Mechanisms of DNA Transposition

ALISON B. HICKMAN and FRED DYDA

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, 5 Center Dr., Bethesda, MD 20892, USA

ABSTRACT DNA transposases use a limited repertoire of structurally and mechanistically distinct nuclease domains to catalyze the DNA strand breaking and rejoining reactions that comprise DNA transposition. Here, we review the mechanisms of the four known types of transposition reactions catalyzed by (1) RNase H-like transposases (also known as DD(E/D) enzymes); (2) HUH single-stranded DNA transposases; (3) serine transposases; and (4) tyrosine transposases. The large body of accumulated biochemical and structural data, particularly for the RNase H-like transposases, has revealed not only the distinguishing features of each transposon family, but also some emerging themes that appear conserved across all families. The more-recently characterized single-stranded DNA transposases provide insight into how an ancient HUH domain fold has been adapted for transposition to accomplish excision and then site-specific integration. The serine and tyrosine transposases are structurally and mechanistically related to their cousins, the serine and tyrosine site-specific recombinases, but have to date been less intensively studied. These types of enzymes are particularly intriguing as in the context of site-specific recombination they require strict homology between recombining sites, yet for transposition can catalyze the joining of transposon ends to form an excised circle and then integration into a genomic site with much relaxed sequence specificity.

In this chapter, we provide an overview of the fundamental concepts of DNA transposition mechanisms. Our aim is to emphasize basic themes and, in this effort, we will focus on specific illustrative cases rather than attempt an exhaustive review of the literature. We hope that the selected references will point the curious reader towards the landmark studies in the field as well as some of the most exciting recent results. We also direct the reader to other recent reviews (1–3).

DNA transposases are enzymes that move discrete segments of DNA called transposons from one location in the genome (often called the donor site) to a new site without using RNA intermediates. DNA transposases are usually encoded by the mobile element itself (in which case they are “autonomous” transposons). However, some transposons are missing a self-encoded transposase yet have ends that can be recognized by a transposase encoded somewhere else in the genome, and thus are “non-autonomous” (Siguier et al., this volume). Although logic suggests that all DNA transposons are moved by transposases, the term was originally reserved for those enzymes that do not require significant regions of homology between any part of the transposon and the sites to which they are moved, the so-called target (or insertion or integration) sites. As biology is not always neat and tidy, transposases can exhibit a spectrum of homology requirements and vagaries of terminology have arisen such that certain transposases are sometimes referred to as “resolvases” or by the generic term “recombinases”.

From a mechanistic perspective, there are only a few ways in which transposases catalyze the required DNA strand breakage and rejoining reactions that comprise transposition (1), so from a structural point of view there are only a few different types of catalytic domain found in transposases. The catalytic domain topology,
or its “fold”, is a convenient way to classify DNA transposases although they have also historically been grouped according to whether or not their strand breakage mechanism involves a covalently-bound transposase/DNA intermediate. Other modes of classification include whether transposition proceeds through a replicative or non-replicative pathway, and whether transposition involves double- or single-stranded forms of DNA. These are useful distinctions, and it is worth noting that even within a single one of these categories, different DNA transposases can exhibit variations in their mechanisms (see also Siguier et al., this volume). It is this “similar-yet-distinct” property of transposition mechanisms that is part of their ongoing fascination.

CHEMISTRY OF DNA CLEAVAGE AND STRAND TRANSFER

There are four distinct types of catalytic nuclease domain folds (4) that are known to be used by DNA transposases to carry out the chemical reactions of transposition (Table 1). The most common is the so-called RNase H-like fold, sometimes also referred to as a DD(E/D) domain or the “retroviral integrase fold” as it has three catalytic acid residues at its active site (5–7). The second major type of catalytic domain is seen for those transposases that act on single-stranded DNA (ssDNA) and is known as an HUH domain (8). The serine transposases (such as those of IS607, Tn5397, and Tn5541 (9, 10)) and tyrosine transposases (exemplified by those of CTnDOT and Tn916 (11)) are predicted to have the same catalytic domain folds as serine and tyrosine site-specific recombinases, respectively. This last aspect is illustrated in Table 1: the four catalytic nuclease domains found in DNA transposases are also used by other enzymes that rearrange DNA such as retroviral integrases, invertases, resolvases, site-specific recombinases, and the RAG-1 recombinase involved in V(D)J recombination.

1. DNA transposases with RNase H-like catalytic domains

At the core of RNase H-like transposases is an active site in which three catalytic acidic residues (DDE or DDD) coordinate two metal ions - no doubt Mg²⁺ in vivo - in order to activate either a water molecule or a 3′-OH group of a nucleotide for nucleophilic attack on a phosphodiester bond (Figure 1A). This type of active site was first structurally investigated and a chemical mechanism proposed in the context of the 3′-5′ exonuclease of the Klenow fragment of E. coli DNA polymerase I (12). More recently, the mechanism has been further investigated and described for RNase H from Bacillus halodurans (13, 14) and analyzed computationally with quantum mechanics/molecular mechanics (QM/MM) methods (15, 16).

There are two types of reactions catalyzed by RNase H-like active sites in transposases: (i) nucleophilic attack by an activated water molecule on a scissile phosphate at or close to a transposon end to break a DNA

<table>
<thead>
<tr>
<th>TABLE 1 Examples of proteins containing the four types of nuclease catalytic domains found in DNA transposases and other enzymes that rearrange DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FUNCTION</strong></td>
</tr>
<tr>
<td><strong>RNase H-like (DDE)</strong></td>
</tr>
<tr>
<td>Transposase</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Bacteriophage or viral integrase</td>
</tr>
<tr>
<td>DNA invertase</td>
</tr>
<tr>
<td>DNA resolvase</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
phosphodiester bond such that a 3'-OH group and a 5'-phosphate are generated at the cleavage site; and (ii) transesterification in which nucleophile attack by the 3'-OH of a terminal nucleotide of a DNA strand rearranges the connectivity of DNA strands by simultaneously cleaving one strand while covalently joining one to another. This latter reaction can be used to join one DNA strand to its opposite strand to form a hairpin, to join one transposon end to the other end to form a circular intermediate, or to integrate a transposon into a new site. Each reaction occurs via an in-line SN2 nucleophilic attack which occurs with inversion of the stereoconfiguration at the scissile phosphate being attacked (17, 18, 19). These two reactions, used in different combinations and on different DNA strands, have been deployed by transposases to generate a plethora of reaction pathways for transposition.

