5 M NaCl, 5% (v/v) glycerol, 1 mM EDTA and 2 mM 2-mercaptoethanol were studied at rotor speeds ranging from 4,000 to 12,000 r.p.m. using six-channel centerpiece cells and data were scanned from a minimum radius of 5.75 cm and a maximum radius of 7.25 cm. Data were acquired as an average of 16 absorbance measurements at a wavelength of 280 nm and a radial spacing of 0.001 cm. In all cases, equilibrium was achieved within 66 h.

Data for the heterotetramer collected at a loading A_{280} of 0.8 (that is, ~0.8 mg ml^{-1}) and 6,000, 8,000 and 10,000 r.p.m. were analyzed globally in terms of a single ideal solute, yielding a buoyant molecular mass of 31,780 Da. This corresponds to a molecular mass of ~131 kDa, which is consistent with the calculated molecular weight of the observed heterotetramer (142 kDa). Improved data fits were obtained when data were analyzed in terms of two non-interacting ideal solutes. In addition to the heterotetramer, a hexameric species was present at 2% concentration.

Two batches of the Hermes$_{79–612}$ hexameric species were characterized at loading concentrations corresponding to A_{280} values of approximately 0.15, 0.30 and 0.60. Data for these species were analyzed both individually and globally using SEDPHAT 3.0 (http://www.analyticalultracentrifugation.com/sedphat/) in terms of a single ideal solute to obtain the buoyant molecular mass, $M(1 - \nu\rho)$. Protein molecular masses were determined using the calculated value of ν based on the amino acid composition and the experimentally determined solution density. Sedimentation equilibrium experiments carried out at 4,000, 6,000 and 8,000 r.p.m. showed that the samples were monodisperse, with identical values of $M(1 - \nu\rho)$ obtained irrespective of rotor speed and sample loading concentration. A global analysis of data collected at all rotor speeds and all loading concentrations returned a molecular mass of 364,650 Da, consistent with a hexameric Hermes$_{79–612}$ species ($n = 5.98$). An independent analysis of each sample at all rotor speeds resulted in an average molecular mass of 362,050 ± 6,300 Da ($n = 5.94 \pm 0.10$), confirming the presence of a hexameric species and providing an estimate of the error in the determination of the molecular mass.

Negative-stain electron microscopy. To obtain negatively stained images, samples were diluted to ~0.015–0.6 mg ml^{-1}, absorbed to an electron microscope grid, washed and stained with 2% uranyl acetate. Images were taken at ×35,000 and ×45,000 magnification using a Philips CM120 transmission electron microscope (FEI) operating at 100 kV with a focus range of 0.5–1.0 μm under focus. Images were recorded digitally on a Gatan 794 MultiScan charge-coupled device camera with the DigitalMicrograph software package (Gatan Inc.).

Accession codes. PDB: Coordinates have been deposited (accession code 2BW3).

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS
We thank M. Gellert, S. Vasudevan and D. Ronning for helpful discussions and comments on the manuscript. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Basic Energy Sciences, Office of Science, under contract no. W-31-109-Eng-38. This study used the high-performance computational capabilities of the Helix Systems at the National Institutes of Health, Bethesda, MD (http://helix.nih.gov). 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.

Received 20 April; accepted 11 July 2005
Published online at http://www.nature.com/nsmb/

