Mechanism of Mos1 transposition: insights from structural analysis

Julia M Richardson, Angela Dawson,
Natasha O'Hagan, Paul Taylor, David
J Finnegan and Malcolm D Walkinshaw*

School of Biological Sciences, University of Edinburgh, Edinburgh, UK

We present the crystal structure of the catalytic domain of *Mos1* transposase, a member of the Tc1/*mariner* family of transposases. The structure comprises an RNase H-like core, bringing together an aspartic acid triad to form the active site, capped by N- and C-terminal α -helices. We have solved structures with either one Mg^{2+} or two Mn^{2+} ions in the active site, consistent with a two-metal mechanism for catalysis. The lack of hairpin-stabilizing structural motifs is consistent with the absence of a hairpin intermediate in *Mos1* excision. We have built a model for the DNA-binding domain of *Mos1* transposase, based on the structure of the bipartite DNA-binding domain of Tc3 transposase. Combining this with the crystal structure of the catalytic domain provides a model for the paired-end complex formed between a dimer of *Mos1* transposase and inverted repeat DNA. The implications for the mechanisms of first and second strand cleavage are discussed.

The EMBO Journal (2006) 25, 1324–1334. doi:10.1038/sj.emboj.7601018; Published online 2 March 2006

Subject Categories: genome stability & dynamics; structural biology

Keywords: eukaryotic transposition; Tc1/*mariner* transposon; two-metal active site binding; transposase

Introduction

Transposases catalyse the movement of fragments of DNA from one place in a genome to another. In doing so, they generate mutations, rearrange chromosomes and change the DNA content of genomes. Transposition occurs via a series of hydrolysis and *trans*-esterification reactions in a mechanism related to the integration of retroviruses such as HIV-1 and the rearrangement of immunoglobulin genes mediated by the RAG1/2 recombinase (Dawson and Finnegan, 2003; Zhou *et al.*, 2004).

Transposition involves an ordered series of events: (1) sequence-specific binding of transposase to the terminal inverted repeats (IRs) present at the ends of the transposon, (2) cleavage of both strands of DNA at each end of the transposon, (3) synapsis of the ends by transposase–transposase interactions, (4) capture of the target DNA and (5) strand transfer to insert the element into the target.

*Corresponding author. School of Biological Sciences, University of Edinburgh, The King's Buildings, Mayfield Road, Edinburgh EH9 3JR, UK. Tel.: +44 131 650 7056; Fax: +44 131 650 7055; E-mail: malcolm.walkinshaw@ed.ac.uk

Received: 22 July 2005; accepted: 1 February 2006; published online: 2 March 2006

Transposases are members of the retroviral integrase superfamily of proteins. Despite the structural similarities in their catalytic domains (Rice and Baker, 2001), these proteins carry out phosphoryl transfer reactions with different specificities. Some cleave only one strand of DNA, while RNase H cleaves one strand of RNA in an RNA:DNA hybrid duplex. Others generate double-strand DNA breaks, and a variety of mechanisms are employed. The transposases of the bacterial transposons Tn5 and Tn10 carry out first-strand cleavage by hydrolysis to form a 3' hydroxyl (3'OH) at each end of the element, while the second strand is cleaved by *trans*-esterification using this 3'OH as the attacking nucleophile. This forms a DNA hairpin at each end of the element, which is hydrolysed by transposase to regenerate the 3'OH required for strand transfer (Kennedy *et al.*, 1998; Bhasin *et al.*, 1999). V(D)J recombination and transposition of the eukaryotic element *Hermes*, a member of the hAT family, proceed by a similar mechanism, except that the order of strand cleavage is reversed and a hairpin is formed on the flanking, rather than on the excised, DNA (Roth *et al.*, 1992; Zhou *et al.*, 2004). Another bacterial transposon, Tn7, utilizes TnsB to perform first-strand cleavage and recruits a second protein, TnsA, to cleave the nontransferred strand (Hickman *et al.*, 2000).

Mos1 transposase catalyses the movement of the *mariner* transposon *Mos1*, first isolated from *Drosophila mauritiana* (Hartl, 2001), a member of the Tc1/*mariner* family of elements that are found in the genomes of most eukaryotes (Plasterk *et al.*, 1999). Elements of this type, including *Mos1* and Sleeping Beauty, are being used as tools for insertional mutagenesis and transgenesis in several species (Robinson *et al.*, 2004; Collier *et al.*, 2005; Dupuy *et al.*, 2005; Williams *et al.*, 2005), and, in the case of Sleeping Beauty (Ivics *et al.*, 2004), for gene therapy (Yant *et al.*, 2000). The *Mos1* transposon is 1.3 kb in length with one imperfect 28 base pair IR at each end. *Mos1* transposase is encoded by a single gene within the element and is the sole requirement for transposition *in vitro* (Lampe *et al.*, 1996). It contains 345 residues within two domains: the N-terminal DNA-binding domain (residues 1–112) and the C-terminal catalytic domain (126–345) joined by a short linker. The C-terminal catalytic domain contains a variant, DD(34)D, of the catalytic DD(35)E motif (Lohe *et al.*, 1997).

We have previously used *in vitro* assays with purified transposase expressed in *E. coli* to determine details of the mechanism by which *Mos1* transposase catalyses double-strand breaks at each end of the element (Dawson and Finnegan, 2003). The 5' end of the nontransferred strand is cut three bases within the element, whereas the 3' end of the transferred strand (TS) is cut precisely at the junction with the flanking DNA (Figure 1). *Mos1* transposase cleaves the nontransferred strand first, unlike the transposases of the bacterial elements Tn5 and Tn10. First- and second-strand cleavage are not tightly coupled: first-strand cleavage occurs within a single-end complex (SEC), before the two ends are brought together to form a paired-end complex (PEC). The

Figure 1 Schematic of the proposed mechanism of *Mos1* transposition. The left and right IRs of the *Mos1* transposon are indicated by green and orange triangles, respectively, and a transposase monomer is represented by a small and a large circle, denoting the N- and C-terminal domains, respectively. Initially, transposase monomer binds sequence-specifically to the IR, forming a SEC. Subsequently, the two ends of the transposon are brought together to form a PEC consisting of a dimer of transposase. A second SEC, composed of a dimer of transposase at one end, may be a precursor to the PEC, and it is yet to be determined if first-strand cleavage, 3 bases within the element, occurs within a SEC composed of a monomer or a dimer of transposase. Cleavage of the TS, to generate a 3'OH at each end of the element, may occur in *cis* or in *trans*, depending on the route of the linker between the transposase domains. Finally, target DNA is recruited (forming a target capture complex) and the transposon is inserted into target DNA, by attack of the 3'OH of the TS on the 5' side of a TA dinucleotide.

proposed pathway for *Mos1* transposition is summarized in Figure 1. Second-strand cleavage occurs within the PEC and takes place without formation of a hairpin intermediate, indicating that double-strand cleavage is the result of two sequential hydrolysis reactions. Target DNA is recruited to form a target capture complex in which strand transfer takes place (Dawson and Finnegan, 2003). Like all other Tc1/*mariner* elements, *Mos1* is inserted at the 5' end of a TA dinucleotide sequence, using the 3'OH on the TS as a nucleophile in a *trans*-esterification reaction. The resulting single-strand gap is repaired by host enzymes.

