Published in final edited form as:

Curr Opin Struct Biol. 2011 June ; 21(3): 370–378. doi:10.1016/j.sbi.2011.03.004.

Moving DNA around: DNA transposition and retroviral integration

Sherwin P. Montañó and Phoebe A. Rice*

Department of Biochemistry and Molecular Biology, The University of Chicago, 929 E. 57th St., Chicago IL 60615

Abstract

Mobile DNA elements are found in all kingdoms of life, and they employ numerous mechanisms to move within and between genomes. Here we review recent structural advances in understanding two very different families of DNA transposases and retroviral integrases: the DDE and Y1 groups. Even within the DDE family which shares a conserved catalytic domain, there is great diversity in the architecture of the synaptic complexes formed by the intact enzymes with their cognate element end DNAs. However, recurring themes arise from comparing these complexes, such as stabilization by an intertwined network of protein-DNA and protein-protein contacts, and catalysis in *trans*, where each active subunit catalyzes the chemical steps on one DNA segment but also binds specific sequences on the other.

Transposases represent the most abundant genes known [1]. The mobility they lend to the genetic elements that encode them has many consequences, such as spreading antibiotic resistance and providing an evolutionary driving force by carrying regulatory elements to new locations [2]. Mobile DNA elements such as transposons and viruses that integrate into their hosts' genomes have co-opted a variety of cellular enzymes to facilitate their own mobility. Conversely, some transposases have been domesticated by their hosts, such as the RAG1/2 complex that catalyzes V(D)J recombination and Metnase, a histone methyltransferase-transposase fusion that plays a role in human DNA repair. This review will focus on recent advances in understanding DNA transposases and retroviral integrases from 2 different families, as classified by their catalytic domains: (1) the DDE family, which is related to RNaseH and includes many DNA transposases and retroviral integrases as well as RAG1 and Metnase (2) the Y1 family, which is related to proteins involved in rolling circle replication, and includes unusual DNA transposases that require a folded ssDNA substrate. Broader views of the mechanisms exploited by mobile DNA elements can be found in [3–5].

Although the chemical mechanisms used to catalyze these reactions vary among families, events on partner DNAs are nearly always coordinated. This is generally enforced by an ordered reaction pathway, in which the chemical steps follow formation of an appropriate “synaptic” protein-DNA complex, containing the two element ends, an oligomer of the recombinase, and in some cases, accessory protein and DNA factors as well. Within the complex, each active site catalyzes 2 or more sequential phosphoryl transfer reactions. Another intriguing feature of many recombination mechanisms, including those reviewed

© 2011 Elsevier Ltd. All rights reserved.

*corresponding author: PRice@uchicago.edu; (773) 834 1723.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

here, is that the final step simply exchanges one high-energy phosphodiester bond for another, and is thus chemically isoenergetic. Such reactions can be pulled forward by product binding energy, by a favorable change in DNA topology, or by subsequent treatment of the product complex [6,7]. The past two years have seen exciting advances in our three-dimensional understanding of several transposition and related recombination systems. For the DDE family, structural information regarding synaptic complexes had been limited to crystal structures of one transposase (Tn5) and an EM reconstruction of another (Mu transposase) [8,9]. New work includes crystal structures of two more (a mariner-family transposase and a retroviral integrase), and an EM reconstruction of a third [10–13]. Comparison of these structures highlights the diversity of architectures that the conserved DDE domain can be harnessed to function within. Additionally, a new series of structures of a Y1-family transposase illustrates an entirely different way to carry out DNA transposition [14].

The DDE family

Many DNA transposases and retroviral integrases belong to a large superfamily of polynucleotide transferases [3]. These feature a catalytic triad of negatively charged amino acids, DDE or DDD, that bind divalent metal ions and are brought together by an RNaseH-like fold. The catalytic domains are often significantly elaborated with additional structures, some of which are insertions within the catalytic domains and some of which are independent N- or C-terminal domains. These are often involved in sequence-specific DNA-binding, oligomerization, and stabilization of unusual DNA structures such as hairpins.

