Annu. Rev. Microbiol. 1999. 53:245–81

INTEGRATING DNA: Transposases and Retroviral Integrases

L. Haren, B. Ton-Hoang, and M. Chandler
Laboratoire de Microbiologie et Génétique Moléculaire, CNRS (UPR 9007) and Université Paul Sabatier, 31062 Toulouse, France; e-mail: mike@ibcg.biotoul.fr

Key Words transposons, insertion sequences, structure, transposition mechanism

■ **Abstract** Transposable elements appear quite disparate in their organization and in the types of genetic rearrangements they promote. In spite of this diversity, retroviruses and many transposons of both prokaryotes and eukaryotes show clear similarities in the chemical reactions involved in their transposition. This is reflected in the enzymes, integrases and transposases, that catalyze these reactions and that are essential for the mobility of the elements. In this chapter, we examine the structure-function relationships between these enzymes and the different ways in which the individual steps are assembled to produce a complete transposition cycle.

CONTENTS

0066-4227/99/1001-0245$08.00

INTRODUCTION

Transposable elements are discrete segments of DNA capable of moving from one locus to another in their host genome or between different genomes. They are distributed across the living world and play a fundamental role as motors of genome plasticity in all three biological kingdoms. They induce various types of genome rearrangement and are a major source of mutation. In the case of prokaryotic organisms, they are implicated in the acquisition of "accessory" functions such as resistance to antibacterial agents, catabolism of "unusual" substances, virulence, and in the control of expression of neighboring host genes (see 16, 134).

One of the major distinguishing features of transposable elements is whether their transposition relies exclusively on DNA intermediates or includes an RNA stage. DNA elements [transposons and insertion sequences (ISs)] can be found in both prokaryotes and eukaryotes, whereas those with RNA intermediates (retroviruses and retrotransposons) are restricted to eukaryotic organisms. RNA elements can be divided into those that carry long-terminal repeats (LTR) (retroviruses and LTR retrotransposons) and those that do not (LINE-type retrotransposons). Retroviruses are distinguished from retrotransposons by the presence of an envelope gene located at the 3′ end of a multigenic polyprotein reading frame; this gene is necessary for the extracellular phase of the viral life cycle (see 20). DNA elements range from the genetically complex (such as bacteriophage Mu, which, like the retroviruses, encodes an extensive set of proteins necessary for its viral life-style) to the genetically elementary ISs composed of two short terminal inverted repeats flanking a single, or two, short open reading frames. Many hundreds of RNA and DNA elements have been identified. For bacterial ISs alone, >17 families containing >500 members are currently known (106). Although at first sight RNA and DNA elements might appear to be fundamentally different, passage of the retroviruses and LTR retrotransposons through a double-strand DNA stage provides a DNA substrate for integration that is analogous to that of the DNA transposons and is processed in a similar manner. This review focuses on the properties and activities of the enzymes that catalyze the DNA transactions necessary for the mobility of these elements. These are called transposases (Tpases) for transposons and ISs; for the LTR RNA elements, they are called integrases (IN). Discussion will be limited to a particular group that uses a conserved triad of acidic amino acids (DDE) for catalysis. Members of this enzyme family are extremely widespread and appear to represent the majority of known Tpases and IN. However, where appropriate, information concerning other well-characterized or promising transposition systems whose Tpases do not exhibit an obvious DDE motif (such as the *Drosophila melanogaster* P element and the plant transposons Ac, En/Spm) is included.

THE TRANSPOSITION CYCLE

Transposition proceeds by endonucleolytic cleavage of the phosphodiester bonds at the ends of the transposable element and transfer of these ends into a target DNA molecule. This recombination reaction requires assembly of a highly organized nucleoprotein ternary complex (the synaptic complex or transpososome), which includes Tpase or IN, the transposon ends, and target DNA. Historically, transposition mechanisms have been separated into two principal modes, conservative and replicative, based on whether the element is copied in the course of its displacement. This is dictated by the nature and order of the cleavages at the ends (Figure 1): whether the transposon is liberated from its donor backbone by double-strand cleavages or whether it remains attached following cleavage of only a single strand.