One important characteristic shared by DNA transposases is that the hydrolysis of high energy cofactors is not required for any of the steps. (Although the Drosophila P element transposase contains a GTP-binding domain, GTP hydrolysis is not required for transposition (Majumdar & Rio, this volume)). That said, transposition reactions often require supercoiling of either the transposon ends or the target DNA, in part because the potential energy stored in supercoils can drive the reaction forward when DNA is cleaved. At the same time, it has been demonstrated that in some cases hydrolysis of high energy cofactors is required for the disassembly of the final product protein-DNA complex (20): transposition reactions are often slow and proceed through several steps of assembling elaborate protein-DNA complexes in which the final assembly is so stable that taking it apart requires energy. It is possible that the process of assembling ever-more stable complexes ensures the directionality of the reaction.

Where it has been examined, evidence suggests that in the context of an active complex in which a transposase is bound to its transposon end DNA (also known as a “synaptic complex” or “transpososome”), a single active site is able to catalyze both types of reactions at one end of the transposon and does so sequentially (19, 21). In other words, there is no evidence that during the reactions at one transposon end, one transposase monomer catalyzes hydrolysis and a separate monomer performs strand transfer.

The only divalent metal ion species that appear to be able to support all the steps of transposition are Mg\(^{2+}\) and Mn\(^{2+}\). Typically, reactions \textit{in vitro} are faster and more robust in Mn\(^{2+}\), although sometimes less accurate. Interestingly, Ca\(^{2+}\) generally does not support hydrolysis or transesterification and acts as an inhibitor (22) despite an ionic radius (1.14Å) that does not differ much from those of Mg\(^{2+}\) (0.86Å) and Mn\(^{2+}\) (0.81Å). It appears that, at least for hydrolysis, the reduced charge transfer from water to Ca\(^{2+}\) results in less effective activation of the nucleophile (22). However, it has been reported that Ca\(^{2+}\) can catalyze strand transfer in the case of the bacteriophage Mu transposase (23; see also Harshey, this volume), perhaps indicating that this step where the nucleophile is a 3'-OH group is less stringent.

The two metal ions that are coordinated by the catalytic acidic residues of the RNase H-like fold and the scissile phosphate (Figure 1A) serve to precisely position the reacting groups, activate the nucleophile, and stabilize the pentacovalent transition state that is presumed to exist as the reactions are believed to be associative. The two metal ions most likely adopt distinct roles at each step of the transposition reaction in which one activates the nucleophile and the other stabilizes the leaving group (24–27). This has been particularly well-characterized for the RNase H-like catalytic domain of the prototype foamy virus (PFV) integrase, a close relative of DNA transposases, which shows that in the presence of metal ions, there is a symmetric organization of the metal ions on either side of the scissile phosphate (where the metals M\(_{A}\) and M\(_{B}\) both coordinate the same oxygen of the scissile phosphate and each binds to one of the side chain carboxyl oxygens of a active site Asp residue (28; Engelman & Cherepanov, this volume). This suggests that this particular nuclease active site is adept at coordinating alternating nucleophilic attack from first one side and then another to choreograph the multiple steps that comprise transposition (19).

2. DNA transposases with HUH catalytic domains

A large number of DNA transposases use an entirely different catalytic domain with an HUH nuclease fold (reviewed in (3)) to cut and join ssDNA (Figure 1B). The best characterized examples are the transposases of the prokaryotic IS200 (see He \textit{et al.} and Sigier \textit{et al.}, this volume) and IS91 families of insertion sequences. In the case of the IS200 family, the transposition mechanism has been established experimentally in detail, and shown to require ssDNA forms of both the donor and target DNAs (29–35). HUH catalytic domains are used by a wide variety of proteins that occupy different biological niches, initiating processes such as plasmid rolling circle replication, the conjugative transfer of plasmids between cells, and the replication of paroviruses such as the adeno-associated virus.
CLEAVAGE

A. DD(E/D)

B. HUH

C. Ser transposase

D. Tyr transposase

STRAND TRANSFER

target DNA

flank from RE

flank from RE

180°
HUH nuclease domains use either one or two active site nucleophilic tyrosine residues (where the nucleophile is the OH group of the side chain) to cleave ssDNA through the formation of a 5′-phosphotyrosine covalent intermediate. If the 5′-phosphotyrosine linkage is subsequently attacked by a terminal 3′-OH group of another DNA strand, the covalent intermediate is said to be “resolved” i.e., the phosphotyrosine link is broken and the two DNA strands become connected in a strand transfer reaction. The name “HUH” refers to two conserved and catalytically required histidines (separated by a hydrophobic residue) that coordinate an essential divalent metal ion cofactor (Figure 1B) which in all likelihood is Mg^{2+} in cells. Mg^{2+} bonds and polarizes the scissile phosphate, setting it up for nucleophilic attack by either tyrosine or a 3′-OH group (36–38). Some HUH transposases require only one active site tyrosine (“Y1 transposases”) yet others need two closely spaced tyrosines (“Y2 transposases”) to complete the cycle of strand cleavage and rejoining.

For some HUH nucleases, the chemical steps can be supported by a wide range of divalent metal ions, suggesting that the active site is relatively tolerant. For example, TrwC, the HUH relaxase of plasmid R388 which catalyzes the initial nicking reaction of conjugative DNA transfer, can cleave ssDNA in the presence of Mg^{2+}, Mn^{2+}, Ni^{2+}, Zn^{2+}, Ca^{2+}, or Cu^{2+} (39).