The mechanism of phosphotransfer falls within the broad category of “two metal ion” – catalyzed reactions, in which two precisely positioned divalent metal ions can orient the reacting groups, lower the pK_a s of the nucleophile and leaving group, and stabilize the geometry and charge of the transition state [15,16]. Two steps are common to all transposition and integration reactions catalyzed by DDE-family members. The first is DNA cleavage by simple hydrolysis at the mobile element – host junction (Fig. 1, arrows 1a or 1b). The second shared step, catalyzed by the same active site, is the attack of the 3'OH at the element's end on a target DNA phosphodiester, yielding the strand transfer product (Fig 1, arrows 2a or 2b). In some cases the strand transfer step immediately follows the initial hydrolysis reaction, while in others additional hydrolysis and/or transesterification reactions occur between these two common steps. Events at the two ends of a mobile DNA element are usually coordinated, with the recombinase inactive until it is incorporated into a synaptic complex, termed a “transpososome” or “intasome,” that contains both element ends and at least a dimer of the recombinase. This coordination is enforced in part by catalysis *in trans*: the subunit that catalyzes the chemical steps on the element's left end binds specific sequences near the right end, and *vice versa*. Thus proteins that are only bound to one DNA segment and not incorporated into the proper synaptic complex are chemically inactive. This prevents biologically unproductive insertion of a single element end into a target DNA. This *trans* architecture, predicted from biochemical studies, is a recurring feature in structural studies [17–19].

DDE recombinases employ diverse mechanisms in generating the intermediates of transposition (Fig. 1). Some cleave only one DNA strand per transposon end (e.g. Mu and retroviruses in Fig 1), some cleave both strands through a hairpin intermediate (e.g. Tn5, Tn10, Rag1/2 and Hermes), and some cleave both strands without a hairpin intermediate (e.g. Mos1). Remarkably, the same pair of catalytic sites usually performs all the required phosphotransfer reactions. For many years our only structural view of a DDE recombinase synaptic complex was provided by a series of structures of the bacterial Tn5 transposase [8].

However, the new findings reviewed below have greatly expanded our perspective of these diverse enzymes.

Mariner transposition

Recently, a crystal structure of Mos1 transposase in complex with its element-end DNA was reported, providing the first such structure for the Tc1/mariner family of eukaryotic transposons [13]. The architectures of the Tn5 and Mos1 transpososomes are remarkably different (Fig. 2). The transposon ends are nearly parallel in the Mos1 complex but antiparallel in the Tn5 complex. Both use helix-turn-helix (HTH) motifs to recognize specific sequences marking their element's ends, but Mos1 transposase has two such domains connected by a long linker, whereas Tn5 transposase has a single DNA-binding domain. Although in both cases dimerization is mediated by a network of protein-protein and protein-DNA contacts, these contacts are mediated by different protein segments. The Mos1 transposase dimer involves two separate protein-protein interactions: dimerization of the N-terminal DNA-binding domain, and the interaction of the "clamp loop" (an insertion in the canonical catalytic domain) of one subunit with the interdomain linker and 2nd DNA-binding domain of the other subunit. In Tn5 transposase dimer, the only protein-protein contacts involve a small helical domain C-terminal to the catalytic one. Remarkably, these very different interaction networks both result in *trans* catalysis.

Mos1 transposase cleaves both strands of each transposon end without forming a hairpin intermediate. It cleaves the nontransferred strand first, but how it does so is not obvious. The crystal structure used pre-cleaved DNA and presumably represents the state immediately preceding the final, strand transfer step. The 5' ends of the nontransferred strands are remote from both active sites, implying that during the initial cleavage step, the conformation of the complex must be quite different. Biochemical data disagree on whether or not cleavage of the nontransferred strand requires synapsis [13,20,21]. As is often the case, this beautiful structure that has answered many questions also raises new ones.

Metnase: a domesticated mariner transposase

Metnase, also called SETMAR, is a product of molecular evolution. Present only in primates, it is a fusion of a SET gene encoding a histone methyltransferase and a mariner-family transposase, Hsmar1 [22]. This chimeric protein is under positive selection and participates in several aspects of DNA metabolism, including double-strand break repair, but its exact roles are unclear. Metnase retains the ability to catalyze DNA nicking and strand transfer, although its DDE motif has been mutated to DDN and it is defective in cleaving the transferred strand, which may prevent unwanted transposition *in vivo*. The recent crystal structure of the Metnase transposase domain reveals a new dimer interface not found in Mos1 structures despite the similarity of their catalytic domains [23]. A mutation that disrupted this dimer also disrupted DNA cleavage and integration activities, implying that it is a functional interaction. However, the clamp loop that forms different interactions between subunits in the Mos1 synaptic complex is conserved in Metnase. It is possible that different protein-protein interactions form at different stages of the reaction, but more structural and functional studies of Metnase, including its complex(es) with DNA are needed to solve these puzzles.