1. Kazazian, H.H. Mobile elements: drivers of genome evolution. *Science* **303**, 1626–1632 (2004).
2. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
3. Ostertag, E.M. & Kazazian, H.H., Jr. Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* **35**, 501–538 (2001).
4. Plasterk, R.H.A. & van Luenen, H.G.A.M. In *Mobile DNA II* (ed. Craig, N.L., Craigie, R., Gellert, M. & Lambowitz, A.M.) 519–532 (ASM Press, Washington, DC, USA, 2002).
5. Calvi, B.R., Hong, T.J., Findley, S.D. & Gelbart, W.M. Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: hobo, Activator, and Tam3. *Cell* **66**, 465–471 (1991).
6. McClintock, B. Chromosome organization and genic expression. *Cold Spring Harb. Symp. Quant. Biol.* **16**, 13–47 (1951).
7. 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).
8. Zhou, L. *et al.* Transposition of *hAT* elements links transposable elements and V(D)J recombination. *Nature* **432**, 995–1001 (2004).
9. 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 thymocytes. *Cell* **70**, 983–991 (1992).
10. McBlane, J.F. *et al.* Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* **83**, 387–395 (1995).
11. Perez, Z.N., Musingarimi, P., Craig, N.L., Dyda, F. & Hickman, A.B. Purification, crystallization, and preliminary crystallographic analysis of the Hermes transposase. *Acta Crystallogr.* **F61**, 587–590 (2005).
12. Dyda, F. *et al.* Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266**, 1981–1986 (1994).
13. Curcio, M.J. & Derbyshire, K.M. The outs and ins of transposition: from Mu to kangaroo. *Nat. Rev. Mol. Cell Biol.* **4**, 865–877 (2003).
14. Michel, K. & Atkinson, P.W. Nuclear localization of the Hermes transposase depends on basic amino acid residues at the N-terminus of the protein. *J. Cell. Biochem.* **89**, 778–790 (2003).
15. Aravind, L. The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. *Trends Biochem. Sci.* **25**, 421–423 (2000).
16. Holm, L. & Sander, C. Dali/FSSP classification of three-dimensional protein folds. *Nucleic Acids Res.* **25**, 231–234 (1997).
17. Davies, D.R., Goryshin, I.Y., Reznikoff, W.S. & Rayment, I. Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **289**, 77–85 (2000).
18. Ason, B. & Reznikoff, W.S. Mutational analysis of the base flipping event found in Tn5 transposition. *J. Biol. Chem.* **277**, 11284–11291 (2002).
19. Feldmar, S. & Kunze, R. The ORFa protein, the putative transposase of maize transposable element *Ac*, has a basic DNA binding domain. *EMBO J.* **10**, 4003–4010 (1991).
20. Marchler-Bauer, A. *et al.* CDD: a conserved domain database for protein classification. *Nucleic Acids Res.* **33**, D192–D196 (2005).
21. Thompson, C.B. New insights into V(D)J recombination and its role in the evolution of the immune system. *Immunity* **3**, 531–539 (1995).
22. van Gent, D.C., Mizuuchi, K. & Gellert, M. Similarities between initiation of V(D)J recombination and retroviral integration. *Science* **271**, 1592–1594 (1996).
23. Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).
24. 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).
25. 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).
26. 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).
27. De, P. & Rodgers, K.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).
28. Larsen, T.A., Olson, A.J. & Goodsell, D.S. Morphology of protein-protein interfaces. *Structure* **6**, 421–427 (1998).
29. 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).
30. Yuan, J.F., Beniac, D.R., Chaconas, G. & Ottensmeyer, F.P. 3D reconstruction of the Mu transposase and the Type 1 transpososome: a structural framework for Mu DNA transposition. *Genes Dev.* **19**, 840–852 (2005).
31. Namgoong, S.Y. & Harshey, R.M. 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).
32. 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).
33. 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).
34. Smit, A.F.A. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9**, 657–663 (1999).
35. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
36. Sheldrick, G.M. SHELX: applications to macromolecules. in *Direct Methods for Solving Macromolecular Structures* (S. Fortier, ed.) 401–411 (Kluwer, Dordrecht, The Netherlands, 1998).
37. Furey, W. & Swaminathan, S. PHASES-95: a program package for processing and analyzing diffraction data from macromolecules. *Methods Enzymol.* **277**, 590–620 (1997).
38. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* **A47**, 110–119 (1991).
39. Brünger, A.T. *et al.* Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr.* **D54**, 905–921 (1998).
40. Christopher, J.A. SPOCK The Structural Properties Observation and Calculation Kit (Program Manual) (The Center for Macromolecular Design, Texas A&M University, College Station, Texas, USA, 1998).
41. Persistence of Vision Pty. Ltd. *Persistence of Vision Raytracer* (Persistence of Vision Pty. Ltd., Williamstown, Victoria, Australia, 2004).
42. DeLano, W.L. *The PyMol Molecular Graphics System* (DeLano Scientific, San Carlos, California, USA, 2002).