

review

Comparative architecture of transposase and integrase complexes

Phoebe A. Rice¹ and Tania A. Baker²

Transposases and retroviral integrases promote the movement of DNA segments to new locations within and between genomes. These recombinases function as multimeric protein–DNA complexes. Recent success in solving the crystal structure of a Tn5 transposase–DNA complex provides the first detailed structural information about a member of the transposase/integrase superfamily in its active, DNA-bound state. Here, we summarize the reactions catalyzed by transposases and integrases and review the Tn5 transposase–DNA co-crystal structure. The insights gained from the Tn5 structure and other available structures are considered together with biochemical and genetic data to discuss features that are likely to prove common to the catalytic complexes used by members of this important protein family.

Genes are supposed to stay in one place. However, cells carry proteins that are powerful antagonists of this founding principle of molecular biology. These are transposases that catalyze the movement of DNA segments, and the associated genes, to new DNA sites. These proteins have dramatic biological and evolutionary consequences that shape the genomes of organisms. Remarkably, we have just learned that a full 45% of the DNA in the human genome is transposable element-derived sequence¹. Although there are several classes of transposition proteins, the best understood are members of the transposase/retroviral integrase superfamily. These proteins are responsible for the movement of most DNA elements and the virus-like (LTR) retrotransposons, as well as for retroviral integration. The new millennium has brought the first structure of a transposase–DNA complex, as well as several new retroviral integrase structures and the structure of a transposase-associated nuclease. These data, along with recent biochemical experiments, reveal sophisticated new views on the mechanisms used by transposases to assemble multimeric protein–DNA complexes and catalyze recombination. Here, we present an overview of the chemical reactions catalyzed by transposases and integrases, and review the Tn5 transposase–DNA co-crystal structure. After this summary, we use information from the Tn5 structure and other available structures, along with biochemical and genetic data, to discuss aspects of the active protein–DNA complexes that are likely to emerge as shared features among members of this protein family.

Introduction to transposases and transposition

Transposases and retroviral integrases (hereafter transposases) share structurally related catalytic domains (Fig. 1). They are members of a larger superfamily of polynucleotidyl transferases that includes RNaseH and RuvC. The latter two enzymes are both nucleases: RNaseH cleaves the RNA strand of hybrid RNA–DNA duplexes and RuvC cleaves the Holliday junction DNA intermediates formed during homologous recombination^{2,3}. Transposases promote three central steps in recombination⁴ (Fig. 2). These proteins (i) recognize specific DNA sequences at the ends of the mobile element and pair them to form a synaptic complex; (ii)

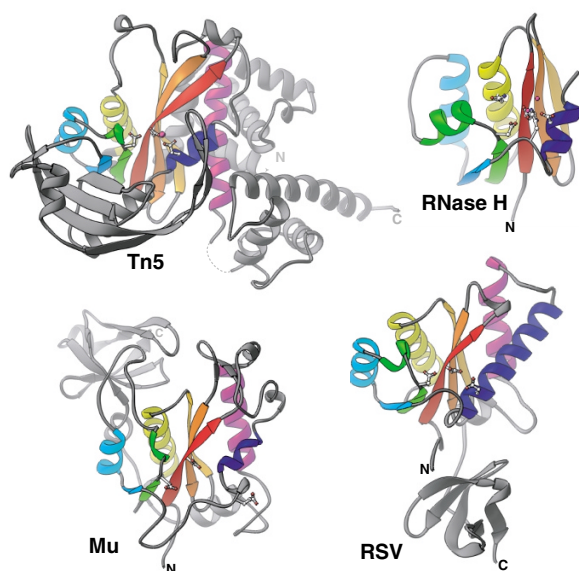


Fig. 1 A conserved catalytic core. Single subunits or fragments of four family members are shown with conserved secondary structure elements colored in rainbow order from red to violet. Shown are: one subunit of intact Tn5 transposase taken from the DNA complex structure¹⁰ (PDB accession code 1F3I); the RNase H domain of HIV reverse transcriptase (RT), taken from an RT–DNA complex¹³ (PDB accession code 1RTD); the core fragment of Mu transposase³⁷ (PDB accession code 1BCM), which contains a small β -barrel in addition to the canonical catalytic domain and the catalytic and C-terminal domains of RSV integrase³² (PDB accession code 1COM). Catalytic residues are shown in ball-and-stick representation with divalent metal ions in pink. In the transposase and integrase structures, the catalytic residues extend from the 1st and 4th β -strands (red and green, respectively) and from the 4th helix (blue). A slightly different array is found in RNase H; in addition to conserved negatively charged residues extending from these 3 structural elements, its active site contains a Glu (here mutated to Gln) extending from the 1st helix (yellow) and a His from the turn just before the 4th helix. Figs 1, 3, 5 and 6 were prepared using RIBBONS⁴⁵.

¹Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, Illinois 60637, USA. ²Department of Biology and the Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA.