Conservative Transposition

In conservative or cut-and-paste transposition (Figures 1*A* and 1*B*), the element is excised from the donor site and reinserted into a target site without replication. This implies cleavage of both DNA strands at the ends of the element and their rejoining to target DNA to generate a simple insertion. Examples of elements shown to have adopted this strategy include Tn*7* (34), IS*10* and IS*50* (60, 83), the Tc/*mariner* family (123), and the P element (50). Transposon excision may introduce a potentially lethal double-strand break in the donor molecule that can be rescued by double-strand break repair or by gene conversion using a sister chromosome as a template (32, 51, 62). For IS*10* and probably other ISs of the IS*4*/IS*5* group (181; see 106), passive replication of the element by the host replicon activates transposition. This ensures that duplication of the IS occurs before transposition (see 83).

Replicative Transposition

Replicative transposition (Figure 1*C*) involves cleavage of only one strand at each transposon end (the transposon is not excised from the donor molecule) and transfer into a target site in such a way as to create a replication fork. Elements that use this strategy include bacteriophage Mu (94, 109), members of the Tn*3* family of bacterial transposons (144), and the IS*6* family of ISs (see 106). If transposition is intermolecular, replication from the nascent fork(s) generates cointegrates in which donor and target replicons are joined but separated by a directly repeated copy of the element at each junction. Intramolecular transposition of this type generates inversions and deletions. Resolution of the cointegrate structures to regenerate the donor and target molecules, each carrying a single copy of the element, is accomplished by recombination between the two elements. This proceeds for some transposons by site-specific recombination promoted by a specialized transposon-specific enzyme distinct from the Tpase, the "resolvase" (e.g. Tn*3* family; 144), or