Retroviral integration

Structural studies of the Prototype Foamy Virus (PFV) intasome were a major breakthrough not only for retroviral research but also in the biochemistry of DNA transposition [11,12]. These structures provide the first view of any DDE recombinase with both element ends and target DNA bound, as well as the first view of a tetrameric one.

The architecture of the PFV intasome is very different from that of the Tn5 and Mos1 transpososomes (Fig. 2). However, it underscores the recurring theme of catalysis *in trans*. Despite the functional requirement for a tetramer, only two subunits participate directly in catalysis, and each of these catalytic domains interacts with the 3' end of the viral DNA that is specifically bound by the other active subunit's N-terminal domains. The C-terminal domain of each active subunit interacts with both viral DNA segments, as well as the target DNA in the strand transfer complex. The role of the additional two subunits is unclear: only their catalytic domains are visible, and they only contact the active subunits, not the DNA. It is possible that their DNA-binding domains would interact nonspecifically with longer target or viral end DNAs.

As in the Tn5 and Mos1 complexes, there are no direct protein-protein contacts between the two active catalytic domains in the PFV intasome. Instead, complex assembly is mediated by protein-protein contacts involving less conserved domains and linker regions, and by the fact that each active subunit binds both element end DNAs. In the PFV intasome, a C-terminal β barrel, not found in Tn5 or Mos1, appears to aid in orienting the two viral end DNAs by interacting with both of them, as well as the target DNA. The target DNA also interacts with a long loop between β strands 1 and 2 of the catalytic domain, which is also where Mos1's "clamp loop" is inserted. Part of the Mos1 loop is also oriented such that it might interact with target DNA.

The target DNA in the PFV intasome is strongly bent, as expected from biochemical data, including insertion into nucleosomal DNA (Fig. 3) [24]. Bending is facilitated by unstacking of the central dinucleotide, which rationalizes the enzyme's preference for weakly stacking pyrimidine-purine steps at this position. A very similar path for target DNA is seen in our new structure of the bacteriophage Mu transpososome (Montaño, Zhang and Rice, unpublished data). That strongly bent target DNA may be a recurring theme is also suggested by the crystal packing of the Mos1 complex. Two additional copies of the element end DNA are associated with each complex, but these are not bound by the specific DNA-binding domains. Instead they are found inserted into the catalytic domains in an overall geometry that resembles that of the PFV target DNA (Fig. 3b) [13].

What prevents reversal of the strand transfer step, which is simply the displacement of one 3'OH by another? Structures of the PFV intasome before and after this step show surprisingly little overall difference. However, in the product complex, the scissile phosphate has shifted 2.3Å away from the active site, thus making it an unlikely target for attack.

V(D)J recombination

To enhance genetic diversity, immunoglobulin and T cell receptor genes are assembled by V(D)J recombination, which pastes together 2 or sometimes 3 gene segments from pools of cassettes [25]. The initial steps, juxtaposition of two cassettes and DNA cleavage, are catalyzed by a pair of proteins, RAG1 and RAG2, that appear to have derived from a domesticated transposon [26]. RAG1 contains a predicted RNaseH-like fold with a DDE motif, but RAG2 is also required for proper complex formation and full activity. The binding sites for these proteins, termed recombination signal sequences, lie at the ends of the intervening genomic segment that will be deleted, and contain two different sequence motifs of 7 and 9 bp, separated by either spacers of either 12 or 23 bp. Efficient V(D)J recombination requires a synaptic complex containing one of each type of end, RAG1, RAG2, and the DNA bending protein HMG1/2. Despite extensive biochemical characterization, structural data on this system has been limited to individual domains outside the recombination core machinery [27,28]. Recent advances include an EM

reconstruction of the full synaptic complex, and a crystal structure of a RAG1 DNA-binding domain in complex with DNA bearing the 9bp sequence motif (the “nonamer”)[10,29].