Correspondence should be addressed to P.A.R. email: price@midway.uchicago.edu

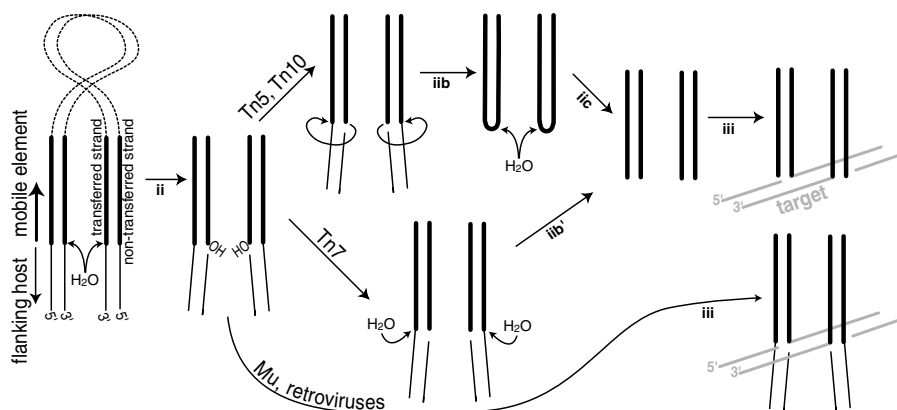


Fig. 2 DNA cleavage and joining reactions catalyzed by transposases. The bold lines represent the ends of a mobile element, and the thin lines are flanking host DNA. The first step described in the text, specific recognition and pairing of the ends by a transposase, is not shown. In step (ii), the transposase nicks the transferred DNA strand, freeing its 3' OH, which in step (iii) attacks a target DNA (gray). Some transposases — for example, Mu transposase — do not cleave the nontransferred strand and, thus, generate a branched DNA product. The equivalent of the flanking host DNA in the retroviral integrase case is only a few base pairs. These sequences on the nontransferred strand are thought to be removed by host enzymes. Some transposases such as Tn10 and Tn5 cleave the nontransferred strand via formation and opening of a 'DNA hairpin' intermediate, shown in steps (iib) and (iic). Alternatively, the nontransferred strand can be directly hydrolyzed, as is done by the Tn7 proteins (step iib'). See refs 46–49 for more detailed reviews of the systems described here.

nick the DNA at each end of the element to liberate the 3'OH ends of the element from the DNA in which it was imbedded; and (iii) promote the direct attack of these cleaved DNA ends on two phosphodiester bonds at a new DNA site, thereby joining the element's genome to a new segment of DNA. Water serves as the attacking nucleophile in the nicking step, whereas the element's 3' hydroxyls do so in the DNA joining step (called DNA strand transfer). Remarkably, both reactions are catalyzed by the same divalent-cation-bound active site. The structural relationship between transposases and some nucleases implies that they may have originated from ancestral nucleases that gained the ability to use a DNA 3'OH as well as water as a nucleophile.

Transposases require no external energy source and do not utilize covalent protein–DNA intermediates. They form very stable complexes with their substrate DNA, such that once properly assembled, they remain bound throughout recombination. The final step, DNA strand transfer, does not change the net number of high-energy bonds and is probably driven forward only by product binding energy. Thus, these proteins usually function stoichiometrically rather than promoting multiple reaction cycles.

The 3' end cleavage and DNA strand transfer reactions are common steps promoted by all characterized members of this transposase family. In addition, many elements (although not all, see Mu in Fig. 2) excise completely from their original DNA location prior to inserting at a new site. This excision demands that both strands of the DNA at the junction between the element and its original host site are severed. Whereas the 'transferred strand' is always cleaved by the transposase, transposons have evolved multiple strategies for cleaving the 'nontransferred strand' (Fig. 2, steps iib, iib' and iic). One strategy, initially discovered for Tn10 transposition and also used by Tn5, involves an interstrand DNA strand transfer reaction that generates a 'DNA hairpin' intermediate⁵ (Fig. 2, step iib). This process proceeds as follows: once the first strand of DNA is cleaved, the newly generated 3'OH end attacks the opposite DNA strand, simultaneously cleaving this strand and generating an interstrand DNA phosphodiester, or 'hairpin'. The same active site then hydrolyzes the hairpin to regenerate a free 3'OH at the element end, now on a DNA broken on both strands^{5,6}. A similar hairpin mechanism is used to intro-

duce a double strand break in the DNA during the assembly of immunoglobulin and T-cell receptor genes by VDJ recombination^{7,8}. A second method of cleaving the nontransferred strand involves simple hydrolysis of this DNA site by a nuclease subunit, other than the transposase, which is part of the transposase–DNA synaptic complex. This method is used during Tn7 transposition⁹. As described below, we now understand these two mechanisms in substantial structural detail.

The first view of an active transposase

The structure of the Tn5 transposase synaptic complex consists of a dimer of ~55 kDa transposase subunits engulfing a pair of 20-base pair DNA fragments¹⁰ (Fig. 3a). These fragments mimic the ends of the Tn5 genome after excision from host DNA. The 3' hydroxyls are engaged in the transposase active sites. Thus, the structure provides a snapshot of the synaptic complex poised to 'pounce' on the target DNA.