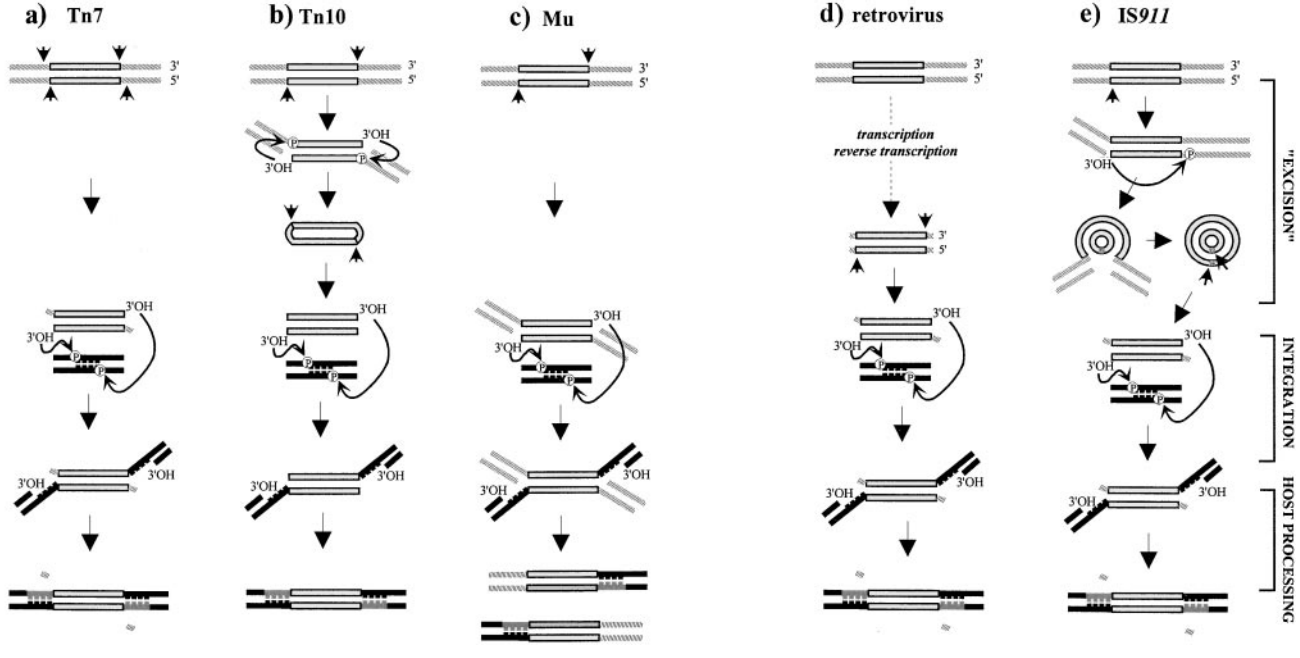


Figure 1 Types of strand cleavage and transfer. Transposon DNA is indicated by light gray boxes, donor DNA by striped boxes, and target DNA by filled boxes. Positions of strand cleavage are indicated by short vertical arrows. Nucleophilic attack of the phosphodiester bond (represented within a circle) by the active 3′ hydroxyl resulting in strand transfer is also indicated by short arrows. The "toothed" region shown in the target DNA represents the small target duplications associated with insertion. The newly synthesized DNA is indicated by dark gray. DNA polarity is shown at the top of each panel. (*A*) elements that undergo double-strand cleavage, e.g. Tn*7*; (*B*) elements that undergo double-strand cleavage by way of a hairpin intermediate, e.g. IS*10*; (*C*) elements that pass through a cointegrate intermediate, e.g. phage Mu (note that the newly replicated transposon strand is indicated in dark gray); (*D*) amplification by transcription and reverse transcription before insertion, e.g. retroviruses; (*E*) amplification before insertion by way of a circular DNA intermediate, e.g. IS*911*. This figure was inspired by Reference (78).

is taken in charge by the host homologous recombination system (e.g. IS*6* family; see 106).

Alternative Routes

Transcription The LTR elements (Figure 1*D*) undergo amplification by transcription from an integrated DNA copy into RNA copies. A free double-stranded cDNA copy, designated for integration, is then synthesized by reverse transcription (20). This generally carries 2-base-pair (-bp) extensions at each end, compared with the integrated copy. Single-strand cleavage occurs at both ends to remove two terminal bases, and the recessed ends are then transferred into a target site to give simple insertions. Thus, in spite of the fact that cleavages are limited to a single strand at each end, integration is a conservative event. This is made possible because the element is separated from the donor backbone by transcription before initiation of the transposition reactions.

Circle Formation Another alternative transposition strategy to circumvent the requirement for double-strand end cleavages has recently been described (Figure 1*E*). In the case of IS*911* IS*2* and IS*3*, Tpase promotes single-strand cleavage at one end of the transposon and its site-specific transfer to the same strand of the opposite end. This circularizes a single transposon strand, leaving the complementary strand attached to the donor backbone. This second transposon strand is then resolved to generate a circular transposon copy in which the transposon ends are abutted. The resolution mechanism is at present unclear but could involve simple cleavage and repair or replication promoted by host proteins. The covalently attached ends can then undergo simultaneous single-strand cleavage and transfer to a suitable target (154 and references therein). Elements that use this type of strategy involving site-specific strand transfer may include other members of the IS*3* family and probably those of the IS*30* and IS*21* families (see 106). Although site-specific strand transfer from one end of the element to the other generates transposon circles, it also may occur between two elements carried by the same molecule. Transfer of ends between the two IS copies in a plasmid dimer, for example, would be expected to generate head-to-tail IS dimers. This type of structure has been observed in the case of IS*21* (127), IS*2* (150), IS*30* (120), and IS*911* (C. Turlan B.T.H. and M.C., unpublished) and is extremely active in transposition.

It is clear from the pathways described above that replicative transposition per se, or more precisely replicative integration, is limited to cases such as Mu and the Tn*3* family of transposons, in which replication of the element is intimately associated with transfer of the cleaved ends into the target. In the other pathways, the preintegration intermediate is separated from the donor backbone; as a consequence, the integration event itself is conservative. The major difference between these pathways is the way in which separation of the preintegration intermediate occurs. This may simply involve double-strand cleavage at each transposon end or

entail more elaborate steps such as reverse transcription or circularisation of the element, as shown in Figure 1.

MECHANISM

The development of defined in vitro systems has been essential in understanding the steps involved in transposition. Such systems are now available for several retroviruses [human immunodeficiency virus (HIV), avian sarcoma virus (ASV), and murine leukemia virus], the P element, members of the Tc/*mariner* group of eukaryotic ISs, the prokaryotic elements Mu, Tn*7*, and bacterial ISs of the IS*4* (IS*10*, IS*50*) and IS*3* (IS*911*) families (13, 20, 34, 60, 83, 109, 110, 126, 154). In all these cases, the cleavage and strand-joining reactions necessary for transposition are remarkably similar. They consist of a series of consecutive Tpase- or IN-catalyzed hydrolysis and transesterification reactions that require no external energy source (Figures 1 and 2). The overall reaction can be divided into three stages: detachment of the transposon from its donor site by single- or double-strand cleavage at the ends, transfer of the cleaved transposon ends to a target site, and processing of the products by host-encoded enzymes.