The structure of Tn5 transposase, complexed with transposon end DNA, has been solved and has given insight into the mechanism by which the DNA strands are cleaved in prokaryotes (Davies *et al*, 2000; Lovell *et al*, 2002). Recently, the structure of the eukaryotic hAT transposase Hermes has been determined (Hickman *et al*, 2005).

We present here the crystal structure of the catalytic domain of *Mos1* transposase. It contains an RNase H-like fold, capped by two additional (N- and C-terminal) helices.

We show that two divalent metal ions can bind in the active site, in the absence of DNA substrate, consistent with a two-metal mechanism for catalysis. We have used this structure, together with a model of the DNA-binding domain of *Mos1* (based on the structure of the bipartite DNA-binding domain of Tc3 transposase bound to DNA (Watkins *et al*, 2004)), to propose a structure for the PEC. Our results reveal the structural basis for DNA strand cleavages by *Mos1* transposase, in the absence of a hairpin intermediate.

Results and discussion

The C-terminal domain of *Mos1* transposase contains an RNase H-like catalytic core

The structure of the C-terminal domain of *Mos1* transposase (Figure 2A) has been determined to a resolution of 2.25 Å. The protein used was the full-length, catalytically active mutant T216A. This protein cleaved during crystallization (by an unknown mechanism) and crystals of the catalytic domain (residues 119–345) were obtained (Richardson *et al*, 2004). The structure contains a four-stranded mixed β -sheet flanked by four helices, which bring together the three residues of the DDD motif (D156, D249 and D284) to form the active site. The N- and C-terminal helices, formed by residues 126–144 (α 1) and 329–339 (α 9), respectively, cap this RNase-H like catalytic core and occupy a similar location to the insertion domain of Tn5 (Figure 2B). The N-terminal residues (119 and 120) and a loop region (residues 165–188) do not lie in continuous density, suggesting disorder or flexibility in these regions.

The RNase H-like catalytic core (Yang *et al*, 1990; Rice and Baker, 2001) is also conserved in integrases (Dyda *et al*, 1994; Bujacz *et al*, 1996), transposases (Rice and Mizuuchi, 1995; Davies *et al*, 2000; Hickman *et al*, 2005) and Argonaute, a component of the RNAi pathway (Song *et al*, 2004). The closest structural homolog of the catalytic domain of *Mos1* is the core domain of avian sarcoma virus (PDBID: 1CXQ), which can be superimposed with an r.m.s.d. of 3.3 Å over 127 C α atoms with 15% shared sequence identity. The next three closest homologues are the catalytic core domain of HIV-1 integrase (PDBID:1B9D), the bacteriophage Mu transposase core (PDBID:1BCO) and Hermes transposase (PDBID:2BW3), respectively.

Comparison of the catalytic domains of *Mos1* and Tn5 (the only transposase for which there is a structure with DNA bound in the active site) reveals similar active site architecture (Figure 2C). The r.m.s. fit of the three active site carboxyl groups is 0.52 Å; however, the C α atom of D284 of *Mos1* is shifted approximately 1 Å closer to the active site than the C α atom of E326 in Tn5 due to the slightly different orientation of helix 5 (Figure 2C).

Mos1 transposase can bind two Mn^{2+} ions in the active site

The crystals of the C-terminal domain of *Mos1* transposase were grown in 5 mM $MgCl_2$ and the structure contains one Mg^{2+} in the active site. The metal ion is coordinated by the carboxyl oxygens of D156 and D249 and four water molecules. This is designated Site 1 (Figure 2D). The preference for Mg^{2+} binding at Site 1 in *Mos1* transposase is similar to the binding of a single Mg^{2+} in ASV (Bujacz *et al*, 1997) and

Figure 2 The crystal structures of (A) the catalytic domain of *Mos1* transposase and (B) the catalytic domain of Tn5 transposase. The RNase-H core is marine blue, the N- and C-terminal helices of *Mos1* and the insertion domain of Tn5 are cyan, the active site residues are highlighted in green and the metal ions are purple spheres. The disordered region of *Mos1* is shown as a dotted line. (C) Comparison of the active site residues of *Mos1* (green) and Tn5 transposases (orange; PDB accession code 1MUH); one Mg^{2+} ion (purple, M1) is bound in *Mos1* and one Mn^{2+} ion (pink, M2) is bound in Tn5. The residues of the YREKK motif in Tn5 are also shown, along with the structurally equivalent residues of *Mos1* transposase. (D) The active site of native crystals of *Mos1* transposase contains one Mg^{2+} in the active site in Site 1, with coordinated waters shown as red spheres; the $2F_o - F_c$ electron density map is shown as a grey mesh. (E) The active site of *Mos1* transposase crystals soaked in $MnCl_2$ contains two Mn^{2+} ions (pink spheres) in Sites 1 and 2. The Mn anomalous difference electron density map, contoured at 3σ , is shown as a pink mesh. (F) Putative dimers of the *Mos1* catalytic domain related by the crystallographic two-fold axis, which is parallel to the vertical in this view. The residues involved in three areas of protein–protein contacts are highlighted in orange and shown in stick representation (see also Supplementary Figures S3A and S3B). (G) Electrostatic surface potential of the catalytic domain dimer, viewed at an orientation tilted from the two-fold axis to show one of the two channels formed in the dimer core. All pictures were produced using PYMOL (<http://www.pymol.org>).