Beyond the IS91 and IS200 families, there are several recently discovered families of mobile elements which have associated proteins with HUH domains and which may turn out to fit the definition of DNA transposons. For example, the ISCRs, or Insertion Sequence Common Regions, are associated with antibody resistance genes and seem likely to be mobile elements with transposases resembling those of the IS91 family, although their mobility has not yet been experimentally shown (40).

More closely related to the IS200 family of transposons are a group of Tnp\textsubscript{AREP} proteins (41, 42) (also known as RAYTS (43)) that are associated with repetitive extragenic palindromic sequences (or REPs). REP sequences form hairpins and have been found scattered throughout many bacterial genomes (41, 43); it seems likely that the REP sequences and Tnp\textsubscript{AREP} proteins are remnants of ssDNA transposons. Finally, the widespread eukaryotic helitron transposons (Thomas & Pritham, this volume) also appear to encode an HUH domain in their transposases but have not yet been demonstrated to be active in the myriad species in which they have been identified (44–46).

### 3. Serine transposases

There are a number of bacterial transposons that encode a serine transposase. These include insertion sequences (ISs) such as IS607 (47) which is proposed to use a circular intermediate of the IS to recombine with a target DNA (48) (Figures 1C & 2), and certain conjugative transposons such as Tn5397 from Clostridium difficile and its relatives which also move using a circular intermediate (49, 50). Serine transposases are assumed to share many catalytic features with the serine site-specific recombinases such as resolvases and invertases (reviewed in (51) and elsewhere in this volume), yet exhibit a relaxed or practically-absent requirement for homology between the recombining sites (i.e., the abutted transposon ends in the circular intermediate and the target site). Nevertheless, specificity of target-site selection is a continuum; for instance, Tn5397 displays a strong target site preference (52) whereas IS607 has very little insertion specificity (53).

Serine recombinases are predicted to have the same catalytic core fold as the structurally characterized γδ resolvase (54, 55). They have been classified into four

---

**Figure 1** Basic chemical reactions catalyzed by DNA transposases. (A) An RNase H-like active site, based on structures of PFV intasomes (28, 94, 95). The green DNA represents the cleaved dinucleotide, and orange is the target strand. Spheres indicate bound metal ions. (B) HUH nuclease active site acting on single-stranded DNA (based on PDB ID 2X06 of IS\textsubscript{DraZ} TnpA). Shown is the reaction that occurs at the transposon Left End (LE). After cleavage, the DNA flanking the LE (black) remains in the active site; upon exchange of α-helices between the two active sites of the dimeric transposase, the cleaved LE moves to the other monomer where it is joined to the cleaved RE to form a circular excised transposon (not shown). At the same time, the flanking DNA from the RE of the transposon switches active sites (as shown here in black) and subsequent joining results in a sealed donor backbone. (C) DNA cleavage catalyzed by a serine recombinase. The active serine is surrounded by many Arg residues. Upon 180° rotation of one dimer within a tetramer, one strand rotates out of the active site (green) while another rotates in (orange). (D) DNA cleavage catalyzed by a tyrosine recombinase. Crucial residues within the active site include a conserved RHR triad (for details, see also (181)). doi:10.1128/microbiolspec.MDNA3-0034-2014.f1
groups based on their domain organization and function (9) and, according to this classification, the Tn5397 and Tn4451 transposases are also known as “large serine” recombinases because - relative to γδ resolvase - they have an additional large C-terminal domain. The group of IS607-like transposases are distinguished from γδ resolvase by a reversal of the order of the DNA-binding domain and the catalytic domain within the primary sequence.

DNA cleavage by serine recombinases involves an active site serine nucleophile that attacks the scissile phosphate forming a covalent 5′-phosphoserine intermediate and a free 3′-OH group (Figure 1C). No divalent metal ion or any other cofactors are required, and key roles in catalysis are played by an array of arginine residues at the active site (56, 57). The resolution of the covalent intermediate by a terminal 3′-OH group of another strand results in strand transfer.

Recombination by serine recombinases is understood to occur in the context of a tetrameric complex in which each subunit cleaves one of the four strands of the recombining DNA duplexes (Figure 2). In the tetramer, resolution of the phosphoserine linkages and strand transfer occur after a dramatic subunit rotation where one of the dimers of the tetramer rotates 180 degrees around the other (55).

4. Tyrosine transposases
Most conjugative transposons (also known as Integrative Conjugative Elements, or ICEs; Figure 3) are mobilized by associated tyrosine transposases which are

![Diagram](https://example.com/diagram.png)

**FIGURE 2** Proposed pathway for transposon circle integration into target DNA catalyzed by a serine transposase. At the top, a tetrameric assembly is shown bringing together the abutted Left End (LE; orange) and Right End (RE; red) of an excised circular transposon with a target DNA (green). The reactions in the dashed box show how four cleavage reactions in which each active site serine becomes covalently attached to one strand of DNA, followed by a 180 degree rotation of the left-most dimer, leads to a re-organization of the strands. Resolution of the four covalent intermediates results in an integrated transposon. doi:10.1128/microbiolspec.MDNA3-0034-2014.f2
FIGURE 3 Pathway of conjugal transposition. Whether catalyzed by a serine or a tyrosine transposase, excision results in a circular intermediate in which the transposon ends are abutted. Only one of the strands of this intermediate is transferred to the recipient cell, and replication (new strands shown in blue) regenerates the double-stranded form in both cells. doi:10.1128/Microbiolspec.MDNA3-0034-2014.f3
believed to be structurally and mechanistically related to the well-characterized site-specific tyrosine recombinases (51, 58, 59) such as Cre (van Duyne, this volume), Flp (Jayaram et al., this volume; 60) and λ integrase (Landy, this volume) (2). The most extensively studied are the transposases of the Tn916 family of conjugative transposons (61) and the CTnDOT (Wood & Gardner, this volume) conjugative transposon. These appear to have adapted a mechanism of site-specific recombination that proceeds through a Holliday junction to catalyze the transposition steps.