Extensive protein–DNA contacts. The protein–DNA contacts in the structure are extensive and interwoven. Transposase is constructed from three functional domains — the N-terminal site-specific DNA binding domain, the catalytic core and the C-terminal protein–protein interaction domain — all of which contact DNA. Each transposase subunit makes multiple contacts with both DNA fragments and approximately two-thirds of the 8,000 Å² of solvent accessible surface buried upon formation of the dimer results from buried DNA surface. This is largely because the DNA fragment, which is bound specifically by the N-terminal domain of one subunit, has its terminal nucleotides in the active site of the other subunit. This interwoven subunit arrangement explains why synapsis of the DNA ends is an essential prerequisite for catalytic activity and helps avoid deleterious recombination events involving only a single transposon end.

Active site of the transposase. The Tn5 transposase–DNA complex structure also provides the first view of a transposase catalytic domain with DNA engaged in the active site. Members of the transposase protein family carry a signature motif of three acidic amino acids, the DDE-motif, that cluster together in the folded catalytic domain. These acidic residues have been proposed to coordinate divalent metal ions (presumably Mg²⁺

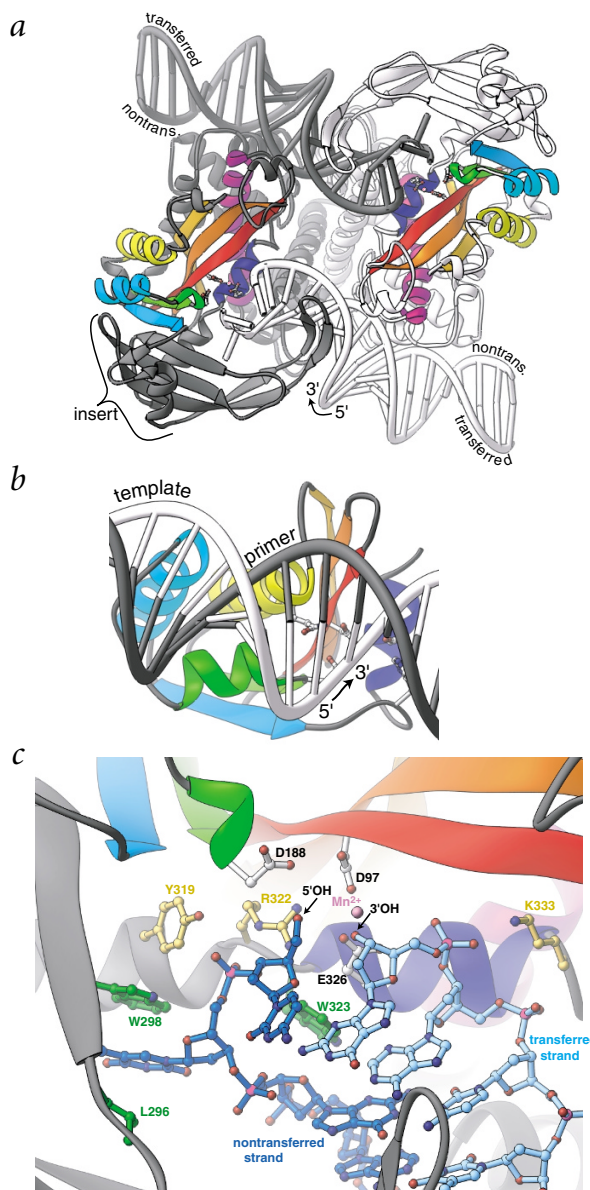
review

Fig. 3 Transposase–substrate interactions. **a**, Tn5 dimer bound to DNA. The conserved portions of the catalytic core domains are colored as in Fig. 1. Other portions of one subunit are in dark gray, with those of the second subunit in pale gray. The gray subdomain on the lower left and top right is the 'insertion region', which is present within the core domain. This region binds the DNA in the vicinity of the active site. The transferred and nontransferred strands are also labeled. Note that each DNA duplex interacts extensively with both subunits. **b**, Substrate interactions of RNase H. The RNase H domain of reverse transcriptase, taken from the structure of the intact enzyme complexed with duplex DNA, is shown along with the portion of the DNA that approaches the RNaseH active site. The conserved secondary structure elements of RNaseH are in approximately the same orientation as those of the left Tn5 transposase subunit in part (a). The template strand, which would be cleaved by RNaseH if it were RNA, is shown in white and the primer strand in dark gray. The polarity of the scissile strand in this structure is nearly opposite to that of the transferred strand in Tn5. **c**, Close-up of the protein–DNA interactions in the active site region of Tn5 transposase. The protein is colored as in Fig. 1; the transferred DNA strand is light blue, whereas the nontransferred strand is dark blue. The 'DDE' motif residues are shown with the bonds and carbon atoms in white and 'YREK' residues yellow (Glu 326 is part of both motifs). Three additional residues that interact with the tip of the hairpin are shown in green: Trp 298 and Leu 296 flank the 'flipped out' thymidine residue, while Trp 323 sits on the minor groove side of the hairpin, packing against both strands. Mutation of the residues analogous to Leu 296 and Trp 323 in Tn10 transposase specifically disrupts cleavage of the nontransferred strand⁵⁰, presumably by destabilizing this hairpin conformation.

in vivo, although Mn^{2+} also supports activity *in vitro*). The Tn5 structure has one Mn^{2+} ion in each active site, coordinated by two DDE-motif residues (Asp 97 and Glu 326) and the 3'OH of the transferred DNA strand. This Mn^{2+} is attractively positioned for a role in activating this 3'OH DNA end for nucleophilic attack on a new DNA phosphodiester.