EM analysis of the full synaptic complex reveals an anchor-shaped object with approximate twofold symmetry, and a stoichiometry of 2RAG1:2RAG2:1HMGB1 [10]. The symmetry was surprising because the two DNA segments’ spacers differ by a full helical turn, but bending by HMGB1 may help the longer DNA fit into the same protein scaffold as the shorter one. In combination with immunolabeling and atomic force microscopy, it was shown that RAG2 lies near the catalytic region of RAG1, and that the two DNAs exit the other end of the complex in a roughly parallel orientation.

The crystal structure of the RAG1 nonamer binding domain with DNA shows a largely helical, intertwined dimer synapsing two DNA segments [29]. Although contacts between DNA-binding domains occur in other recombinases [13,30,31], this degree of intertwining has only been seen in the DNA-binding domain of Hermes transposase, which follows a similar hairpin-formation pathway [32]. This crystal structure implies that the two DNAs are antiparallel, in apparent contradiction to the EM/AFM study. However, the DNA in the crystals actually bound in both of two possible orientations. Further investigation is needed to understand exactly how this substructure fits into the full complex.

Y1 transposases: small but clever

Despite the widespread nature and versatility of the DDE domain, several other phosphotransferase domains have also been co-opted by mobile elements for use as transposases. Y1 transposases are intriguing examples. They are encoded by the IS200/605 family of insertion sequences and are related to proteins that initiate rolling circle replication [4,33]. These single-domain, dimeric enzymes include a conserved HUH (His – hydrophobic- His) motif that binds a single divalent cation and a nucleophilic tyrosine that forms a covalent bond to the DNA in the reaction intermediates. This family includes the smallest known DNA transposases. Unlike the DDE recombinases, which usually show little to no specificity for their targets, these enzymes catalyze insertion just 3’ of a defined 4-5nt sequence.

Y1 transposases are unique in that they recognize a folded single-stranded version of their element ends, and even more so in that they choose their target sites based primarily on DNA-DNA contacts. Preference for single-stranded DNA substrates allows targeting of particular aspects of DNA metabolism, such as the lagging strand during DNA replication and the intermediates of double-strand break repair [14,34]. Previous structures of IS608 transposase illustrated how the enzyme recognizes the element ends, and how it displays them to select a target DNA entirely through DNA-DNA contacts (including what may be the first biologically relevant base triplet in DNA) [33]. These structures also showed how the element could be engineered to select different targets solely by changing the DNA sequences near the element ends, which should make it a very useful genetic engineering tool [35]. A recent series of structures of a different Y1 transposase, from the ISDra2 insertion sequence, shows that while many features are conserved, the detailed rules for target site selection are not [14]. The new work also provides the first view of a fully assembled active site. In both the IS608 and the ISDra2 complexes, both subunits contribute to each active site: a single helix bearing the nucleophilic tyrosine from one subunit docks next to the metal-binding HUH motif of the other (Fig. 4). The metal ion also binds the scissile phosphate, which is attacked by the tyrosine, displacing the 3’OH and creating a 5’ phosphotyrosine linkage. The two tyrosine-bearing helices are then proposed to swap places, juxtaposing their phosphotyrosine linkages to the opposite 3’OHs. Religation in the new configuration follows in the chemical reverse of the cleavage reaction. During the excision

of the element, initial cleavage of the DNA is thus in *trans* (the nucleophilic Y of one protomer attacks the DNA bound to the other), but after helix swapping, ligation in *cis* forms a circular single-stranded element and reseals the host DNA (See cartoons in Fig. 4). After dissociation of the resealed host DNA, the circular element – transposase complex can bind a new target sequence, and in the reverse of the excision process, integrate into it.

Perspectives

Comparison of the several structures of DDE recombinase synaptic complexes now available highlights some common features. Reflecting a need to coordinate events at the two element ends, the complexes are arranged such that catalysis much occur in *trans*, and they are held together not by direct contacts between catalytic domains but by an intertwined network involving extensive protein-DNA contacts. The overall geometry varies greatly, which probably reflects different spacings between attack sites on the target DNA (from 4bp for PFV integrase to 9 for Tn5 transposase), different regulatory details, and widely divergent evolution. This is partially analogous to the type II restriction enzyme family, where the conserved fold of the catalytic domain is found in a variety of dimers, resulting in different spacings between cut sites. However, in restriction enzymes the catalytic domain is generally responsible for oligomerization and DNA sequence specificity as well, whereas these functions are distributed among different domains for the DDE recombinases.