HIV-1 integrases (Goldgur *et al*, 1998), but contrasts with the preferential binding of a single Mn^{2+} between the first and third residues of the catalytic triad in the first Tn5 PEC structure (Davies *et al*, 2000 and Figure 2C).

A second structure was determined from crystals soaked in 20 mM $MnCl_2$, which diffracted to 3.4 Å (see Table I). Anomalous data, collected at the peak of the K-edge of Mn ($\lambda = 1.89$ Å), revealed two Mn^{2+} ions in the active site: one Mn^{2+} ion occupies Site 1 and the second Mn^{2+} ion is coordinated by D249 and D284 in Site 2 (Figure 2E). The conformation of the protein remains largely unchanged upon binding of the second metal, apart from rotation of the side-chain atoms of D284. Thus, *Mos1* transposase can bind two Mn^{2+} ions in the active site, consistent with a two-metal mechanism for catalysis, as first described for DNA polymerase I (Beese and Steitz, 1991) and shown for Tn5 transposase (Lovell *et al*, 2002) and RNase H (Nowotny *et al*, 2005). Tn5 binds a second Mn^{2+} only in the presence of DNA substrate and a similar DNA substrate requirement for two-metal (Mg^{2+} or Mn^{2+}) binding by HIV-1 and ASV integrases has been suggested (Goldgur *et al*, 1998). By contrast, *Mos1* transposase can bind two Mn^{2+} ions in the absence of DNA substrate.

The essential role of each of the three catalytic Asp residues (D156, D249 and D284) of *Mos1* transposase, in both first- and second-strand cleavage, has been confirmed in biochemical assays. Mutation of each Asp to Asn results in a loss of both first- and second-strand cleavage activity (AD and DJF, unpublished results). While the third catalytic aspartic acid (D284) is conserved in *mariner* elements, Tc elements contain a glutamic acid at this position. The mutation D284E in *Mos1* abolishes transposase activity (Lohe *et al*, 1997) and the structure presented here suggests that, because the backbone atoms of D284 are positioned closer to the active site than in a DDE enzyme, there is a lack of space to coordinate a metal ion in Site 2 when D284 is replaced by E284.

***Mos1* transposase contains no hairpin-stabilizing structural motifs**

Some transposases utilize a hairpin intermediate during transposition, enabling one active site to carry out two coupled DNA cleavage reactions, via a two-metal mechanism. The crystal structure of the PEC of the bacterial transposase Tn5 revealed the molecular basis for DNA excision *via* a hairpin intermediate on transposon DNA (Davies *et al*, 2000). Two structural features for stabilization of this intermediate were highlighted: a base flipping event, involving stacking interactions between the flipped base and W298 (Ason and Reznikoff, 2002), and the interaction of the YREKK motif with DNA bases close to the active site. These features are conserved in the IS4 bacterial transposases (Rezsohazy *et al*, 1993), and conserved tryptophan residues, implicated in base flipping events, have been identified in the hAT transposases and RAG1 recombinase (Zhou *et al*, 2004).

Previously we have shown, by *in vitro* assays, that there is no hairpin intermediate during *Mos1* transposition (Dawson and Finnegan, 2003) and comparison of the structures of the catalytic domains of *Mos1* and Tn5 transposase supports this conclusion. *Mos1* transposase does not contain the YREKK motif for hairpin stabilization, formed in Tn5 transposase by residues Y319, R322, E326, K330 and K333; the structurally equivalent residues in *Mos1* transposase are H273, A281, D284, A289 and G292 (Figure 2C). *Mos1* transposase also lacks a structural equivalent to W298 in Tn5, and the hydrophobic cleft into which base T2 is flipped in Tn5 is replaced by the N- and C-terminal capping helices in *Mos1*. We have also confirmed using site-directed mutagenesis that W159 (a possible mimic for W298 in Tn5) is not required for transposase activity (data not shown).

The crystal structure provides a putative dimerization interface for the catalytic domain of *Mos1* transposase

Gel filtration studies of full-length (T216A) *Mos1* transposase (at 2 mg/ml) indicate that the protein is dimeric at this

Table I Data and refinement statistics

	Native	Semet peak	SeMet infl	SeMet remote	Mn peak
<i>Data set</i>					
Space group	P4(1)2(1)2	P4(1)2(1)2	P4(1)2(1)2	P4(1)2(1)2	P4(1)2(1)2
Unit cell (<i>a</i> , <i>c</i>) (Å)	44.2, 206.0	44.7, 205.8	44.7, 205.8	44.7, 205.8	44.2, 205.7
Resolution (Å)	20.0–2.25 (2.35–2.25)	30.0–3.0 (3.09–3.0)	30.0–3.0 (3.09–3.0)	30.0–3.0 (3.09–3.0)	20.0–3.40 (3.5–3.4)
Completeness	97.1 (85.4)	99.7 (99.5)	99.7 (99.4)	99.9 (99.4)	96.5 (88.8)
Total reflections	111301	97828	98051	98273	35370
Unique reflections	10257	4766	4797	4782	3123
% Reflections for R_{free}	4.8	—	—	—	—
R_{sym} (%)	13.0 (45.0)	6.9 (10.0)	6.5 (10.0)	6.5 (10.0)	9.7 (20.7)
$I/\sigma(I)$	13.0 (6.3)	29.4 (22.2)	29.35 (21.3)	29.48 (21.6)	9.95 (3.8)
Wavelength (Å)	0.97624	0.97868	0.97939	0.97500	1.8925
f'	—	–7.5	–10.0	–4.3	–7.2
f''	—	5.4	2.7	4.08	6.1
<i>Refinement</i>					
R_{work}/R_{free}	0.193 (0.267)				
Bond length r.m.s.d. from ideality (Å)	0.026				
Bond angle r.m.s.d. from ideality (deg)	2.04				
Ramachandran plot	88.7% core; 11.3% allowed				
Average B factor (Å ²)	34.5				
Protein atoms in the model	1863				
Water molecules	117				
Mg ions in the model	1				

concentration, and that no higher-order multimers are formed (Supplementary Figure 4a). The crystals of the catalytic domain of *Mos1* contain molecular dimers related by the crystallographic two-fold axis, suggesting a possible dimerization interface (Figure 2F). In this packing arrangement, the active sites face towards each other on the inside of the dimer, as in the Tn5 synaptic complex (Davies *et al*, 2000) and in the hexamer model of *Hermes* transposase (Hickman *et al*, 2005; Supplementary Figure 2a and b). By contrast, the dimeric core domains of retroviral integrases (e.g. HIV-1) have the active sites on the outside of the dimer, facing away from each other (Chen *et al*, 2000). Despite the close structural homology of *Mos1* and integrase core domains, the retroviral integrase dimer interface is not conserved in the *Mos1* transposase crystals (see Supplementary Figure 2c) and this may reflect the involvement of dimers in *Mos1* transposition, whereas higher-order oligomers are relevant for integrase activity.