How many metal ions? Although only one metal ion is observed in the current Tn5 structure, many enzymes that catalyze phosphoryltransfer reactions use a two-metal ion mechanism¹¹. Currently, there is evidence favoring both one- and two-metal models for the transposase active center. Evidence supporting a one-metal mechanism includes data indicating that the transposase cousin RNase H may require only a single catalytic metal ion¹². However, RNase H may not be a typical representative of a transposase catalytic domain; the makeup of the RNase H active site is somewhat different from that of transposases (see Fig. 1 legend). Furthermore, the structure of nucleic acid-bound HIV reverse transcriptase, which contains an RNase H domain, indicates that the cleaved strand of the substrate crosses the RNase H active site in the opposite direction to that seen in the Tn5 structure¹³ (Fig. 3b). Crystallographic studies of metal binding to retroviral integrases have been ambiguous: in the presence of Mg^{2+} , Mn^{2+} or Ca^{2+} , a single cation is seen in the active site but with Zn^{2+} or Cd^{2+} , two ions are seen (reviewed by Wlodawer)¹⁴. However, as has been seen with other phosphotransferases^{15,16}, the scissile phosphate of the substrate may play an important role in coordinating metal ions in the active site. There are no RNase H or integrase structures available with substrate docked in the active site, and the particular substrate DNA used in the Tn5 structure lacks the terminal 5' phosphate on the nontransferred strand, which should lie in the active site. Therefore, further experiments are needed to resolve how metal ions are bound and the exact roles they play in the transposase catalytic strategy.

How the hairpin intermediate forms. The view of the DNA termini within the active site of the Tn5 structure provides a clear picture of how one potentially difficult step in the reaction pathway is achieved: cleavage of the nontransferred strand *via* the DNA hairpin intermediate (see Fig. 2). The DNA caught in the active site of Tn5 transposase is held in a hairpin structure, despite the lack of a covalent connection between the two DNA



strands (Fig. 3c). The nucleotides in the final two positions of the DNA are unpaired, and the phosphodiester backbone of the nontransferred strand is distorted such that its 5' end is within 3.5 Å of the active 3'OH of the other strand. Multiple protein–DNA contacts stabilize this bent conformation. Tn5 transposase contains a large insertion (compared to many other family members) in the catalytic core domain, and this region plays a central role in stabilizing the hairpin. It provides a hydrophobic binding pocket for the penultimate base of the nontransferred strand, which is 'flipped out' of the double helix. In addition, the 5' phosphate of this nucleotide is positioned by contacts with several residues, including Tyr 319 and Arg 322 of the YREK motif (where the E is the catalytic glutamate, here Glu 326). This motif has long been recognized as conserved in some transposase subfamilies, but its function was unknown¹⁷. It is now clear that the first two residues contribute functional groups to the 'hairpinning' active site. Arg 322 is also poised to interact with the scissile phosphate and

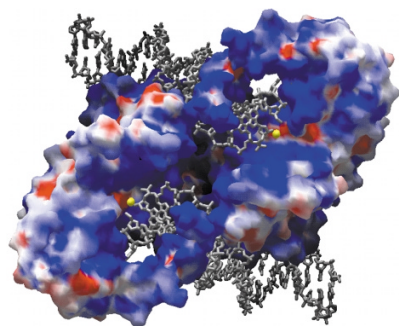


Fig. 4 Potential target-binding groove in the Tn5 transposase–DNA complex. The molecular surface of the protein portion of the Tn5 transposase complex is shown, colored according to electrostatic potential (with red = negative and blue = positive). The orientation is the same as in Fig. 3a. DNA is shown in white, and the Mn^{2+} ions in the active sites are yellow. A deep, positively charged cleft extends between the two active sites. This figure was prepared with GRASP⁵¹.

may play a catalytic role in other reaction steps as well. Mutation of the analogous arginine in Tn10 transposase obliterates all catalytic activities of the protein¹⁸.

Questions that remain. The Tn5–DNA complex gives us a wonderfully informative snapshot of the synaptic complex at one point along the reaction pathway, but many questions remain regarding the other steps in transposition shown in Fig. 2. For example, how does the synaptic complex interact with the flanking host DNA and the target DNA, both absent from the structure? The current structure provides some hints regarding target DNA interactions. The scissile phosphates of the two strands of the target duplex are 9 bp apart, and the two 3' hydroxyls within the Tn5 active sites lie 41 Å apart, a distance only slightly longer than the 35 Å expected for this 9 bp spacing. There is a long, positively charged trench stretching between these active sites (Fig. 4) that, with some movement of loops, could accommodate duplex DNA. Thus, the structure suggests an attractive hypothesis for how to dock the central portion of the target DNA within the active synaptic complex. Beyond this portion, significant conformational changes must occur in either the target DNA or the protein to avoid steric clashes. However, there is good evidence for large DNA distortions just outside the scissile phosphates in the related Tn10–target DNA complex¹⁹.