Tyrosine transposases cleave DNA using an active site tyrosine residue to attack the scissile phosphate and to form a covalent 3'-phosphotyrosine linkage (Figure 1D). This cleavage polarity is the opposite to that exhibited by the other three types of transposases. By analogy to tyrosine recombinases, a tetrameric complex is believed to assemble in which each subunit cleaves one of the four strands of the recombining sites (Figure 4). In the excision step, recombination between the two transposon ends (attL and attR) results in the formation of a free circular intermediate in which the ends are abutted (Figure 4A). To generate this intermediate, tyrosine transposases make staggered cuts at the transposon ends in which pairs of active site tyrosine residues are sequentially covalently bound (Figure 4A). The circular intermediate is then transferred into the recipient cell through conjugation. Once there, another recombination event takes place and the intermediate is integrated into a target at a attB site (Figure 4B).

Conjugative transposons display a spectrum of targeting specificity in their requirements for homology between the attL/attR junction and the attB site (11). Some insert essentially randomly, others display relatively strict specificity reflecting the site-specific recombinase machinery at work, and yet others occupy a middle ground with strict albeit usually very short sequence requirements for where they will integrate (11, 61, 62). While most conjugative transposons use tyrosine or serine transposases, it has recently been discovered that some conjugative transposons are mobilized by RNase H-like transposases (63, 64).

**TRANSPOSITION PATHWAYS**

1. RNase H–like transposases
   (i) Replicative transposition

Replicative transposition couples transposition to extensive DNA replication and, in doing so, generates a second copy of the transposon at a new target site. This is an obvious mechanism that allows a mobile element to expand and proliferate within a genome. One of the best studied of these elements is bacteriophage Mu which uses replicative transposition to increase its copy number in infected cells (reviewed in 65 and Harshey, this volume). This mechanism is also employed by the Tn3 family of transposons (Nicolas et al., this volume). During phage Mu replicative transposition (66), the phage-encoded transposase, MuA, catalyzes two hydrolysis and transesterification reactions that occur sequentially on each transposon end with no intervening steps (Figure 5a): the first nucleophilic attack by water on the transposon end generates a 3'OH group that is then used to attack the site for transposon insertion. This strand of the transposon end is therefore known as the “transferred strand”. The end result are branched DNA structures at each transposon end; these are then substrates for a complex set of “handover” reactions between the transpososome and the replication fork (67, 68).

One key feature of replicative transposition is that it does not generate double-strand breaks (DSBs) at the transposon ends. In this sense, there is a mechanistic analogy with retroviral integration (see Engelman & Cherepanov, this volume): the donor DNA of retroviral integration is the blunt-ended linear product of reverse transcription and while integrase processes this by removing two nts from the viral transferred strand, no DSBs are needed. This mechanistic parallel is reflected in the close structural similarity between the catalytic domains of retroviral integrase (5) and the phage MuA transposase (6).

A different type of replicative transposition is carried out by the so-called “copy-and-paste” insertion sequences such as those of the large IS3 family, and has been extensively studied for the representative element IS911 (see Chandler et al., this volume). Replication is required by these transposons to generate an excised transposon circle that is the substrate for the integration step (Figure 5b). After the initial generation of a 3'OH group at one end of the transposon, this then attacks the same strand at the opposite end of the transposon to generate a “figure-of-eight” intermediate (i.e., what it looks like when the transposon is contained on a plasmid). This results in the joining of the two transposon ends by a ssDNA bridge. Replication by the host cell machinery then converts this intermediate to a transposon circle (69), which is the substrate for subsequent insertion into a new site. The integration step of IS911 transposition (70) requires two more strand cleavage events at the junction of the joined transposon ends to generate the 3'OH nucleophiles for insertion into a target.
In the case of IS911, either end can be used to initiate transposition (71). In a remarkably clever twist, formation of an IS911 transposon circle concomitantly generates a strong promoter at the site of the junction between the two abutted transposon ends (72). This promoter drives expression of the transposase, high levels of which are thought to be particularly important for integration as it would increase the probability of rapid insertion of the transposon into a new site before loss of the transposon circle.

**FIGURE 4** Proposed pathways of excision and integration by tyrosine transposases. (A) Transposon excision. (B) Transposon integration. doi:10.1128/microbiolspec.MDNA3-0034-2014.f4
Replicative

A. MuA
B. IS911
C. Mos1
D. Tn5
E. Hermes

Non-replicative

H2O
H2O
H2O
H2O
H2O

F.

two replication forks

= 3’-OH group

strand transfer

gap repair
Mechanisms of DNA Transposition

(ii) Cut-and-paste transposition: Excision

Many RNase H-like transposases catalyze “non-replicative” or “cut-and-paste” transposition. These enzymes must generate DSBs at their transposon ends, the prerequisite for liberating the mobile element from its donor site. After an initial hydrolysis reaction that cleaves one strand, different cut-and-paste transposases cleave the second strand and generate a DSB in a variety of different ways (73), differences that are also reflected in structural variations in the transposases (74). Discovering the details of second-strand processing is a particularly informative approach to understand transposition mechanisms.

After the initial nucleophilic attack by a water molecule to form the first 3′-OH group at a transposon end, one way in which the second strand break can be introduced is by another activated water molecule, as illustrated in Figure 5c for the Tc1 mariner pathway (see Tellier et al., this volume). Thus, the generation of DSBs is simply a case of two sequential strand cleavage events on opposite strands at the same end (75–77). The eukaryotic P element also uses this pathway, but is unusual in that there is an atypically large offset in the position of cleavage events on the two strands: whereas the transferred strand is cleaved at the transposon-donor junction, the other is cleaved 17 bp into the transposon end (78).

Alternatively, the 3′-OH generated during the first strand cleavage can serve as the nucleophile to attack the opposite, second strand at the same transposon end, in which case a hairpin is formed. There are two possible variations on this step. If the first strand that is cleaved is the one that will eventually be the transferred strand, then the resulting 3′-OH is on the transposon end, and attack by this on the second strand generates a hairpin on the transposon end (Figure 5d); this is the case for the prokaryotic IS4 family of transposases such as Tn10 and Tn5 (Haniford & Ellis, this volume; 79), as well as for the eukaryotic piggyBac transposases (Yusa, this volume; 80). For the transposition reaction to continue, the hairpin must be opened so that a 3′-OH group is available at the transposon end for the final transesterification step of integration, and this occurs by yet another nucleophilic attack by an activated water molecule.