The progress of the past few years reviewed here, will hopefully pave the way toward answering further questions as well. For example, although some DDE recombinases function efficiently using two identical element ends as substrates, others, like the RAG1/2 complex, prefer two somewhat different ends. This can help ensure the correct choice of substrates, but molecular explanations remain elusive for how the enzymes can specifically pair one left end with one right end. Another poorly explained feature of some systems is their apparent use of seemingly extra subunits. Although PFV integrase will perform coordinated integration of a pair of short viral DNA ends such as those in the crystal structure, HIV integrase requires much longer DNA substrates to form a proper intasome [36], suggesting that the outer subunits, and maybe many other weakly interacting ones as well, might form a complex network of stabilizing protein-protein and protein-DNA interactions. Tn7 and bacteriophage Mu transposases are examples of systems that include specific binding sites for subunits not needed for the chemicals steps of transposition [37,38]. In these cases the binding sites arrays differ between the left and right element ends, and the “extra” subunits may somehow aid in pairing one left with one right end, as well as in stabilizing the full complex.

Transposons are useful genetic tools, and the possibility of designing specificity for particular target sequences would make them more so. Recent structural work provides foundations for doing so. In the PFV intasome case, careful examination of the protein – target interactions provides the necessary background for altering target preferences through mutation and selection. The Mos1 structure showed where a sequence-specific DNA-binding domain could be fused to the transposase, which has recently been demonstrated to provide target specificity [39]. Finally, structural studies of Y1 transposases have guided redesign of target specificity by simply altering the element’s DNA ends rather than changing the protein [35].

References cited

1. Aziz RK, Breitbart M, Edwards RA. Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res.* 2010; 38(13):4207–4217. [PubMed: 20215432]
2. Shapiro JA. Mobile DNA and evolution in the 21st century. *Mob DNA.* 2010; 1(1):4. [PubMed: 20226073]