How universal is the Tn5 structure model?

Transposases use a common mechanism for DNA cleavage and joining and have structurally conserved catalytic core domains (Fig. 1). There is, however, substantial mechanistic and structural diversity within this protein family. From a structural viewpoint, domains outside the catalytic core are not highly conserved. Mechanistically, the spacing between attack sites on the two target DNA strands varies — for example, 5 bp is more common than the 9 bp used by Tn5 — as does the fate of the nontransferred DNA strand. Thus, it is interesting to consider what features of the Tn5 transposase–DNA complex are likely to be conserved among many family members and what features will emerge as specific for elements occupying a particular mechanistic and biological niche.

Common DNA binding and synopsis strategies

Like Tn5 transposase, many transposases have a three-domain organization. However, aside

from the catalytic core, these domains are not highly conserved. The C-terminal domains are especially diverse and sometimes absent, and no unifying structural or functional features of these regions have emerged, although many C-terminal domains are implicated in DNA contacts. The N-terminal regions are more similar, with most family members containing at least one small DNA binding domain N-terminal to the catalytic core. This region is often known to be critical in specific recognition of the ends of the transposon DNA. Where the structure of these N-terminal domains is known, it is always a variation on a helix–turn–helix motif. HIV integrase has a single N-terminal domain that contains a zinc-binding site and a helix–turn–helix^{20,21}. The N-terminal regions of the transposases from phage Mu (MuA)^{22,23} and *Caenorhabditis elegans* Tc3 (ref. 24) both carry two sequential subdomains involved in DNA recognition. Both subdomains of MuA and one from Tc3 contain a homeodomain-like helix–turn–helix motif. For comparison, Tn5 transposase carries a single N-terminal domain, whose 2nd, 3rd, and 4th helices correspond roughly to the helices of the homeodomains of MuA and Tc3, although the angles among these helices differ significantly from the canonical structural motif.

The striking interwoven arrangement of the N-terminal DNA binding domains and the catalytic cores on the two DNA ends in the Tn5 transposase structure is very likely to be found in other transposase synaptic complexes. This type of crisscrossed architecture has been proposed for the MuA–DNA synaptic complex^{25–28}. Extensive biochemical experiments reveal that the N-terminal DNA binding domain of one MuA subunit is bound specifically to one end of the Mu DNA, whereas the catalytic domain of that same subunit promotes recombination of the other DNA end^{25–28}. Because the Tn5 and MuA transposases are not close relatives, the fact that the two synaptic complexes share this structural feature suggests that this will be a common theme among transposase–DNA complexes. This subunit organization clearly provides a structural basis for the requirement of substrate DNA for complex assembly and insures that the two DNA ends are paired prior to the irreversible step of DNA cleavage.

The number of subunits is not conserved

One aspect of the Tn5 complex is almost certain to prove universal:

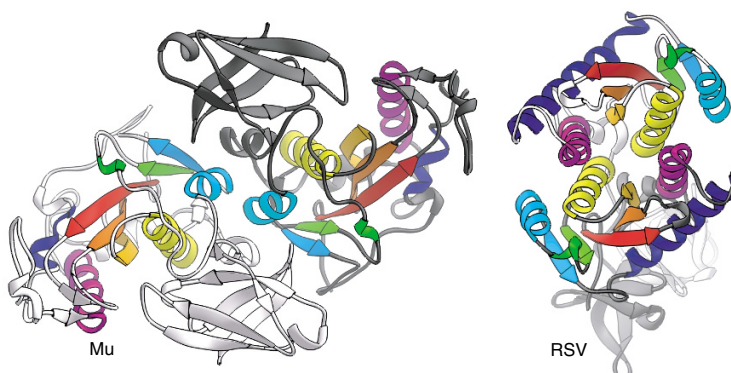
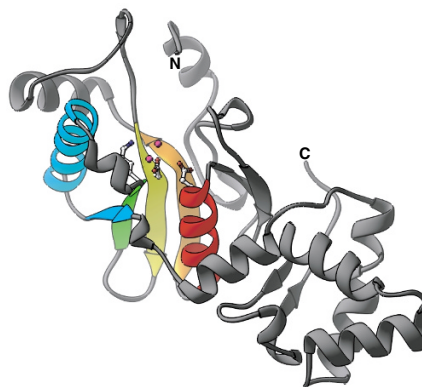


Fig. 5 Dimer interfaces in other transposase structures. Left, pairing of MuA catalytic domains found in the crystal. Right, the recurring dimer found in retroviral integrase structures. (RSV integrase is shown; PDB accession code 1COM)³². The catalytic cores are colored as in Fig. 1.

review

Fig. 6 TnsA from Tn7. The secondary structure elements conserved among type II restriction enzymes are colored in rainbow order from red to blue, the catalytic residues are shown in ball-and-stick and the two magnesium ions found in the active site are shown in pink⁴¹ (PDB accession code 1FLZ).