In contrast, if the first cleavage reaction is on the “non-transferred strand”, then the resulting free 3′-OH is on the flanking DNA, and the hairpin intermediate is also formed on the flanking donor DNA (Figure 5e). Examples of transposases that use this pathway include the eukaryotic bat (Atkinson, this volume; 81) and CACTA transposases, as well as the related V(D)J RAG1/2 recombinase (Roth, this volume; 82) which almost certainly evolved from an ancient Transib transposon (83). In these cases, the hairpin on the flanking end DNA does not need to be opened up for the purposes of transposition, as the transferred strand 3′-OH is generated at the same time as hairpin formation. There are no known examples of prokaryotic transposases that form flanking hairpins during transposition.

It is clear from the accumulated experimental work over decades and on a variety of systems that, no matter how the second strand is processed, there is only one DD (E/D) active site involved at each transposon end, despite the observation that transpososomes always contain multiple transposase monomers. (One possible exception may be the P element with its 17-nucleotide staggered cuts; it has been proposed that the simplest model to explain this would be for two monomers of the tetrameric transposase to act on each end (84)). Mechanistic studies suggest that in the case of transposases that form transposon end hairpins, the cleavage and transesterification steps are accomplished with the transferred strand remaining in the transposase active site for all four chemical steps (19, 21), and major conformational rearrangements appear unnecessary. On the other hand, when the hairpin is formed on flanking DNA, the active site must somehow switch from the non-transferred strand where the first nicking reaction takes place to the transferred strand to catalyze the subsequent transesterification reaction. Thus, it seems likely that significant conformational rearrangements will be needed for this type of reaction.

One cut-and-paste transposition system, Tn7 from E. coli (Peters, this volume), has solved the issue of DSB generation in a unique way. Tn7 encodes five

---

**FIGURE 5** Transposition pathways for RNase H-like transposases. Arrows indicate sites of strand cleavage and the black dots indicate 3′-OH groups. Many pathways converge on essentially the same form of the excised transposon (highlighted with grey boxes). This linear intermediate is then integrated into target DNA as shown in (f). Target site duplications (TSDs) are generated when the cell repairs the gaps introduced by staggered strand transfer reactions. Adapted from (J). doi:10.1128/microbiolspec.MDNA3-0034-2014.f5
transposition-related proteins (TnsA, B, C, D and E) with TnsA/TnsB forming the heteromeric transposase (85, 86). TnsB, which contains a predicted RNase H-like catalytic domain, cuts the transferred strand generating the 3'-OH but the non-transferred strand cut is by TnsA, which is a restriction endonuclease-like nuclease (87). In a clear demonstration of the deep mechanistic relationship between cut-and-paste transposition and DSBs, TnsA active site mutants that render the transposase unable to cut the non-transferred 5' ends of Tn7 turn the system into a replicative transposon (88).

(iii) Insights from structure
The stage of the transposition reaction where the transposon ends have been cleaved but the target DNA is not yet bound has been captured in three-dimensional structures (Figure 6) of three different cut-and-paste transposases bound to DNA (89, 90, 91; reviewed in (92)). The structures have been extremely valuable in illuminating several aspects of DNA transposition mechanisms relevant to excision. To date, only one transpososome - that of MuA - has been structurally characterized at a later step in the reaction in which its ends have been inserted into target DNA (93; Figure 6D). The retroviral intasome has also been structurally characterized at various stages along the integration pathway and is discussed in detail elsewhere (28, 94, 95; Engelman & Cherepanov, this volume; Figure 6B).

How does a transposase recognize its own DNA? Clearly, all transposases must be able to specifically recognize both ends of their transposons. The types and organization of sequences at transposon ends that are important for transposase binding vary widely. In general, binding sites responsible for transposon end recognition do not extend to the very tip of the transposon where the cleavages occur but are slightly subterminal. Among the characterized systems, some transposases recognize a single stretch of DNA that contains enough basepairs to uniquely define a binding site (e.g., Tn5 (89)). Others have been shown to bind two different short sites that are close together using two distinct site-specific DNA binding domains within a single protein monomer (e.g., MuA (93, 96) and the mariner elements (90, 97)), or to bind multiple sites close to the transposon ends where each site binds a separate transposase monomer (e.g., Tn7 (98)).

The Tn5 transposase binds identical 19 bp sequences (known as Terminal Inverted Repeats (TIRs) or Inverted Terminal Repeats (ITRs)) that are found in reverse orientation at each transposon end. The crystal structure of the Tn5 transposase/TIR complex (Figure 6A) revealed that almost all of the 19 bp are contacted by protein (89), certainly sufficient for unique recognition of the ends. There are three distinct domains within the Tn5 transposase (99) (as there are for its close relative, the Tn10 transposase (100)), and residues contributed by all of them participate in DNA binding.

In contrast, the Mos1 transposase uses two small site-specific DNA binding Helix-Turn-Helix (HTH) domains to recognize two distinct subterminal segments within its ends (Figure 6C). This is most likely a feature of all Tc/mariner transposases. HTH domains and their variants are encountered in a wide range of DNA binding proteins (101, 102), and are similarly employed by another structurally characterized transposase, bacteriophage MuA (93, Figure 6D). Many IS families and several eukaryotic transposon superfamilies have transposases with N-terminal HTH domains implicated in DNA binding (103–105).
Other transposons also feature two different binding sites at their ends (106). For example, analysis of the structure of the Hermes hAT transposase bound to its TIRs (91; Figure 6E) suggests a bipartite binding mode in which the TIRs are bound by multiple domains of the transposase whereas an N-terminal BED-finger domain (missing in the structure in Figure 6E; 107) recognizes short subterminal repeats which are a characteristic feature of hAT transposons (Atkinson, this volume).