First-Strand Donor Cleavage

This important first chemical step initiates separation of transposon and donor DNA. For elements such as bacteriophage Mu, the retroviruses, and members of the IS*3* family (Figures 1*C*, 1*D*, and 1*E*, respectively), the Tpases catalyze single-strand cleavage at the ends liberating a 3′OH (see 109). For the retroviruses, the equivalent reaction is called processing and results in the liberation of two bases from the 3′ end of the double-stranded cDNA, leaving a two-base 5′ overhang. This is a hydrolysis reaction with H_2O as an attacking nucleophile and it generates a free 3′OH (Figure 2*A*) (see 109, 111).

Second-Strand Donor Cleavage

For elements in which transposition occurs by excision from the donor molecule, cleavage of the complementary strand must also take place. Whereas cleavage at the 3′ end appears to be common to all transposable elements analyzed to date, this is not true for second-strand cleavage. For IS*10*, this occurs opposite the cleaved 3′ strand to generate flush transposon ends (Figure 1*B*); for Tn*7*, second-strand cleavage takes place three nucleotides to the 5′ side, resulting in a three-base 5′ overhang (Figure 1*A*). For elements of the Tc/*mariner* family and for the P element, second-strand cleavage occurs several nucleotides within the transposon (2 for Tc*1/3* [162]; 3 for the *mariner* element, Himar*1* [93]; and 17 for P [13]), generating a molecule with a 3′ overhang.

In the case of IS*10*, the two breaks are not analogous: 5′ cleavage occurs subsequent to 3′ cleavage (17), and the free 3′OH generated by 3′ cleavage is itself used