transposase–DNA complexes will undoubtedly all have two catalytic domains that promote the cleavage and DNA strand transfer reactions on the two ends of the element DNA. In support of this assertion, biochemical experiments argue that both the MuA and Tn10 transposase synaptic complexes contain two functional active centers^{18,25,27}. However, among transposases the pair of active sites appears to reside in several different sized oligomers and to be positioned by different protein–protein contacts at different spacing along the target DNA. Furthermore, the C-terminal region of Tn5 transposase, which lies at the center of the dimer interface, is quite poorly conserved when compared to other transposase sequences²⁹. Thus, the transposase family is similar to the type II restriction enzymes, where a common catalytic module is found arranged in a variety of different oligomers, depending on the relative locations of the two scissile bonds in the DNA substrate¹⁵. Below, we discuss the protein interfaces observed in transposase structures that are not docked on DNA and the possible roles of these interfaces in synaptic complexes.

In all crystal structures of retroviral integrases to date, including the recently described structures carrying the C-terminal domain^{30–32}, the catalytic domains form a dimer involving the helices colored yellow and purple in Figs 1 and 5 (ref. 14). As the corresponding helices are not involved in protein–protein contacts in the Tn5 complex, and as the DNA plays a pivotal role in Tn5 dimerization, one might question the biological significance of this interface. However, the relevant oligomer of integrase is probably larger than a dimer and has been suggested to be a tetramer (reviewed in refs 33,34) or even an octomer³⁵. This recurring interface may therefore represent a pair of subunits that bind the same viral end, essentially taking the place of a single Tn5 transposase monomer. Using the new two-domain integrase structures, several models of how DNA may dock with integrase have been proposed. The flexibility of the tether between the catalytic and C-terminal domains plays a key role in these models by allowing the C-terminal domains of different subunits to bind different portions of the DNA substrate.

As with the retroviral integrases, MuA functions as a larger multimer than Tn5 transposase³⁶. A minimal active transpososome contains a tetramer of MuA and the two ends of the Mu DNA. In crystals of the MuA catalytic domain, weak dimers were noted with an interface centered on the turquoise helix in Figs 1 and 5 (ref. 37). It is intriguing that when the catalytic domain of one subunit of this MuA dimer is aligned with the catalytic domain of Tn5 transposase, the C-terminal β -barrel of the other MuA subunit lies roughly in the same place as the inserted region present in the Tn5 catalytic domain (described above, and labeled in Fig. 3a.) A homologous inserted region is absent from the MuA core domain. The barrel in MuA is important in target DNA interactions and can be donated by a different subunit than provides the DDE-motif residues of the active site^{38,39}. Thus, again, one such MuA dimer may substitute for a single Tn5 protomer.

An alternative solution to cleaving the nontransferred strand

The ‘hairpinning’ mechanism used by Tn5 and Tn10 transposases to generate DNA double-strand breaks is an elegant and economic solution to this need to cleave the transposon DNA out of

the host cell’s genome. This mechanism is, however, far from universal⁴⁰. Recent work gives us structural insight into a second cleavage mechanism, that used by another bacterial transposon, Tn7. Unlike Tn5 and Tn10, which carry a single transposition protein, Tn7 encodes five proteins that participate directly in recombination. One of these enzymes is a site-specific nuclease dedicated to cleaving the nontransferred strand. This protein, TnsA, is an essential component of the Tn7 transposase–DNA synaptic complex along with the DDE-motif protein, TnsB⁹. The catalytic domain of TnsA, however, is structurally unrelated to transposases but has a topology and active center highly reminiscent of a restriction enzyme⁴¹ (Fig. 6). The second domain of TnsA contains a helix–turn–helix motif, which is probably involved in protein–protein rather than protein–DNA interactions. It will be fascinating to see how this nuclease is docked with the TnsB transposase on the DNA in what promises to be at least a $\alpha_2\beta_2$ tetrameric complex.

Further directions: complex dynamics and regulation

A central unanswered question regarding transposase structure and function is how the different DNA segments are positioned and repositioned within the active sites as recombination progresses. As discussed above, it is now very clear that a single active center promotes multiple distinct chemical steps. For Tn5 and Tn10, there are a total of four reaction steps (Fig. 2): cleavage of the transferred strand, hairpinning, cleavage of the DNA hairpin and DNA strand transfer to covalently join the cleaved DNA to a new target site. The active center, which binds and stabilizes the intermediates during these phosphotransfer reactions, must therefore be occupied by three different phosphate groups in succession: the first one from the transferred DNA strand, then one from the nontransferred strand and finally one from the target DNA. Moreover, when a particular phosphate group leaves the active site, it does not fully dissociate but remains in the complex during the next reaction step. An interesting series of conformational changes thus appears to be required to rearrange the scissile DNA strands between subsequent recombination steps. Are these different scissile DNA strands always bound by the active site in the same fashion? Or are there multiple binding ‘pockets’ within the active site? Recent experiments with Tn10 transposase using chiral phosphorothioates at the scissile phosphate have shed light on this question⁴². Considered together with the structure of the Tn5 complex¹⁰, these biochemical data suggest that the scissile phosphate of the target DNA is not bound in exactly the same orientation as the scissile phosphate of the nontransferred strand during hairpin formation or cleavage. Furthermore, viewing the Tn5 structure suggests that the 3′ end



of the transferred strand may remain almost stationary throughout recombination, while the various phosphates 'come and go' from the catalytic center. Clearly, it will be essential to have structures of transposase family members 'caught in the act' of catalyzing all the reaction steps necessary for a round of recombination to fully understand this process.