The distance between the Mn^{2+} ions in Site 2 in the two active sites of the *Mos1* dimer is 21 Å. Interestingly, this is consistent with the 2-bp spacing between the TA/AT sites (on opposite strands of target DNA) into which the excised *Mos1* element is inserted. This is closer than the separation of the two active sites in Tn5 (~41 Å) and in the *Hermes* hexamer model (~40 Å) which catalyse target-site insertions spaced 9 and 8 bp apart, respectively.

There are three areas of protein–protein contacts across the interface. Helices $\alpha 6$ of each monomer interact, primarily via residues H293, which form intermolecular hydrogen bond interactions as well as hydrophobic interactions with A296 (Supplementary Figure S3a). The structurally equivalent helix in *Hermes* transposase (residues 573–584) is involved in protein–protein interactions across interface 3 in the hexamer model. There is also a salt bridge between K190 and E345 (Supplementary Figure 3b) and a symmetry-related equivalent. The mutation K190A disfavours the formation of full-length protein dimers in gel filtration studies, and is defective in PEC formation in biochemical assays (Supplementary Figure 4). Thus, this interaction, seen in the putative dimerization interface of catalytic domains, is also likely to have been part of the original assembly of full-length dimers. The N-terminal six residues of the crystal structure, WVPHEL (residues 119–124), are conserved within the *mariner* family and may play a role in the interaction between transposase monomers as they interact across the dimerization interface found in the crystal (Supplementary Figure 3c). Mutations V120A and L124S resulted in decreased protein–protein interactions in a yeast two-hybrid assay (Zhang *et al*, 2001). In biochemical assays, the mutant proteins W119P and L124S are defective in PEC formation and second strand cleavage (A Dawson and N O'Hagan, unpublished data). Thus, this region (which links the N and C domains) is likely to mediate the formation of critical protein–protein interactions within the PEC. Although the buried surface area of the catalytic domain dimer interface is small (630 Å²), the dimer interface of the full-length protein will be stabilized further by interactions involving the N-terminal domain (Zhang *et al*, 2001).

Two channels, lined with positively charged surfaces, are formed in the core of the dimer of catalytic domains. These channels are approximately 9 Å in diameter and an active site is positioned near the top of each channel (Figure 2G). Three similarly charged, but broader, channels were observed in the

hexamer model of *Hermes* transposase and it was speculated that DNA might bind within these channels (Hickman *et al*, 2005).

The structure of Tc3 bipartite DNA-binding domain provides a model for the *Mos1* transposase DNA-binding domain

The first 112 amino acids of *Mos1* transposase are predicted to contain two helix–turn–helix motifs, for DNA recognition: one, between residues 88 and 108, is required for sequence-specific binding (Petrokovski and Henikoff, 1997; Zhang *et al*, 2001; Bigot *et al*, 2005), and another between residues 25 and 54 (Figure 3A; Plasterk *et al*, 1999). R106 is critical for sequence-specific recognition, as the mutation R106A abolishes binding (Zhang *et al*, 2001). The first 30 amino acids are also essential for DNA binding (Zhang *et al*, 2001). The two predicted HTH motifs are joined by a sequence (GKPPK; residues 66–70) likely to form a variant of the AT-hook motif, which makes contacts with the DNA minor groove (Aravind and Landsman, 1998). The consensus sequence for mariner-like elements contains the AT-hook signature tripeptide GRP (Robertson, 1995), but the arginine is replaced by lysine (K67) in *Mos1* transposase.

AT-hook motifs have also been identified in two other Tc1/*mariner* transposases: a true AT hook in the Tc1 homolog Ant1, with the sequence RGRPR (residues 62–66) (Aravind and Landsman, 1998) and a variant (with the sequence GRRR: residues 59–62) in Sleeping Beauty, between the PAI and RED subdomains (Izsvak *et al*, 2002). Thus, the AT-hook-like linker may be a recurring feature in the Tc1/*mariner* family of transposases, but with considerable degeneracy in the motif sequence.

The DNA-binding domain of Tc3 transposase is also composed of two HTH motifs, although it is the first HTH motif that is critical for sequence-specific recognition (Thompson and Woodbury, 2001), not the second as for *Mos1*. The two HTH motifs are joined by a sequence RAPR (residues 54–57) and the two arginines are inserted into the DNA minor groove in a manner reminiscent of an AT hook. An alignment (Figure 3A and B) of the equivalent domains of *Mos1* and Tc3 indicates that the structure of the bipartite DNA-binding domain of Tc3 in complex with DNA (Watkins *et al*, 2004) is a good template for building a model of the N-terminal DNA-binding domain of *Mos1* transposase. The structure of DNA in this complex is distorted from B-DNA, predominately by local narrowing of the minor groove in two TA rich regions at the centre of the IR. The naked *Mos1* IR DNA is also predicted to be bent, by approximately 20°, with the bending propensity dependent upon the base at position 16, the centre of an AT-rich region (Bigot *et al*, 2005). This bend may be accentuated upon binding of protein. We have used the bent Tc3 DNA structure as the template for the *Mos1* IR model, as it is the best available experimentally determined IR structure.

Figure 4A shows the proposed structure of the DNA-binding domain of *Mos1* transposase bound to IR DNA. This model shows that the overall dimensions of the *Mos1* DNA-binding domain/IR complex are consistent with footprinting data, which suggest that most of the IR bases are protected (Auge-Gouillou *et al*, 2001; Lampe *et al*, 2001). The model is also consistent with the results of binding studies using truncated IR sequences (Auge-Gouillou *et al*, 2005).