How do transposases synapse their two transposon ends?
In their active forms, both the Tn5 and Mos1 transposase are dimeric, yet they arrive at this point through different assembly pathways. For Tn5, the transposase is a monomer in solution in the absence of DNA (108), as are most prokaryotic DNA transposases that contain an RNase H-like catalytic domain and act on double-stranded DNA, and protein dimerization is believed to occur after each has been bound by one transposase monomer (79). The Tn5 transposase binds its transposon ends “in trans”, which means that the active site that is engaged in processing one transposon end is part of the polypeptide chain that encodes the DNA binding domain(s) that binds the other. Thus, dimerization is concomitant with transpososome formation. In contrast, Mos1 is a dimer in solution prior to DNA binding (109), and mariner elements such as Mos1 sequentially capture their transposon ends (110) in a defined order as the relative binding affinities to the two ends are different (111, 112). The Mos1 transpososome structure reveals that catalysis is again in trans (90), suggesting a recurring regulation mechanism that would ensure that both ends are located and bound before any chemical reactions begin. Regardless of the assembly pathway, for these transpososomes, there is one transposase monomer per end and, at each, one active site per end appears to perform all of the chemical steps.

Transposases that bind to multiple asymmetrically-organized sites within their transposon ends tend to form transpososomes that contain more than two transposase protomers. For example, among the prokaryotic systems, the MuA transpososome forms a series of distinct complexes with DNA - some involving DNA binding sites far removed from the bacteriophage ends - before it arrives at a tetrameric synaptic complex (Figure 6D) capable of initiating the chemical steps of transposition (66). The Tn7 transposase incorporates its ultimate target site into the synaptic complex before any chemical reactions begin (113, 114) and, after strand transfer, the Tn7 transpososome has been reported to contain one molecule of TnsD, at least 6 protomers of TnsB, and multiple copies of TnsA2C2 (115).

Among the euukaryotic transposons, the transposon ends of the Hermes transposase are typical of other hAT elements as its asymmetric ends are several hundred bp long and contain multiple apparently haphazardly arranged subterminal binding sites (91, 116). The active form of Hermes is a ring-shaped octamer in which eight N-terminal site-specific DNA binding domains are available to interact with these interior sites while presenting the two transposon TIRs to the catalytic sites of one of the dimers of the octameric assembly (91; Figure 6E). The P element transposase, which also has asymmetric ends (Majumdar & Rio, this volume), is reportedly tetrameric both prior to DNA binding and upon end synapsis (84). Sleeping Beauty, a resurrected vertebrate transposase of the Tc1/mariner family (117), is also proposed to form a tetrameric transpososome (118). It has been suggested that multimerization prior to DNA binding might be a way to down-regulate transposition activity (sometimes called over-production inhibition or OPI) (119).

(iv) Target binding and integration
In addition to binding its transposon ends, a DNA transposase must also bind the DNA into which it is going to insert its transposon, and the structures of the MuA transpososome and the PFV intasome bound to target DNA have been particularly instructive regarding this step of transposition (93, 95). For many transposition systems, target DNA binding is non-specific and the transposon can integrate essentially anywhere. On the other hand, a rare few integrate into specific sites such as Tn7 which integrates into a precise location downstream of the E. coli glmS gene (see Peters, this volume). This targeting by Tn7 is dependent on TnsD which site-specifically binds the target sequence (120, 121). However, Tn7 is very resourceful as it also encodes TnsE (122). When TnsE is incorporated in the Tn7 transpososome instead of TnsD, transposition is directed to chromosomal regions where replication is terminated and to DSBs (123). This target selection process is likely mediated by a direct physical interaction between TnsE and the β clamp replication processivity factor (124).

Other transposons exhibit varying degrees of target specificity. Among the RNase H-like transposases, the Tc1/mariner transposases insert at TA sites (125) and the piggyBac transposases always insert into a TTTA sequence (126). Other RNAse H-like transposases have preferred sites of integration. These involve distinct and often palindromic (127) patterns of base pairs (128–137), suggesting that these target site preferences might
reflect some other property other than sequence specificity, for example perhaps DNA bendability. Indeed, target DNA has been repeatedly observed to be bent when bound by transpososomes (e.g., for Tn7 (138), Tn10 (139), Mos1 (140), MuA (93)), and such bending has been proposed to be an effective mechanism to ensure the directionality of the reaction (93, 141). In the crystal structure of the MuA transpososome (Figure 6D), target DNA is bent through a total of 140° (93) whereas in the retroviral PFV intasome (Figure 6B), the target DNA is also bent but to less of an extent (25).

In another manifestation of the variability of transposition mechanisms, the point during the reaction at which target DNA is bound by the transpososome can differ. For some transposases, target DNA cannot be bound until both transposon ends are cleaved and the flanking DNA has been released; this might be a consequence of using the same protein surface for binding both flanking and target DNA (142, 143). For other transposases such as that of Tn7, the requirement is precisely the opposite and target DNA must be bound before strand cleavage is initiated (113, 138).

Integration into target DNA (Figure 5f) occurs via two transesterrification reactions involving the coordinated attack of the two free 3’-OH groups at the transposon ends on opposite strands of target DNA a defined distance (i.e., number of bp) apart. This staggered strand transfer generates short gaps at both sides of the integrated transposon that must be subsequently repaired, giving rise to target site duplications (TSDs). The length of the TSDs is a characteristic of each particular family of transposon, and is generally between 2 and 11 bp (Siguier et al., this volume and (105)). In mechanistic terms, the constancy of TSD size reflects that integration into two target strands occurs in the context of a transpososome assembly with a fixed structural relationship between two active sites within the transpososome.

Sometimes, the TSDs reflect aspects of prior steps of the transposition pathway. For example, piggyBac excision results in a four bp 5’ overhang on each end that is 5’TAAA (80), identical to its TATA target site requirement and, due to a four bp offset in the sites of strand transfer into target, identical to its TSDs (126). Upon insertion, the overhangs basepair with the offsets generated by the four bp stagger in nucleophilic attacks (80), and there are no gaps that have to be filled but simply a bond to be formed. Thus, the TSDs are only temporary as, when piggyBac excises, the flanking DNA is restored to its original sequence without the need for any DNA synthesis. This property has made piggyBac a particularly attractive system for in vivo genomic applications: during a cycle of insertion and excision, it leaves no permanent genomic marks behind (144). Other transposons, notably Sleeping Beauty and Tol2, are also finding wide use for genomic manipulation experiments (145–147).