- 3**. Hickman AB, Chandler M, Dyda F. Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. *Crit Rev Biochem Mol Biol.* 2010; 45(1):50–69. This work carefully integrates genomics, known structures, and structural predictions to categorize and predict the mechanisms of a plethora of transposases. Such work will become increasingly more important as the sequence database expands much faster than biochemical and structural knowledge can. [PubMed: 20067338]
4. Filee J, Siguier P, Chandler M. Insertion sequence diversity in archaea. *Microbiol Mol Biol Rev.* 2007; 71(1):121–157. [PubMed: 17347521]
5. Dyda, F.; Hickman, AB. DNA transposases. In: Rice, PA.; Correll, CC., editors. *Protein-nucleic acid interactions : structural biology.* RSC Pub; Cambridge: 2008. p. xviip. 397
6. Abdelhakim AH, Sauer RT, Baker TA. The AAA+ ClpX machine unfolds a keystone subunit to remodel the Mu transpososome. *Proc Natl Acad Sci U S A.* 2010; 107(6):2437–2442. [PubMed: 20133746]
7. Grindley ND, Whiteson KL, Rice PA. Mechanisms of site-specific recombination. *Annu Rev Biochem.* 2006; 75:567–605. [PubMed: 16756503]
8. Steiniger-White M, Rayment I, Reznikoff WS. Structure/function insights into Tn5 transposition. *Curr Opin Struct Biol.* 2004; 14(1):50–57. [PubMed: 15102449]
9. Yuan JF, Beniac DR, Chaconas G, Ottensmeyer FP. 3D reconstruction of the Mu transposase and the Type 1 transpososome: a structural framework for Mu DNA transposition. *Genes Dev.* 2005; 19(7):840–852. [PubMed: 15774720]
- 10**. Grundy GJ, Ramon-Maiques S, Dimitriadis EK, Kotova S, Biertumpfel C, Heymann JB, Steven AC, Gellert M, Yang W. Initial stages of V(D)J recombination: the organization of RAG1/2 and RSS DNA in the postcleavage complex. *Mol Cell.* 2009; 35(2):217–227. A biochemical and technical tour-de-force that provides a low-resolution map of the protein-DNA complex that initiates V(D)J recombination. [PubMed: 19647518]
- 11**. Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature.* 2010; 464(7286):232–236. The first crystal structure of any retroviral integrase in a functional complex with viral DNA ends. [PubMed: 20118915]
- 12**. Maertens GN, Hare S, Cherepanov P. The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature.* 2010; 468(7321):326–329. The first crystal structure of any DDE recombinase with target DNA bound. Structures are reported for the target-bound intasome before and after the strand transfer, showing how the reverse reaction is inhibited. [PubMed: 21068843]
- 13**. Richardson JM, Colloms SD, Finnegan DJ, Walkinshaw MD. Molecular architecture of the Mos1 paired-end complex: the structural basis of DNA transposition in a eukaryote. *Cell.* 2009; 138(6):1096–1108. The first structure of an intact Tc1/mariner family transposase in complex with transposon end DNA, poised for target capture and strand transfer. [PubMed: 19766564]
- 14**. Hickman AB, James JA, Barabas O, Pasternak C, Ton-Hoang B, Chandler M, Sommer S, Dyda F. DNA recognition and the precleavage state during single-stranded DNA transposition in D. radiodurans. *EMBO J.* 2010 A series of crystal structures of a Y1 transposase synaptic complex provides the first view of a fully assembled active site, and shows unexpected complexity in target recognition showing that the rules cannot be directly extrapolated from one Y1 transposase to the next.
15. Kim EE, Wyckoff HW. Reaction mechanism of alkaline phosphatase based on crystal structures. Two-metal ion catalysis. *J Mol Biol.* 1991; 218(2):449–464. [PubMed: 2010919]
16. Steitz TA, Steitz JA. A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A.* 1993; 90(14):6498–6502. [PubMed: 8341661]
17. Aldaz H, Schuster E, Baker TA. The interwoven architecture of the Mu transposase couples DNA synapsis to catalysis. *Cell.* 1996; 85(2):257–269. [PubMed: 8612278]
18. Savilahti H, Mizuuchi K. Mu transpositional recombination: donor DNA cleavage and strand transfer in trans by the Mu transposase. *Cell.* 1996; 85(2):271–280. [PubMed: 8612279]
19. Davies DR, Goryshin IY, Reznikoff WS, Rayment I. Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science.* 2000; 289(5476):77–85. [PubMed: 10884228]