Figure 3 (A) Alignment of the sequences of the DNA-binding domains of *Mos1* and *Tc3* transposases. The secondary structure elements highlighted above the alignment are those predicted for *Mos1* transposase and those below are from the crystal structure of *Tc3* (PDB ID: 1U78). Orange triangles mark residues involved in the putative dimerization interface of *Mos1*. (B) Alignment of the sequences of the catalytic domains of *Mos1*, *Tc3* and Sleeping Beauty transposases. Purple stars mark the catalytic residues, and the residues structurally equivalent to the YREKK motif in *Tn5* are highlighted by green circles. Residues marked with an orange triangle are involved in the putative dimerization interface. Secondary structure elements correspond to the *Mos1* crystal structure presented here. (C) Sequence of the right IR of *Mos1* transposon and the outer left IR of *Tc3* transposon, which is numbered according to the *Tc3*/DNA cocrystal structure (Watkins *et al*, 2004). The *Mos1* right IR is numbered from 1 to 28 on the lower (nontransferred) strand and from 29 to 56 on the upper (transferred) strand. The bases which differ between the left and right IRs of *Mos1* (1, 16, 18 and 26) are highlighted in bold italics. The flanking bases are displayed in lower case italics and the cleavage sites are marked by dotted lines.

Contacts made by the two HTH motifs and the major groove of the DNA position the transposon end correctly for cleavage by the catalytic domain (Figure 4A). The first HTH binds between bases 20 and 25 of the IR, and the second between bases 3 and 8. The AT-hook-like linker makes contacts with the minor groove of DNA between bases 15 and 18, an AT-rich region. Pyrimidine-specific contacts are made between K67:NZ and T39:O2 (paired with A18) and between K70:NZ and T42:O2 (paired with A15). These are equivalent to those observed in the *Tc3A* structure between R54:Nε and T42:O2 and between R57:NH1 and T39:O2. However, there is no equivalent to the purine-specific contact

seen in *Tc3* between R54:NH2 and A43:N3. The DNA bases contacted by the second HTH and the linker correspond to regions highly conserved in mariner-like element IRs and are therefore likely to be critical for sequence-specific DNA binding (Lampe *et al*, 2001).

Dimerization of the DNA-binding domain of *Mos1* transposase via the second HTH motif

In yeast-two-hybrid experiments we have shown that the mutations Q91R, L92H or A93V give rise to decreased interactions between mutant and wild-type *Mos1* transposase (Zhang *et al*, 2001). These results are consistent with dimer-

Figure 4 (A) Model of the DNA-binding domain of *Mos1* transposase bound to right-hand IR DNA. The nontransferred DNA strand (bases G4–T28) is coloured pale purple and the TS (bases A29–A56) is coloured hot pink. The numbering of bases is that shown in Figure 3C. (B) Closeup of the putative dimerization interface of the *Mos1* DNA-binding domain, which is proposed to dimerize via the second HTH motif. One chain is coloured gold and the other pink. Hydrophobic interactions are indicated by green lines, whereas electrostatic interactions are highlighted by blue lines.

ization of *Mos1* DNA-binding domains via the second HTH motif (residues 72–110), the domain critical for sequence-specific recognition, as shown in Figure 4B. In this model, based on the dimer found in the Tc3 crystal structure (Watkins *et al*, 2004), there are symmetry-related sets of electrostatic and hydrophobic interactions. The N ϵ of Q78 and the carbonyl oxygen of L77 form a hydrogen bond across the dimerization interface, which is equivalent to that observed in both the Tc3A (van Pouderooyen *et al*, 1997) and Tc3 structures between the N ϵ of Q14 and the carbonyl oxygen of A13. There are further electrostatic interactions between the side-chain carboxyl oxygens of D85 and the N ϵ atoms of Q91 and Q87. In addition, the methyl groups of L77, L81 and L96 are involved in hydrophobic interactions across the interface.

A model of the PEC

To model the PEC, formed by a dimer of transposase and two transposon ends (Auge-Gouillou *et al*, 2005), we manually docked the model of the dimeric N-terminal DNA-binding domain of *Mos1* transposase onto the dimeric structure of the catalytic domain (Figure 5A and B). These two domains are mutually compatible and come together neatly; only some

small adjustments of bases 55 and 56 of the TS are required to avoid steric clashes with the protein side chains. The precise position of the DNA in the catalytic domain will depend on the extent to which *Mos1* IR DNA is bent, and clarification of this will require the X-ray structure of the PEC. The short gap between the modelled N- and C-terminal domains is spanned by a linker of nine amino acids (I113–P121), and there are two possible routes for this linker, as shown in Figure 1. In one case it joins the catalytic domain and the N-terminal domain bound to the same IR end that is in the active site, giving rise to *cis* cleavage. In the second case, the linker connects the catalytic and N-terminal domains to promote *trans* cleavages.

The two transposon ends approach the active sites in parallel in contrast to the antiparallel arrangement of the IRs in the Tn5 synaptic complex structure (Figure 5C). The DNA mimics the double-strand break product. The TS DNA is positioned close to the active site in this model, whereas the nontransferred strand DNA lies near the N-terminal end of helix 4, close to an outer face of the molecule. Double-stranded flanking DNA cannot be easily accommodated in this model of the PEC. However, the two channels lined with positively charged surfaces, which form in the centre of the dimer of catalytic domains (Figure 2G), suggest a possible

Figure 5 (A) Stereo view of the model of the *Mos1* PEC formed by a dimer of transposase and two transposon ends and (B) alternative stereo view of the *Mos1* PEC model. The two-fold crystallographic axis is parallel to the vertical in both views. (C) Stereo view of the structure of the Tn5 PEC. The colour scheme for the catalytic domains is identical to that in Figure 2A and B; the DNA-binding domains are coloured wheat; the nontransferred strand of the IR DNA is pale purple and the TS is hot pink. The C-terminal dimerization domain of Tn5 is in grey.

route for single-stranded flanking DNA of the TS of each IR (not modelled) to pass through the core of the PEC. The DNA flanking the nontransferred strand could drape over the outer face of the complex, or may be released by transposase prior to PEC formation.

Flanking duplex DNA is a pre-requisite for first-strand cleavage, but inhibits second-strand cleavage

To assess flanking DNA requirements for first- and second-strand cleavage, DNA cleavage assays were performed using substrates with and without flanking DNA on the NTS or TS

strand (Figure 6). The 33-nucleotide product of first-strand cleavage of the NTS is only detected when duplex flanking DNA is present (Figure 6, lane 4). No cleavage takes place if the flanking DNA is single stranded (Figure 6, lane 1), indicating that transposase requires duplex DNA downstream of the cleavage site to carry out first-strand cleavage. In contrast, second-strand cleavage (of the TS) is enhanced, relative to intact or prenicked substrate, after the removal of flanking DNA of the NTS (Figure 6, lanes 7, 10 and 13). As second-strand cleavage via hairpin formation would require that the 3'OH on the flanking DNA was available to act as the nucleophile for attack of the TS, these results further confirm the absence of a hairpin intermediate in *Mos1* transposition. These observations also support the suggestion that flanking DNA of the TS is single-stranded within the PEC and are therefore consistent with our model of the PEC, which could accommodate the flanking DNA strand through a pore in the structure.