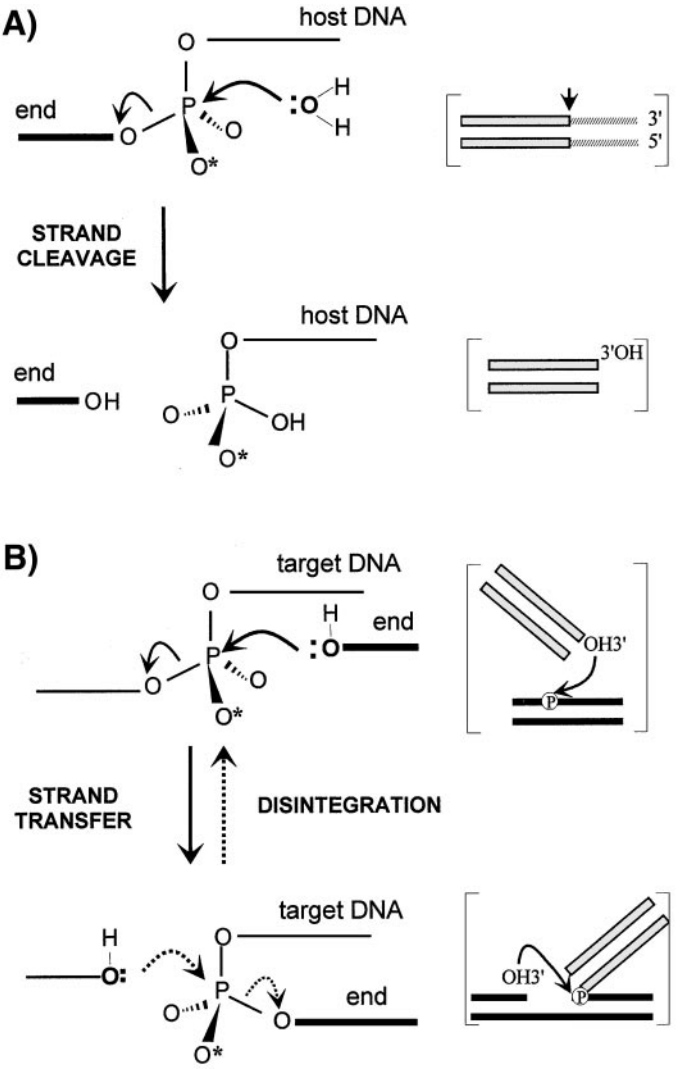


Figure 2 The chemistry of strand cleavage and transfer: (*A*) strand cleavage, (*B*) strand transfer and disintegration. Only one of the relevant DNA strands is shown in each case. The transposon DNA is represented as a bold line. Nucleophilic attack either by H_2O (Figure 2*A*) or a 3′ hydroxyl from the exposed transposon end (end) for strand transfer or from the exposed 3′ hydroxyl in the target DNA for disintegration (Figure 2*B*) is shown from the right and indicated by curved arrows. The phosphodiester bond that undergoes cleavage in these reactions is shown as a chiral form. Although the chirality of the phosphate is not normally fixed, introduction of a sulphur atom to replace a nonbridging oxygen (O^*) fixes one or the other chiral forms. The hydrolysis (Figure 2*A*) and transesterification (Figure 2*B*) reactions shown here result in an inversion of chirality. The corresponding steps involving double-strand donor and target DNA together with their polarity are presented on the right of this figure for clarity.

as the nucleophile in attacking the second strand (78). This generates a hairpin structure at the transposon ends that is subsequently hydrolyzed to regenerate the 3′OH and 5′ phosphate ends (Figure 1*B*). This mechanism is reminiscent of V(D)J recombination, which generates the immunoglobin repertoire, although the V(D)J hairpin is generated on the equivalent of the donor backbone ends (159). Interestingly, the V(D)J recombination system can promote inter- and intramolecular transposition and may indeed have derived from an ancestral transposon (2, 69). A similar mechanism has been proposed for the excision/transposition of a variety of other eukaryotic transposons, such as Ac, Tam*3*, hobo, and Ascot-*1* (8, 32, 33).

Few details are available concerning second-strand cleavage of the other elements. Transposon Tn*7* (Figure 1*A*) uses a different strategy. Its Tpase is composed of two distinct polypeptides, TnsA and TnsB, each dedicated to cleavage of a particular strand (137). For members of the Tc family, early studies with purified Tpase were unable to detect 3′ strand cleavage but revealed specific 5′ cleavage (167), showing that cleavage of the two strands can be uncoupled. As described above, for IS*911* and probably a relatively large group of bacterial ISs, the necessity for a 5′ cleavage step is circumvented by the use of asymmetric strand cleavage and transfer. In the resulting circular intermediate, both ends are abutted and only two single-strand 3′ cleavages are required to liberate them for strand transfer into the target (Figure 1*E*) (154).

Strand Transfer

Once cleaved, the 3′ ends of the transposon are transferred into the target DNA molecule. Similar to strand cleavage, strand transfer proceeds by a Tpase- or IN-catalyzed nucleophilic attack. Here the attacking nucleophile is the 3′OH group at the free transposon end (Figure 2*B*) (see 109, 111). Like strand cleavage, the reaction does not require an external energy source, suggesting that the energy in the target phosphodiester bond is used directly in the formation of the new transposon-target joint. Experiments with Mu and HIV-1, in which chirality was imposed on the phosphate of the scissile target phosphodiester bond by substitution of a nonbridging oxygen for a sulfur atom to generate a phosphorothioate group (49, 112), imply that this reaction does not involve the formation of a covalent protein-DNA intermediate. The target phosphate underwent stereochemical inversion in the course of the reaction, implying a direct single-step in-line nucleophilic attack (Figure 2). A two-step mechanism using a protein-DNA intermediate would have regenerated the original stereochemical configuration as a consequence of the second step, as shown for recombination reactions promoted by phage λ Int (112). For HIV-1, both cleavage and strand transfer reactions behaved in the same way (49). It is the direct attack by the 3′ transposon end(s) that couples transposon to target DNA without prior target cleavage. This creates a new phophodiester bond between transposon and target and leaves a 3′OH group in the target DNA at the

point of insertion. Concerted insertion of each 3′ transposon end into the target generally occurs in a staggered way (Figure 1). Although the transfer reaction, like the initial cleavage reaction, is isoenergetic, several in vitro transposition systems show a requirement for high-energy cofactors [ATP for Mu and Tn*7* (9, 107); GTP for P (77)]. These appear to have a regulatory role, however, and do not intervene in catalysis. Furthermore, the fact that precleaved substrates can undergo relatively efficient integration (9, 14, 35, 76, 98, 135, 152) shows that the strand transfer reaction is not energetically coupled to donor strand cleavage.

IN and Tpases also catalyze a reaction called disintegration that can be considered as a reversal of integration (14, 30, 126, 167). Joining of the 3′ transposon end to the target generates a branched molecule and uncovers a 3′OH on the target DNA. In the disintegraton reaction, it is this 3′OH that attacks the newly formed transposon target junction to liberate the inserted end (Figure 2*B*). Although this reaction has been useful in the analysis of IN and Tpase function, it is unclear whether it is biologically relevant.