One final aspect of target site selection that should be mentioned is the notion of target immunity. Certain transposons such as phage Mu, Tn7, and those of the Tn3 family possess the ability to distinguish self from non-self, and can avoid the suicidal step of integrating into themselves. In the case of Mu, immunity depends on a Mu-encoded ATPase, MuB (148, 149). For Tn7, the ATPase TnsC and the TnsB protein of the transposase work together to establish target immunity (150). Curiously, a similarly-functioning protein has not yet been identified for Tn3 transposons (Nicolas et al., this volume; 151) and the mechanism of target immunity is not currently understood for Tn3 and its relatives.

(v) Host proteins
Some RNase H-like transposases have been shown to require host proteins to carry out transposition, very often to bend or deform the transposon ends. This may be important when transposon binding sites are separated along the DNA yet need to be incorporated within the transpososome. Classical examples are the MuA, Tn5, and Tn10 transposases which rely on highly expressed DNA bending proteins such as IHF (“Integration Host Factor”) and HU (“Histone-like protein” from E. coli strain U93). For example, Mu transposition (Harshey, this volume) requires both IHF and HU; an IHF binding site is located within an enhancer sequence which is ∼900 bp away from the phage left end, and HU is needed to assemble a catalytically active transpososome but is not needed thereafter (152). The structure of the MuA transpososome has provided a valuable starting point for modeling how HU may participate in the assembly process (93). The Tn10 transposase also relies on IHF for transpososome assembly (153, 154). In contrast, the closely related Tn5 transposase does not appear to require IHF whereas both Tn5 and Tn10 use the host protein H-NS (histone-like nucleoid-structuring protein) to assist proper transpososome assembly in roles that are different in detail (155, 156).

There are also examples of eukaryotic transposases that rely on a DNA binding protein, HMGB (“high mobility group box”), in a similar assisting role. Sleeping Beauty uses HMGB1 (157), probably due to two far-spaced binding sites on both transposon ends although this does not appear to be a general property of mariner transposases. HMGB plays a similar role in V(D)J
recombination in the assembly of RAG1/2-RSS (Recombination Signal Sequence) complexes (158, 159). It would be surprising if HMGB did not participate in other as-yet biochemically uncharacterized eukaryotic transposition systems.

2. HUH transposases

(i) Transposon end recognition

Single-stranded DNA transposases of the IS200/IS605 family mobilize their transposons when they become accessible in single-stranded form, for example on the lagging strands during replication or during certain types of DNA repair (160, 161). Other HUH DNA transposases such as those of the IS91 family or helitrons may actively assist in generating ssDNA, either by recruiting a cellular helicase as has been suggested for IS91 (162) or by encoding a helicase domain as proposed for helitrons (44), but the mechanisms of these “rolling circle” transposons (163) are far from being firmly established.

Recently, the mechanism of IS200/IS605 transposition has been intensively studied through a series of genetic, biochemical, and structural experiments with the Helicobacter pylori IS608 and Deinococcus radiodurans ISdra2 transposases (29–35; He et al., this volume). One of the most important concepts to emerge is that neither the cleavage sites at the transposon ends nor the target site is recognized by a site-specific DNA binding domain of the transposase. Rather, the transposase mediates DNA-DNA interactions involving base-pairing between two ssDNA regions, and this directs the appropriate scissile phosphate into the active site. This same mechanism also ensures site-specificity of integration, which precisely targets either a tetra- or pentanucleotide sequence (164, 165).

TnpA is an obligatory dimer even in the absence of DNA, and locates its transposon ends by binding DNA hairpins that are formed by palindromic sequences located subterminal to the transposition tips (30; Figure 7). Binding is neither strictly in cis or in trans, as both monomers contribute to the binding of both hairpins. Directly 5′ of the recognition hairpins are “guide sequences” that basepair with bases at the cleavage sites (32). This arrangement is provocatively reminiscent of the use of RNA guide sequences located 5′ of CRISPR hairpins in microbial immune systems (166). Figure 7 shows a model of how the IS608 dimer would bind its two transposon ends. Each ssDNA end is folded into a distinct secondary structure similar to the types of complicated secondary structures formed by RNA, featuring two layers of base triplets and the subterminal hairpin. Upon cleavage at each end, the 5′-ends become covalently attached to the protein, one at each active site (Figure 1B). It is proposed that exchange of the 5′ ends between subunits through a dramatic conformational change involving the two α-helices bearing the active site tyrosine residues, followed by attack of the two free 3′ ends on the active site phosphotyrosines, leads to the generation of an excised single-stranded circle in which the two transposon ends are directly abutted (32). This is the substrate for subsequent insertion into a new location.

The IS200 TnpA active sites are composite in the sense that, for cleavage, the HUH motif is provided by one monomer while the nucleophilic tyrosine is provided by the other. For strand transfer, the arrangement changes as the tyrosines covalently attached to DNA strands through a 5′-phosphotyrosine linkage move from one monomer to the other; strand transfer is therefore catalyzed by active sites that are now composed of residues from the same polypeptide chain.

Other ssDNA transposons have subterminal palindromic sequences capable of forming hairpins, suggesting that some aspects of end recognition may be conserved. For example, IS91 has dissimilar hairpins at its two transposon ends, and they have been proposed to have distinct roles during rolling circle transposition (162). The IS91 transposase is a Y2 transposase member of the HUH superfamily, and is likely a functional homolog of the plasmid ΦX174 rolling circle replication protein, gpA (163, 167). For the eukaryotic helitrons, alignments of reconstructed consensus sequences show that while a palindromic sequence followed by a highly conserved tetranucleotide at the transposon 3′ end is always present, the only common feature at 5′ transposon ends is a conserved dinucleotide (44). Thus, understanding how these transposases are able to recognize and cleave their two ends awaits experimental work to establish their mechanism.