Figure 6 The effect of flanking DNA on first- and second-strand cleavage. DNA denaturing gel (10%) of the products of DNA cleavage assays, carried out under standard conditions (A), without DMSO (B) and without transposase (C). Five variants of the IRR100 substrate were used in the assays, as described in the text and shown schematically above the corresponding lanes on the gel; the position of the radiolabel on each substrate is shown by an asterisk and the transferred and nontransferred strands are marked TS and NTS, respectively. Lane numbers are indicated below the gel. Arrows indicate the positions of the labelled products of first-strand cleavage (33-mer) and second-strand cleavage (70-mer), which have been calibrated with DNA markers of known size. The additional cleavage bands (between the 70- and 100-mer) are assumed to be secondary cleavage sites. The extra bands observed in lanes 10 and 11, migrating more slowly than free substrate, correspond to strand transfer complexes, as described previously (Dawson and Finnegan, 2003).

A two-metal mechanism for hydrolysis, accompanied by a conformational change, can account for the two sequential cleavage events in *Mos1* transposition

The active site of *Mos1* transposase (comprising the DDD motif) can bind two divalent metal ions. A two-metal mechanism for hydrolysis (Beese and Steitz, 1991) can therefore be invoked to account for the sequential cleavage of two DNA strands of opposite polarity, with the roles of the two metals in the active site being reversed for the second hydrolysis reaction, compared with the first. During first-strand cleavage, within a SEC, one metal activates a water molecule for nucleophilic attack on the NTS of DNA, which is stabilized in the active site by the second metal. Hydrolysis of the NTS DNA generates a 5' phosphate on the transposon DNA and a 3'OH on flanking DNA, which is not required for or involved in further reactions in *Mos1* transposition (Dawson and Finnegan, 2003).

A conformational change is required after first-strand cleavage, to release the NTS, thereby vacating the active site, and to position a TS in the active site for second-strand cleavage. We postulate that synaptic complex formation (Figure 1) promotes the repositioning of the IR DNA in the active site. This could be driven by protein dimerization and aided by local melting of DNA flanking the NTS following first-strand cleavage. The DNA-binding domain and bound DNA would rotate with respect to the catalytic domain, with the linker region between the DNA-binding domain and catalytic domain acting as a hinge. The extent of the required rotation is not known at this stage, particularly as second-strand cleavage could occur in *cis* or in *trans*. This conformational change would be similar to that used by the homing endonuclease I-TevI, which acts as a hinged monomer to perform double-strand DNA breaks with only one active site (Mueller *et al*, 1995; Van Roey *et al*, 2002).

Second-strand cleavage proceeds by hydrolysis of the O3'-P bond between residue A56 and T57, generating a 3'OH on the TS transposon DNA. This can remain bound to one of the metals in the active site, where it is poised to act as the nucleophile for the strand transfer step. Following the binding of target DNA to the *Mos1* PEC, the strand-transfer reaction can proceed, by reversal of the roles of the two metals in the active site for a second time. The two active sites present in the dimeric PEC can therefore carry out all three reaction steps of *Mos1* transposition (Figure 1); four active sites are not essential in this regard. However, in the absence of X-ray structures of PEC intermediates, a higher-order assembly cannot be excluded and it has been suggested that a dimer of transposase is required to bind to each *Mos1* terminal IR, leading to the formation of a tetrameric PEC (Auge-Gouillou *et al*, 2005).

The crystal structure of the catalytic domain of *Mos1* transposase and the model for the PEC presented here provide insight into the mechanism of *Mos1* transposition. *Mos1* transposase differs from other transposases characterized so far in that one active site carries out the sequential cleavage of two DNA strands with opposite polarity by two successive hydrolysis reactions, rather than a DNA hairpin intermediate. Comparison of the sequences of the catalytic domains of *Mos1* with other Tc1/*mariner* transposases reveals the absence of structural motifs for hairpin stabilization (Figure 3B; Robertson, 1995) in all members of this family. It therefore seems likely that all members of the Tc1/*mariner*

family, including Sleeping Beauty, use a mechanism for DNA excision similar to that proposed here for *Mos1*.

Materials and methods

Structure determination

Mos1 transposase was purified and crystals were grown as described previously (Richardson *et al*, 2004). Analysis of the protein crystals by electrospray mass spectrometry gives a major peak with molecular mass 27.3 kDa. N-terminal amino-acid sequencing confirmed that the full-length 345-amino-acid protein had undergone proteolysis during crystallization, leaving a 226-amino-acid C-terminal domain (Figure 3B).

Diffraction data were collected at the ESRF (station ID29) and Daresbury SRS (Station 14.2). The crystal data are given in Table I. Data were indexed and scaled using DENZO and SCALEPACK (Otwinowski and Minor, 1997). Multiple-wavelength anomalous dispersion (MAD) X-ray diffraction data of crystals grown from seleno-methionine-substituted protein were used to find the positions of three selenium atoms with SOLVE (Terwilliger and Berendzen, 1999) (Table I) to give a figure of merit of 50.5%. RESOLVE (Terwilliger, 2000) was used to improve phases by density modification, giving a final figure of merit of 75.5%. Subsequent manual model building using the program WITNOTP (Widmer, 1999) and refinement with REFMAC within the CCP4 suite (Bailey, 1994) confirmed that only the C-terminal catalytic domain of the protein was present. A representative section of the experimental electron density map produced by RESOLVE is shown in Supplementary Figure S.1A and the $2F_o - F_c$ electron density map after refinement is shown in Supplementary Figure S.1B. The closest structural homologues of the catalytic domain of *Mos1* transposase were identified by the DALI server (Holm and Sander, 1996).