Post-Transfer Processing

The staggered insertion of the transposon ends into the target site results in the creation of short, complementary single-strand regions flanking the inserted element. Those regions in which the transposon has been separated from the donor DNA (Figures 1*A*, 1*B*, 1*D*, and 1*E*) are thought to be repaired by the host repair/replication machinery to generate the characteristic short direct-target repeats, a hallmark of many types of insertion. Where the transposon remains connected to the donor backbone by its 5′ ends (Figure 1*C*), a potential replication fork is created at each end in which the flanking 3′OH of the vector can act as a primer. Replication across the element would then result in a cointegrate molecule.

Although the transposable elements considered here appear quite diverse in their transposition cycles, their Tpases or IN catalyze similar types of reaction: hydrolysis for strand cleavage and transesterification for strand transfer. The distinctiveness of these elements is not derived from the chemistry of the reactions but rather in the way these reactions are coordinated during transposition. The fact that these enzymes carry out similar chemical reactions implies a degree of functional and structural similarity. This is examined below.

ENZYMES

Tpases and IN are multifunctional enzymes that must accomplish a series of tasks. To ensure the reactions described above, the enzymes must be able to locate the ends of the element, bring them together into a synaptic complex, recognize the phosphodiester bonds to be cleaved, ensure the incorporation of a target DNA molecule at an appropriate stage in the pathway, direct strand transfer, and relinquish their

place to host enzymes recruited for post-transfer processing. These functions involve specific and nonspecific DNA binding, catalysis, the capacity to multimerize and, in some cases, the capacity to interact with accessory proteins.

Overall Organization

Initial dissection of IN and of various Tpases has often used partial proteolysis in vitro and deletion analysis in vitro and in vivo to define topologically independent functional domains. The Mu Tpase (MuA) has been divided in this way into three major domains: I, II, and III (see 94). These have been further sectioned into subdomains I(α,β,γ), II(α,β), and III(α,β), each with specific activities. Similar types of study have been undertaken with IN, in which partial proteolysis yields N-terminal and C-terminal fragments together with a central core region (see 5) and with the Tpases of IS*10* (91) and IS*50* (19) (Figure 3).

Although it is convenient to analyze Tpase and IN functions in terms of domain structure, it should be kept in mind that individual functions are not necessarily accomplished by a single isolated domain but may be assumed by several different regions of a single polypeptide or by more than one polypeptide in a multimeric complex. However, a general pattern for the functional organization of Tpases is emerging: sequence-specific DNA-binding activities are generally located in the N-terminal region, whereas the catalytic domain is often localized toward the C-terminal end (Figure 3). This arrangement has been observed for IS*1* (104, 183), IS*30* (146), Mu (see 94), Tn*3* (105), IS*50* (175), IS*903* (151), IS*911* (66, 126), Tc*1/3* (see 123), P (14), and Ac (15).

DNA Recognition

DNA recognition operates at several levels. It intervenes in discriminating the ends of the element from nontransposon DNA, in assembling the transpososome, in fitting the transposon substrate into the catalytic pocket, and in sequestering target DNA.

Transposon Ends A key feature of Tpases is their capacity to specifically recognize and bind the transposon ends. In discussing how these enzymes may accomplish this, it is important to consider the organization of the ends themselves. With several notable exceptions (see 106), the majority of ISs exhibit short terminal inverted repeat sequences (IR) of between 10 and 40 bp, whereas other types of transposon may carry an array of such sequences. Many transposon ends are composed of two functional domains. One, located internally, ensures correct sequence-specific positioning of Tpase on the ends. This corresponds to a short nucleotide stretch within the single terminal repeats of ISs (or Tn*3* family elements) or the terminal Tpase-binding sites of certain of the more complex elements, such as Mu and Tn*7*. The other domain corresponds generally to the 2–4 terminal base pairs necessary for cleavage and strand transfer. This extreme terminal domain is identical at both ends of the element and tends to be conserved between related