20. Claeys Bouuaert C, Chalmers R. Transposition of the human Hsmar1 transposon: rate-limiting steps and the importance of the flanking TA dinucleotide in second strand cleavage. *Nucleic Acids Res.* 2010; 38(1):190–202. [PubMed: 19858101]
21. Dawson A, Finnegan DJ. Excision of the Drosophila mariner transposon Mos1. Comparison with bacterial transposition and V(D)J recombination. *Mol Cell.* 2003; 11(1):225–235. [PubMed: 12535535]
22. Shaheen M, Williamson E, Nickoloff J, Lee SH, Hromas R. Metnase/SETMAR: a domesticated primate transposase that enhances DNA repair, replication, and decatenation. *Genetica.* 2010; 138(5):559–566. [PubMed: 20309721]
- 23*. Goodwin KD, He H, Imasaki T, Lee SH, Georgiadis MM. Crystal structure of the human Hsmar1-derived transposase domain in the DNA repair enzyme Metnase. *Biochemistry.* 2010; 49(27):5705–5713. Structure of the catalytic domain of this domesticated transposase. [PubMed: 20521842]
24. Pruss D, Bushman FD, Wolffe AP. Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proc Natl Acad Sci U S A.* 1994; 91(13):5913–5917. [PubMed: 8016088]
25. Swanson PC, Kumar S, Raval P. Early steps of V(D)J rearrangement: insights from biochemical studies of RAG-RSS complexes. *Adv Exp Med Biol.* 2009; 650:1–15. [PubMed: 19731797]
26. Schatz DG. Antigen receptor genes and the evolution of a recombinase. *Semin Immunol.* 2004; 16(4):245–256. [PubMed: 15522623]
27. Matthews AG, Kuo AJ, Ramon-Maiques S, Han S, Champagne KS, Ivanov D, Gallardo M, Carney D, Cheung P, Ciccone DN, Walter KL, et al. RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature.* 2007; 450(7172):1106–1110. [PubMed: 18033247]
28. Bellon SF, Rodgers KK, Schatz DG, Coleman JE, Steitz TA. Crystal structure of the RAG1 dimerization domain reveals multiple zinc-binding motifs including a novel zinc binuclear cluster. *Nat Struct Biol.* 1997; 4(7):586–591. [PubMed: 9228952]
- 29*. Yin FF, Bailey S, Innis CA, Ciubotaru M, Kamtekar S, Steitz TA, Schatz DG. Structure of the RAG1 nonamer binding domain with DNA reveals a dimer that mediates DNA synapsis. *Nat Struct Mol Biol.* 2009; 16(5):499–508. The structure of a RAG1 DNA binding domain in complex with part of the RAG1/2 binding site shows a highly intertwined DNA that mediates synapsis of two DNA segments. [PubMed: 19396172]
30. van Pouderooyen G, Ketting RF, Perrakis A, Plasterk RH, Sixma TK. Crystal structure of the specific DNA-binding domain of Tc3 transposase of *C. elegans* in complex with transposon DNA. *EMBO J.* 1997; 16(19):6044–6054. [PubMed: 9312061]
31. Mouw KW, Rowland SJ, Gajjar MM, Boocock MR, Stark WM, Rice PA. Architecture of a serine recombinase-DNA regulatory complex. *Mol Cell.* 2008; 30(2):145–155. [PubMed: 18439894]
32. Hickman AB, Perez ZN, Zhou L, Musingarimi P, Ghirlando R, Hinshaw JE, Craig NL, Dyda F. Molecular architecture of a eukaryotic DNA transposase. *Nat Struct Mol Biol.* 2005; 12 (8):715–721. [PubMed: 16041385]
33. Barabas O, Ronning DR, Guynet C, Hickman AB, Ton-Hoang B, Chandler M, Dyda F. Mechanism of IS200/IS605 family DNA transposases: activation and transposon-directed target site selection. *Cell.* 2008; 132(2):208–220. [PubMed: 18243097]
34. Ton-Hoang B, Pasternak C, Siguier P, Guynet C, Hickman AB, Dyda F, Sommer S, Chandler M. Single-stranded DNA transposition is coupled to host replication. *Cell.* 2010; 142(3):398–408. [PubMed: 20691900]
- 35*. Guynet C, Achard A, Hoang BT, Barabas O, Hickman AB, Dyda F, Chandler M. Resetting the site: redirecting integration of an insertion sequence in a predictable way. *Mol Cell.* 2009; 34(5): 612–619. This work demonstrates how the target specificity for a Y1 family member can be altered by changing only the DNA substrates, not the protein. [PubMed: 19524540]
36. Li M, Craigie R. Processing of viral DNA ends channels the HIV-1 integration reaction to concerted integration. *J Biol Chem.* 2005; 280(32):29334–29339. [PubMed: 15958388]

37. Gueguen E, Rousseau P, Duval-Valentin G, Chandler M. The transpososome: control of transposition at the level of catalysis. *Trends Microbiol.* 2005; 13(11):543–549. [PubMed: 16181782]
38. Holder JW, Craig NL. Architecture of the Tn7 posttransposition complex: an elaborate nucleoprotein structure. *J Mol Biol.* 2010; 401(2):167–181. [PubMed: 20538004]
- 39*. Feng X, Bednarz AL, Colloms SD. Precise targeted integration by a chimaeric transposase zinc-finger fusion protein. *Nucleic Acids Res.* 2010; 38(4):1204–1216. Based on the Mos1-DNA complex structure, another Tc1/mariner family transposase was fused with a sequence-specific Zn finger domain to create a chimeric transposase with designer target specificity. [PubMed: 19965773]