Sequence alignment and model building of the N-terminal DNA-binding domain

Sequence alignment of the N-terminal DNA-binding domains of *Mos1* and Tc3 transposases was carried out using ClustalW, in combination with manual alignment guided by the secondary structure suggested by PredictProtein (Rost and Liu, 2003). The two HTH motifs in *Mos1* transposase were predicted using the program HelixTurnHelix within the EMBOSS software suite (Rice *et al*, 2000). The first 54 residues of *Mos1* were aligned with the second HTH domain (residues 62–104) of Tc3 and residues 72–112 of *Mos1* transposase were aligned with the first HTH domain of Tc3 (residues 1–48), as shown in Figure 3A. The palindromic, AT hook-like sequence KPPK (residues 67–70) has been aligned with residues 54–57 (RAPR) of Tc3, and during the model building the chain direction of this region was reversed so that the two HTH domains could be linked. Residues 72–76 form a linker between the AT-hook-like region and the second HTH domain. Residues 53–66 are not predicted to form any secondary structure, and have been excluded from the model.

References

- Aravind L, Landsman D (1998) AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res* **26**: 4413–4421
- Ason B, Reznikoff WS (2002) Mutational analysis of the base flipping event found in Tn5 transposition. *J Biol Chem* **277**: 11284–11291
- Auge-Gouillou C, Brillet B, Hamelin MH, Bigot Y (2005) Assembly of the mariner *Mos1* synaptic complex. *Mol Cell Biol* **25**: 2861–2870
- Auge-Gouillou C, Hamelin MH, Demattei MV, Periquet G, Bigot Y (2001) The ITR binding domain of the Mariner *Mos-1* transposase. *Mol Genet Genomics* **265**: 58–65
- Bailey S (1994) The ccp4 Suite—Programs for protein crystallography. *Acta Crystallogr D* **50**: 760–763
- Beese LS, Steitz TA (1991) Structural basis for the 3′–5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J* **10**: 25–33
- Bhasin A, Goryshin IY, Reznikoff WS (1999) Hairpin formation in Tn5 transposition. *J Biol Chem* **274**: 37021–37029

The sequence of the N-terminal domain of *Mos1* transposase and the right-hand IR DNA sequence were threaded manually onto the three-dimensional structure of the Tc3 bipartite DNA-binding domain DNA complex (Figure 4A). The TS was extended by two residues to produce a three base overhang, mimicking the staggered double-strand break product formed in the PEC.

DNA cleavage assays

DNA cleavage assays were performed using wild-type *Mos1* transposase (purified as described previously (Zhang *et al*, 2001)) and five synthetic oligonucleotide substrates based on the IRR100 substrate (Dawson and Finnegan, 2003). IRR100 comprises a 70-bp sequence (containing 28 bp of the right-hand IR of *Mos1*) plus 30 bp of flanking DNA. It is cleaved sequentially by *Mos1* transposase *in vitro*, 3 bases from the end of the IR on the NTS and precisely at the end of the IR on the TS (Supplementary Figure 5); cleavage of the TS only occurs following cleavage of the NTS (Dawson and Finnegan, 2003). If IRR100 is radio-labelled on the NTS at the 5′ end, first-strand cleavage generates a labelled 33-mer, which can be detected following separation under denaturing conditions. Conversely, if IRR100 is 5′ radiolabelled on the TS, second-strand cleavage generates a 70-mer. The DNA substrates were prepared as described previously (Dawson and Finnegan, 2003).

To assess flanking DNA requirements for first-strand cleavage, two substrates, labelled on the NTS, were used: wild-type IRR100 and IRR100 lacking the 30 bp flanking DNA on the TS. To assess flanking DNA requirements for second-strand cleavage, three substrates, labelled on the TS, were used: wild-type IRR100, IRR100 pre-nicked on the NTS at the site of first-strand cleavage, and IRR100 lacking the 33 bases of flanking DNA on the NTS, as shown in Figure 6.

In each assay 150 fmol of substrate was incubated with 7 nM protein for 2 h at 302 K. As already reported, *in vitro* cleavage of linear DNA substrates by *Mos1* transposase requires DMSO, although this requirement is more stringent for first-strand cleavage than second. Thus, each experiment was carried out with and without DMSO (20%), as well as without transposase, as a control.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online.

Acknowledgements

We thank the staff of the ESRF and Daresbury SRS, as well as Lindsay Tulloch and Liam Worrall for help with data collection. This work was supported by a Caledonian Research Foundation fellowship to JMR, project grants from the Wellcome Trust to MDW and DJF (074522) and from the Medical Research Council (G0301082). AD and JMR received funds from a Wellcome Trust VIP award to the University of Edinburgh. The structure has been submitted to the Protein Data Bank with accession code 2F7T.

- Bigot Y, Brillet B, Auge-Gouillou C (2005) Conservation of palindromic and mirror motifs within inverted terminal repeats of mariner-like elements. *J Mol Biol* **351**: 108–116
- Bujacz G, Alexandratos J, Wlodawer A, Merkel G, Andrake M, Katz RA, Skalka AM (1997) Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity. *J Biol Chem* **272**: 18161–18168
- Bujacz G, Jaskolski M, Alexandratos J, Wlodawer A, Merkel G, Katz RA, Skalka AM (1996) The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations. *Structure* **4**: 89–96
- Chen JC-H, Krucinski J, Miercke LJW, Finer-Moore JS, Tang AH, Leavitt AD, Stroud RM (2000) Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc Natl Acad Sci USA* **97**: 8233–8238
- Collier LS, Carlson CM, Ravimohan S, Dupuy AJ, Largaespada DA (2005) Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature* **436**: 272–276