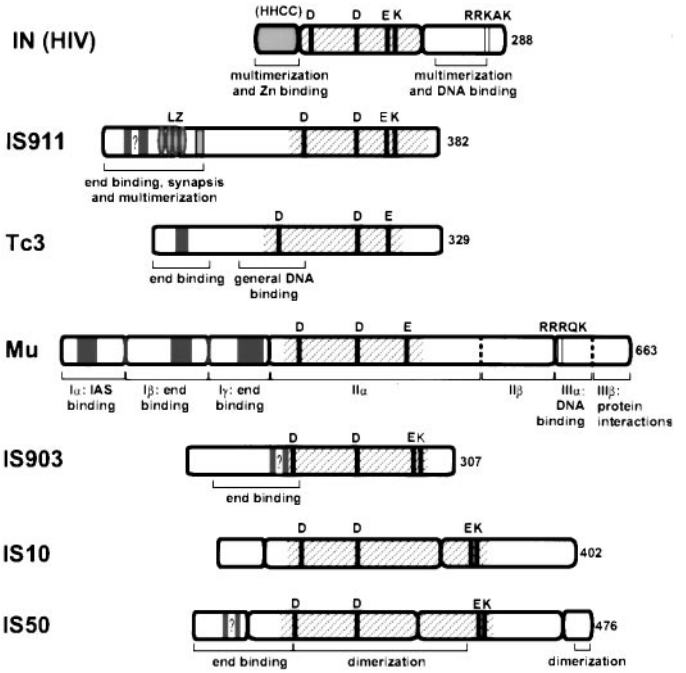


Figure 3 Transposase (Tpase) organization. The positions of protease sensitive sites are delimited by the open boxes. Potential or real helix-turn-helix (HTH) motifs are shown as dark grey boxes. Potential HTH motifs are indicated by "?." The catalytic core is indicated by pale grey and carries the DDE motif. These residues, together with a number of others referred to in the text, are indicated in uppercase letters above each Tpase molecule. LZ indicates the leucine zipper motif with the four repeating heptads observed in the IS*911* Tpase. A second region involved in multimerization is also shown slightly downstream. In those cases investigated, the catalytic core is also capable of promoting multimerization. Known functions of the different Tpase regions are indicated below. Tpase alignments are centered on the second aspartate residue. The length of each protein in amino acids is indicated at the right. The function of each region, where known, is indicated under the respective proteins.

transposons. One of the most common is the terminal CA-3′ dinucleotide found at the ends of Mu, Tn*7*, IS*30*, Tn*552*, and the IS*3* family elements, as well as the retroviral genomes (see 125). Binding sites for host proteins are also sometimes found within or close to the ends (see 106). Such proteins play a role in modulating Tpase expression or in transposition activity by influencing transpososome assembly and progression through the cycle (see "Assembling the Pieces" below).

Simple single terminal Tpase-binding sites observed for most IS elements and members of the Tn*3* family of transposons are to be contrasted with multiple and asymmetric protein-binding sites carried by Mu (36), Tn*7* (34) and Tn*552* (133),

or the eukaryote transposons Ac (89), En/Spm (57), and P (14). These are often arranged in a different way in the left and right ends and can provide a functional distinction between the ends either in the assembly or in the activity of the synaptic complex. The subterminal sites presumably ensure the correct synapsis of two ends or positioning of Tpase on the terminal site.

Tpase End Binding The bifunctional organization of the terminal Tpase-binding sites may be reflected in the overall organization of the cognate Tpases, where the N-terminal sequence-specific DNA recognition domain is frequently topologically independent of the catalytic domain. The arrangement may allow for flexibility in the protein, enabling simultaneous contact of the N-terminal region with the internal domain and the catalytic domain with the external end domain. A common helix-turn-helix (HTH) motif is often found at the heart of the sequence-specific DNA-binding domain both in prokaryotic and eukaryotic elements (Figure 3). This is potentially able to provide binding specificity. The domain may be simple, as appears to be the case for Tpases of IS elements (see 106), or bipartite and able to recognize different DNA sequences, as found in the Mu, Ac, and Tc Tpases (15, 27, 123). In MuA, the end-binding domain corresponds to the adjacent subdomains Iβ and Iγ. Mutations in this region eliminate or reduce Mu end binding (81). Both subdomains contain an HTH motif, and each recognizes one half of the binding site (31, 140). Note that domain Iα is also a DNA-binding domain that recognizes another transposon site, the enhancer (or internal activation sequence; Figure 3) involved in transpososome assembly (see "Assembling the Pieces" below). For the Tc*1/3* Tpases, an N-terminal subdomain recognizes an internal binding site, whereas a second subdomain contacts nucleotides proximal to the terminal nucleotides. The domain resembles the paired domain found in transcription factors and, like them, also carries HTH and HTH-like motifs (122, 163, 167). For Ac, binding to the terminal sites also requires two subdomains, (one in the N-terminal and the other in the C-terminal regions of the protein) whereas binding to the internal sites requires only the second, C-terminal, subdomain (15).