Figure 1. Diversity of DDE recombinase-catalyzed reactions

The mobile element could be a DNA transposon, a retroviral cDNA, or the intergenic region to be deleted in V(D)J recombination. The element ends are cartooned as bold black lines; the rest of the mobile element (not to scale) as dashed lines (omitted in the central panels), and the flanking DNA as blue lines. The recombinase recognizes specific sequences at or near the ends and assembles them into a synaptic complex. The first chemical step is simple hydrolysis at or near the element – flanking DNA junction, freeing either the 3' (top; arrow 1a) or 5' (bottom; 1b) end of the element. The final chemical step, termed strand transfer, also shared by all these systems (arrows 2a and 2b), is the coordinated attack of the 3' OHs of both element ends on a target DNA (gray) (right). Depending on the element, there are ~4-9bp between joining sites on the target DNA. In some cases there are no intervening steps, and the 3'OH that was the leaving group in the hydrolysis reaction directly becomes the nucleophile in the strand transfer reaction (arrow 2a). For transposons following this pathway, the branched strand transfer product can be converted into replication forks that copy the element (new DNA shown in red). For retroviruses, the “flanking DNA” is only a few nucleotides added to the ends during reverse transcription of the viral RNA, and the resulting small flaps and gap can be repaired by host pathways. Other elements use a variety of methods to cleave both DNA strands at each end before strand transfer. Some form hairpin intermediates on either the element end (e.g. Tn5 and Tn10) or flanking DNA (e.g. Rag1/2 and Hermes) by the direct attack of the 3'OH end of one strand on the paired strand. In the former case, these hairpins must be opened by a second hydrolytic reaction to free the transferred strand's 3'OH for strand transfer. Cleavage of the second strand can also be accomplished by direct hydrolysis, either by a second enzyme (e.g. Tn7) or by the DDE recombinase itself (e.g. Mos1). Note that the Rag1/2 complex does not normally perform strand transfer, which could generate genomic translocations. Further variations on these mechanism can be found in [3].

Figure 2. Comparison of synaptic complexes

Ribbon drawings are shown of the synaptic complexes of a) Mos1 transposase, b) Tn5 transposase, and c) and d) PFV integrase (from two orthogonal viewpoints). In each complex, one active subunit is blue with its catalytic domain purple and the other yellow with its catalytic domain green. The 3'OH of the transferred strand is highlighted in red, as are any metal ions found in the active site, and the element end DNAs are gray. In the PFV intasome structure, the two additional copies of the catalytic domain are shown in gray, and the Zn ions found in the N-terminal domains are colored to match the protein. Relevant domains and features of the blue/purple subunit are labeled in each complex, with DNA-binding domain abbreviated as DBD. Note these DBDs often bind other protein domains as well as DNA. PDBids 3hot (Mos1), 1mus (Tn5) and 3os0 (PFV) [8,11–13]. This and figures 3 and 4b were made with Pymol (www.pymol.org).

Figure 3. Target DNA interactions

a) The strand transfer complex of PFV integrase is shown, with the target DNA in black, and the scissile phosphates highlighted in red (other colors are the same as in figure 2). b) The Mos1 synaptic complex is shown, with additional DNA segments that mediate crystal packing and may mimic target binding shown in black. The views in panels a and b here are rotated $\sim 90^\circ$ around a vertical axis from those in Figure 2a and 2d respectively.

Figure 4. Y1 transposases

a) Cartoon of the proposed transposition pathway for Y1 transposases (based on [14]). The host DNA is drawn in blue, the transposon in black, target site motifs as green boxes, and transposon DNA motifs that recognize the target as gray boxes. The position of the scissile phosphate is marked with a small red star. (1) The transposase binds to hairpins found at each element end, and pairs the gray recognition motifs with the green target motifs; (2) DNA cleavage occurs in *trans*, with the tyrosine from the blue subunit attacking the element-host junction bound to the yellow subunit and *vice versa* (colored circles represent phosphotyrosine intermediates); (3) the Y-bearing helices swap places, carrying covalently linked DNA with them; (4) religation occurs in *cis* to generate (5) resealed host DNA and a circular, single-stranded transposon that can re-integrate into a new copy of the target motif through the reverse of this process. b) Structure of an ISDra2 transposase dimer (blue and yellow) bound to two folded, single-stranded left ends (gray) and a short target DNA (green). The active site His and Tyr (here mutated to Phe) residues are red, and the single metal ion (here Cd⁺⁺) is yellow. PDBid 2xo6 [14].