- Davies DR, Goryshin IY, Reznikoff WS, Rayment I (2000) Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **289**: 77–85
- Dawson A, Finnegan DJ (2003) Excision of the *Drosophila* mariner transposon Mos1. Comparison with bacterial transposition and V(D)J recombination. *Mol Cell* **11**: 225–235
- Dupuy AJ, Akagi K, Largaespada DA, Copeland NG, Jenkins NA (2005) Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature* **436**: 221–226
- Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266**: 1981–1986
- Goldgur Y, Dyda F, Hickman AB, Jenkins TM, Craigie R, Davies DR (1998) Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc Natl Acad Sci USA* **95**: 9150–9154
- Hartl DL (2001) Discovery of the transposable element Mariner. *Genetics* **157**: 471–476
- Hickman AB, Li Y, Mathew SV, May EW, Craig NL, Dyda F (2000) Unexpected structural diversity in DNA recombination: the restriction endonuclease connection. *Mol Cell* **5**: 1025–1034
- Hickman AB, Perez ZN, Zhou L, Musingarimi P, Ghirlando R, Hinshaw JE, Craig NL, Dyda F (2005) Molecular architecture of a eukaryotic DNA transposase. *Nat Struct Mol Biol* **12**: 715–721
- Holm L, Sander C (1996) Mapping the protein universe. *Science* **273**: 595–602
- Ivics Z, Kaufman CD, Zayed H, Miskey C, Walisko O, Izsvak Z (2004) The Sleeping Beauty transposable element: evolution, regulation and genetic applications. *Curr Issues Mol Biol* **6**: 43–55
- Izsvak Z, Khare D, Behlke J, Heinemann U, Plasterk RH, Ivics Z (2002) Involvement of a bifunctional, paired-like DNA-binding domain and transpositional enhancer in Sleeping Beauty transposition. *J Biol Chem* **277**: 34581–34588
- Kennedy AK, Guhathakurta A, Kleckner N, Haniford DB (1998) Tn10 transposition via a DNA hairpin intermediate. *Cell* **95**: 125–134
- Lampe DJ, Churchill MEA, Robertson HM (1996) A purified mariner transposase is sufficient to mediate transposition *in vitro*. *EMBO J* **15**: 5470–5479
- Lampe DJ, Walden KK, Robertson HM (2001) Loss of transposase-DNA interaction may underlie the divergence of mariner family transposable elements and the ability of more than one mariner to occupy the same genome. *Mol Biol Evol* **18**: 954–961
- Lohe AR, De Aguiar D, Hartl DL (1997) Mutations in the mariner transposase: the D, D(35)E consensus sequence is non-functional. *Proc Natl Acad Sci USA* **94**: 1293–1297
- Lovell S, Goryshin IY, Reznikoff WR, Rayment I (2002) Two-metal active site binding of a Tn5 transposase synaptic complex. *Nat Struct Biol* **9**: 278–281
- Mueller JE, Smith D, Bryk M, Belfort M (1995) Intron-encoded endonuclease I-TevI binds as a monomer to effect sequential cleavage via conformational changes in the td homing site. *EMBO J* **14**: 5724–5735
- Nowotny M, Gaidamakov SA, Crouch RJ, Yang W (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **121**: 1005–1016
- Otwinowski Z, Minor W (1997) Processing X-ray diffraction data collected in oscillation mode. *Macromol Crystallogr, Pt A* **276**: 307–326
- Petrokovski S, Henikoff S (1997) A helix–turn–helix DNA-binding motif predicted for transposases of DNA transposons. *Mol Gen Genet* **254**: 689–695
- Plasterk RH, Izsvak Z, Ivics Z (1999) Resident aliens: the Tc1/mariner superfamily of transposable elements. *Trends Genet* **15**: 326–332
- Rezsohazy R, Hallet B, Delcour J, Mahillon J (1993) The IS4 family of insertion sequences: evidence for a conserved transposase motif. *Mol Microbiol* **9**: 1283–1295
- Rice P, Longden I, Bleasby A (2000) EMBOS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**: 276–277
- Rice P, Mizuuchi K (1995) Structure of the Bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell* **82**: 209–220
- Rice PA, Baker TA (2001) Comparative architecture of transposase and integrase complexes. *Nat Struct Biol* **8**: 302–307
- Richardson JM, Zhang L, Marcos S, Finnegan DJ, Harding MM, Taylor P, Walkinshaw MD (2004) Expression, purification and preliminary crystallographic studies of a single-point mutant of Mos1 mariner transposase. *Acta Crystallogr D* **60**: 962–964
- Robertson HM (1995) The Tc1-mariner superfamily of transposons in animals. *J Insect Physiol* **41**: 99–105
- Robinson AS, Franz G, Atkinson PW (2004) Insect transgenesis and its potential role in agriculture and human health. *Insect Biochem Mol Biol* **34**: 113–120
- Rost B, Liu J (2003) The PredictProtein server. *Nucleic Acids Res* **31**: 3300–3304
- Roth DB, Menetski JP, Nakajima PB, Bosma MJ, Gellert M (1992) V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in scid mouse thymocytes. *Cell* **70**: 983–991
- Song JJ, Smith SK, Hannon GJ, Joshua-Tor L (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**: 1434–1437
- Terwilliger TC (2000) Maximum likelihood density modification. *Acta Crystallogr D* **56**: 965–972
- Terwilliger TC, Berendzen J (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D* **55**: 849–861
- Thompson M, Woodbury NW (2001) Thermodynamics of specific and nonspecific DNA binding by two DNA-binding domains conjugated to fluorescent probes. *Biophys J* **81**: 1793–1804
- van Pouderooyen G, Ketting RF, Perrakis A, Plasterk RHA, Sixma TK (1997) Crystal structure of the specific DNA-binding domain of Tc3 transposase of *C. elegans* in complex with transposon DNA. *EMBO J* **16**: 6044–6054
- Van Roey P, Meehan L, Kowalski JC, Belfort M, Derbyshire V (2002) Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. *Nat Struct Biol* **9**: 806–811
- Watkins S, van Pouderooyen G, Sixma TK (2004) Structural analysis of the bipartite DNA-binding domain of Tc3 transposase bound to transposon DNA. *Nucleic Acids Res* **32**: 4306–4312
- Widmer A (1999) WITNOTP. Basel, Switzerland: Novartis, AG
- Williams DC, Boulin T, Ruaud AF, Jorgensen EM, Bessereau JL (2005) Characterization of Mos1-mediated mutagenesis in *Caenorhabditis elegans*: a method for the rapid identification of mutated genes. *Genetics* **169**: 1779–1785
- Yang W, Hendrickson WA, Crouch RJ, Satow Y (1990) Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. *Science* **249**: 1398–1405
- Yant SR, Meuse L, Chiu W, Ivics Z, Izsvak Z, Kay MA (2000) Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. *Nat Genet* **25**: 35–41
- Zhang L, Dawson A, Finnegan DJ (2001) DNA binding activity and subunit interaction of the mariner transposase. *Nucleic Acids Res* **29**: 3566–3575
- Zhou L, Mitra R, Atkinson PW, Hickman AB, Dyda F, Craig NL (2004) Transposition of hAT elements links transposable elements and V(D)J recombination. *Nature* **432**: 995–1001