Tn5 transposase is an outstanding example from which to learn the basic structural and mechanistic features of recombination largely because of its relative simplicity. Only the transposase and its DNA binding sites are required for activity. However, mobile elements often require accessory proteins that aid in complex formation and regulation. For example, Tn10, despite its relatively close relationship to Tn5, requires the DNA bending protein integration host factor (IHF) to form a functional synaptic complex. Likewise, assembly of MuA transposase involves the two DNA architectural proteins HU, a histone-like DNA binding protein, and IHF. Furthermore, both Mu and Tn7 transposition utilize regulatory activator proteins

that are ATP-dependant molecular switches⁴³. These proteins have central roles in choosing an appropriate insertion site by providing the transposable elements a means to avoid destroying their own DNA by 'auto-insertion'. A similar function appears to be provided by the barrier-to-autointegration factor (BAF) protein during retroviral integration⁴⁴. How such accessory factors interact with the core transposition machinery and affect activity are fascinating subjects, begging for structure-based insights.

Acknowledgments

We thank B. Burton, I. Goldhaber-Gordon, L. Roldan, T. Sokolsky, T. Williams, K. Swinger and C. Correll for helpful comments on the manuscript. Work on transposases in T.A.B.'s lab is supported by the NIH and in P.A.R.'s lab by the University of Chicago Faculty Research Fund. T.A.B. is an employee of the Howard Hughes Medical Institute.

Received 22 January, 2001; accepted 20 February, 2001.

- Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Yang, W. & Steitz, T.A. Recombining the structures of HIV integrase, RuvC and RNase H. *Structure* **3**, 131–134 (1995).
- Rice, P., Craigie, R. & Davies, D.R. Retroviral integrases and their cousins. *Curr. Opin. Struct. Biol.* **6**, 76–83 (1996).
- Mizuuchi, K. Polynucleotidyl transfer reactions in site-specific DNA recombination. *Genes Cells* **2**, 1–12 (1997).
- Kennedy, A.K., Guhathakurta, A., Kleckner, N. & Haniford, D.B. Tn10 transposition via a DNA hairpin intermediate. *Cell* **95**, 125–134 (1998).
- Bhasin, A., Goryshin, I.Y. & Reznikoff, W.S. Hairpin formation in Tn5 transposition. *J. Biol. Chem.* **274**, 37021–37029 (1999).
- Fugmann, S.D., Lee, A.I., Shockett, P.E., Villey, I.J. & Schatz, D.G. The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu. Rev. Immunol.* **18**, 495–527 (2000).
- Kim, D.R., Park, S.J. & Oettinger, M.A. V(D)J recombination: site-specific cleavage and repair. *Mol. Cell* **10**, 367–374 (2000).
- Sarnovsky, R.J., May, E.W. & Craig, N.L. The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *EMBO J.* **15**, 6348–6361 (1996).
- 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).
- Steitz, T.A. & Steitz, J.A. A general two-metal ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. USA* **90**, 6498–6502 (1993).
- Keck, J.L., Goedken, E.R. & Marqusee, S. Activation/attenuation model for RNase H. A one-metal mechanism with second-metal inhibition. *J. Biol. Chem.* **273**, 34128–34133 (1998).
- Huang, H., Chopra, R., Verdine, G.L. & Harrison, S.C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282**, 1669–1675 (1998).
- Wlodawer, A. Crystal structures of catalytic core domains of retroviral integrases and role of divalent cations in enzymatic activity. *Adv. Virus Res.* **52**, 335–350 (1999).
- Kovall, R.A. & Matthews, B.W. Type II restriction endonucleases: structural, functional and evolutionary relationships. *Curr. Opin. Chem. Biol.* **3**, 578–583 (1999).
- Steitz, T.A. DNA polymerases: structural diversity and common mechanisms. *J. Biol. Chem.* **274**, 17395–17398 (1999).
- Rezsohazy, R., Hallet, B., Delcour, J. & Mahillon, J. The IS4 family of insertion sequences: evidence for a conserved transposase motif. *Mol. Microbiol.* **9**, 1283–1295 (1993).
- Bolland, S. & Kleckner, N. The three chemical steps of Tn10/IS10 transposition involve repeated utilization of a single active site. *Cell* **84**, 223–233 (1996).
- Pribil, P.A. & Haniford, D.B. Substrate recognition and induced DNA deformation by transposase at the target-capture stage of Tn10 transposition. *J. Mol. Biol.* **303**, 145–159 (2000).
- Cai, M. *et al.* Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nature Struct. Biol.* **4**, 567–577 (1997).
- Eijkelenboom, A.P. *et al.* The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: a three-helix bundle stabilized by zinc. *Curr. Biol.* **7**, 739–746 (1997).
- Schumacher, S. *et al.* Solution structure of the Mu end DNA-binding I β subdomain of phage Mu transposase: modular DNA recognition by two tethered domains. *EMBO J.* **16**, 7532–7541 (1997).
- Clubb, R.T., Schumacher, S., Mizuuchi, K., Gronenborn, A.M. & Clore, G.M. Solution structure of the I γ subdomain of the Mu end DNA-binding domain of phage Mu transposase. *J. Mol. Biol.* **273**, 19–25 (1997).
- van Pouderooyen, G., Ketting, R.F., Perrakis, A., Plasterk, R.H. & Sixma, T.K. Crystal structure of the specific DNA-binding domain of Tc3 transposase of *C. elegans* in complex with transposon DNA. *EMBO J.* **16**, 6044–6054 (1997).
- 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).
- Savilahti, H. & Mizuuchi, K. Mu transpositional recombination: donor DNA cleavage and strand transfer in trans by the Mu transposase. *Cell* **85**, 271–280 (1996).
- 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).
- Aldaz, H., Schuster, E. & Baker, T.A. The interwoven architecture of the Mu transposase couples DNA synthesis to catalysis. *Cell* **85**, 257–269 (1996).
- Haren, L., Ton-Hoang, B. & Chandler, M. Integrating DNA: transposases and retroviral integrases. *Annu. Rev. Microbiol.* **53**, 245–81 (1999).
- Chen, J.C. *et al.* 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 (2000).
- Chen, Z. *et al.* X-ray structure of simian immunodeficiency virus integrase containing the core and C-terminal domain (residues 50–293) — an initial glance of the viral DNA binding platform. *J. Mol. Biol.* **296**, 521–533 (2000).
- Yang, Z.N., Mueser, T.C., Bushman, F.D. & Hyde, C.C. Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase. *J. Mol. Biol.* **296**, 535–548 (2000).
- Asante-Appiah, E. & Skalka, A.M. HIV-1 integrase: structural organization, conformational changes, and catalysis. *Adv. Virus Res.* **52**, 351–369 (1999).
- Esposito, D. & Craigie, R. HIV integrase structure and function. *Adv. Virus Res.* **52**, 319–333 (1999).
- Heuer, T.S. & Brown, P.O. Photo-crosslinking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex. *Biochemistry* **37**, 6667–6678 (1998).
- Lavoie, B.D., Chan, B.S., Allison, R.G. & Chaconas, G. Structural aspects of a higher order nucleoprotein complex: induction of an altered DNA structure at the Mu-host junction of the Mu type 1 transpososome. *EMBO J.* **10**, 3051–3059 (1991).
- Rice, P. & Mizuuchi, K. Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell* **82**, 209–220 (1995).
- Kremetsova, E., Giffin, M.J., Pincus, D. & Baker, T.A. Mutational analysis of the Mu transposase. Contributions of two distinct regions of domain II to recombination. *J. Biol. Chem.* **273**, 31358–31365 (1998).
- Namgoong, S.Y. *et al.* Mutational analysis of domain I β of bacteriophage Mu transposase: domains I α and I β belong to different catalytic complementation groups. *J. Mol. Biol.* **275**, 221–232 (1998).
- Turlan, C. & Chandler, M. Playing second fiddle: second-strand processing and liberation of transposable elements from donor DNA. *Trends Microbiol.* **8**, 268–274 (2000).
- Hickman, A.B. *et al.* Unexpected structural diversity in DNA recombination: the restriction endonuclease connection. *Mol. Cell* **5**, 1025–1034 (2000).
- 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).
- Stellwagen, A.E. & Craig, N.L. Mobile DNA elements: controlling transposition with ATP-dependent molecular switches. *Trends Biochem. Sci.* **23**, 486–490 (1998).
- Lee, M.S. & Craigie, R. A previously unidentified host protein protects retroviral DNA from autointegration. *Proc. Natl. Acad. Sci. USA* **95**, 1528–1533 (1998).
- Carson, M. Ribbons. *Macromol. Crystallogr. B* **277**, 493–505 (1997).
- Lavoie, B.D. & Chaconas, G. Transposition of phage Mu DNA. *Curr. Top. Microbiol. Immunol.* **204**, 83–102 (1996).
- Kleckner, N., Chalmers, R.M., Kwon, D., Sakai, J. & Bolland, S. Tn10 and IS10 transposition and chromosome rearrangements: mechanism and regulation *in vivo* and *in vitro*. *Curr. Top. Microbiol. Immunol.* **204**, 49–82 (1996).
- Reznikoff, W.S. *et al.* Tn5: A molecular window on transposition. *Biochem. Biophys. Res. Commun.* **266**, 729–734 (1999).
- Craig, N.L. Transposon Tn7. *Curr. Top. Microbiol. Immunol.* **204**, 27–48 (1996).
- Bolland, S. & Kleckner, N. The two single-strand cleavages at each end of Tn10 occur in a specific order during transposition. *Proc. Natl. Acad. Sci. USA* **92**, 7814–7818 (1995).
- Nicholls, A., Sharp, K.A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).