(ii) Integration

The IS200 DNA transposases integrate immediately 3′ to specific tetra- or pentanucleotide sequences dictated by the guide sequences located subterminal to the 5′ transposon end. For example, the target site requirement for the IS608 transposase is 5′-TTAC. IS91 also has a specific tetranucleotide target sequence, either 5′-CTTG or 5′-GTTC. Eukaryotic helitrons preferentially integrate between A and T nucleotides (44, 46), suggesting that a slightly different mechanism for target site recognition may be at work. None of these transposases generates TSDs upon integration.
For IS200 transposases, an excised ssDNA transposon circle is inserted into a new target site through a mechanism that requires that the target site possess the same sequence as the original Left End cleavage site (reviewed in He et al., this volume); thus, the DNA-DNA recognition step of initial cleavage is repeated but it is the single-stranded target that is directed into the active site rather than the uncleaved transposon-flank junction. It is proposed that integration occurs by a second set of strand cleavages, followed by another exchange of covalently attached 5′-ends between subunits of the transposase dimer. This reorganization of DNA segments results in an integrated transposon. One consequence of this mechanism is that the integration site-specificity can be manipulated at will by simply changing the bases comprising the guide sequence (168).

Thus, the key to strand transfer by HUH enzymes is that, upon nucleophilic attack on a phosphotyrosine intermediate, the covalent linkage to the active site tyrosine is resolved and the enzyme resumes its initial unbound state. It is this ability to cycle between covalent attachment, strand movement, and resolution that makes HUH enzymes particularly adept for repetitive processes such as generating multiple plasmid copies through rolling circle replication or breaking DNA bonds and rejoining them in a different configuration to comprise transposition (2).

Helitrons and ISCRs are notable in that most appear to have captured host genes (or fragments of genes). This may be a consequence of their proposed rolling circle-like mechanism which sporadically may not always terminate at the 5′ end of the element but continue beyond. When termination finally occurs, the mobilized DNA can now include sequences that were located beyond the authentic 5′ end. This capture may be an important contributor to the evolution of species where helitrons are particularly abundant, for example in maize (169, 170), and to the spread of antibiotic resistance genes by the ISCRs (40). Gene capture is likely a variation of one-ended transposition previously described for IS91 (162).

**GENOME REPAIR AFTER TRANSPOSON INSERTION**

Once a transposon has been mobilized and re-integrated, there are two genomic sites that must be repaired: the empty donor site from which the transposon has excised,
CONCLUDING REMARKS

The powerful combination of genetic, biochemical, and structural studies has illuminated many aspects of DNA recombination mechanisms. However, vast areas of great interest remain to be investigated. For example, although the field has a solid foundation for understanding of how catalysis is likely to proceed for the simpler RNase H-like transposases, we still do not have a clear, detailed sense of how protein structure mediates flanking hairpin formation and the evident need to act sequentially on first one strand and then the other. Similarly intriguing is the question of how certain transposases mediate the formation of circular intermediates. It remains unclear how serine and tyrosine transposases have taken the protein building blocks of site-specific recombinases and repurposed them for reactions that are considerably less stringent in terms of sequence specificity. Also, although at least 20 different superfamilies of eukaryotic transposases have been defined, few have yet been shown to be amenable to biochemical studies and we know very little about what distinguishes them and how this might relate to mechanism and structure. Notably, the helitron transposases are most curious as they surely must bear some functional and structural similarities to their well-characterized prokaryotic HUH relatives. Similarly intriguing are the ISCRs, and as yet there is no direct experimental evidence regarding the mechanisms of either of these types of presumptively mobile elements. In all these areas, many important discoveries and answers to fundamental questions of structure and function await.

ACKNOWLEDGMENTS

This work was funded by the NIH Intramural Program of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). We thank Phoebe Rice for the inspiration for Table 1.

REFERENCES


and the nicks and gaps that were introduced (if any) upon target site integration. Surprisingly little is known about these repair steps for most transposons. Among prokaryotic transposition systems, the donor site is believed to be repaired by homologous recombination as long as a sister chromosome is available i.e., after DNA replication but before cell division (171). This timing also leads to the regeneration of the transposon at the donor site, thereby leading to an increase in the number of transposon copies per cell. Gap repair has recently been investigated for the non-replicative pathway of Mu transposition and shown to involve proteins of both the replication restart and homologous recombination pathways (172).

Information on how donor sites are repaired in eukaryotic systems arises predominantly from studies on the P element (Majumdar & Rio, this volume) and the V(D)J recombination system (Roth, this volume). RAG1/2-mediated cleavage and excision leaves flanking hairpins at the sites of the DSBs (173), and the cellular Artemis complex is responsible for the initial step of repair which is hairpin opening (174). The NHEJ (non-homologous end-joining) proteins such as the Ku70/80 heterodimer and DNA-PKcs are also involved (175). The Drosophila Ku70 homolog has been similarly implicated in the repair of gaps introduced by P element excision (176), and Ku70/80 is reported to be important in the proper execution of ciliate programmed genome rearrangements by the domesticated DNA transposase PiggyMac (182). It seems likely that the same proteins and repair pathways will be required for opening flanking hairpins generated by other eukaryotic transposases such as the hAT and CACTA family transposases.

MuA transposition has been particularly well-studied from the perspective of what happens to the transpososome itself once the chemical steps of transposition have been completed. The version of the MuA transpososome which remains bound to DNA once strand transfer is completed is exceedingly stable. To complete transposition, the MuA complex must be removed from the two branched junctions so that the host cell replication machinery can take over. This is accomplished by a host-encoded remodeling protein, ClpX, which is a member of the Clp/Hsp100 family of AAA+ ATPases (20, 177). ClpX recognizes a specific sequence at the C-terminus of MuA, and unfolds one particular subunit of the tetramer in an ATP-dependent reaction; this appears to destabilize the transpososome enough to allow functional replication forks to be assembled (178–180).
Hickman and Dyda


Mechanisms of DNA Transposition


102. Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM. 2004. Analysis of the transposase and reconstitution of transposition activity from pro- 