Direct DNA binding of the N-terminal Tpase domain to the ends of its cognate transposon has also been observed for IS*1* (183), IS*30* (146), IS*903* (151), IS*50* (19), IS*911* (66), Tn*3* (105), the P element (96), and Ac (15). Although the protein domain responsible has not been determined, Tpase binding has also been demonstrated for Tn*1000* (174), Tn*552* (133), and one of the two Tn*7* enzymes (TnsB) but not the other (TnsA) (6).

Nonspecific Tpase Binding Sequence "independent" DNA binding is thought to be necessary for target sequestration and to contribute to positioning of the transposon ends in the catalytic pocket. The catalytic core domain of several Tpases is involved in interaction with the terminal nucleotides of the ends (see "Catalytic Domain" below) and/or target DNA. The isolated core domain of MuA and the Tpases of IS*10* and Tc*1/3* show nonspecific DNA-binding properties (91, 117, 168; see 123). Like IN (see below), the isolated core domain of the IS*911* Tpase is able

to carry out efficient disintegration, but not strand transfer, indicating that it must be capable of efficiently recognizing the branched disintegration substrate (Figure 2; 126). This type of discrimination is also likely to be true for Tc*1* (167) and P (14).

In addition, other domains of these proteins can contribute to sequence-independent DNA binding. For example, domain IIIα of the MuA Tpase (Figure 3) exhibits a nonspecific DNA-binding activity associated with a string of basic amino acids, RRRQK (177). These types of motif can be found in many Tpases, although their role has not been determined. Domain IIIα collaborates with the core domain (II) for catalysis and might stabilize interactions between the catalytic domain and its substrate (see "Assembling the Pieces" below).

Integrase-DNA Interaction IN-DNA interactions have received considerable attention. In contrast to the Tpases, IN shows no conspicuous specific DNA-binding activity, as judged by standard gel retardation and footprinting procedures (92, 157), although cross-linking and competition experiments indicate that it is able to form stable complexes with viral DNA in vitro (44, 164). Moreover, in spite of the fact that the N-terminal domain carries an HTH motif, this appears to be involved in protein multimerization (26). The absence of pronounced binding specificity in the case of IN may reflect the difference in the biology of the LTR-RNA elements and DNA transposons. The DNA substrates of IN are contained within a large nucleoprotein particle (preintegration complex) in close contact with the enzyme (172; see 20), whereas Tpases must recognize and bind the ends of their cognate DNA elements while these are embedded in the host DNA. An important component in recognition of viral DNA ends may simply be the fact that they are the physical ends of the molecule.

Close DNA contacts have been mapped in all three proteolytically defined IN domains. Photo-cross-linking experiments have shown that the N-terminal domain is in close proximity to both viral DNA and target DNA 5′ to the site of integration (67). As for Tpases, the catalytic core domain also participates in viral and target DNA binding (48, 52, 68, 71). The use of chimeric IN proteins derived from HIV and feline immunodeficiency virus demonstrated that this domain determines target site selection (145). The observation that the isolated core domain, IN_{HIV} [50–212], is capable of catalyzing disintegration but not the forward processing or strand-transfer reactions suggests that it might have enhanced affinity for the branched substrate (25). Photo-cross-linking has also revealed an interaction with the terminal nucleotides of the viral LTRs (52, 67). This specificity is provided by several residues observed to be exposed in the crystal structure (see "Catalytic Domain" below). The C-terminal domain is also involved in viral DNA binding (48, 80, 103) and recognizes DNA features proximal to the terminal nucleotides (52, 67, 68). This domain is organized in an SH3-like fold, which produces a large cleft proposed to accept DNA (42, 100). Analysis of IN_{HIV} by mutagenesis revealed that a stretch of basic residues, R262, R263, and K264 (RRKAK, similar to the RRRQK stretch in MuA), is important for DNA binding (103). The C- and N-terminal domains appear to be involved in stabilization of IN interactions with

the LTRs by allowing, for example, tighter binding of the core domain to the terminal sequences (44, 52, 68, 164).

Catalytic Domain

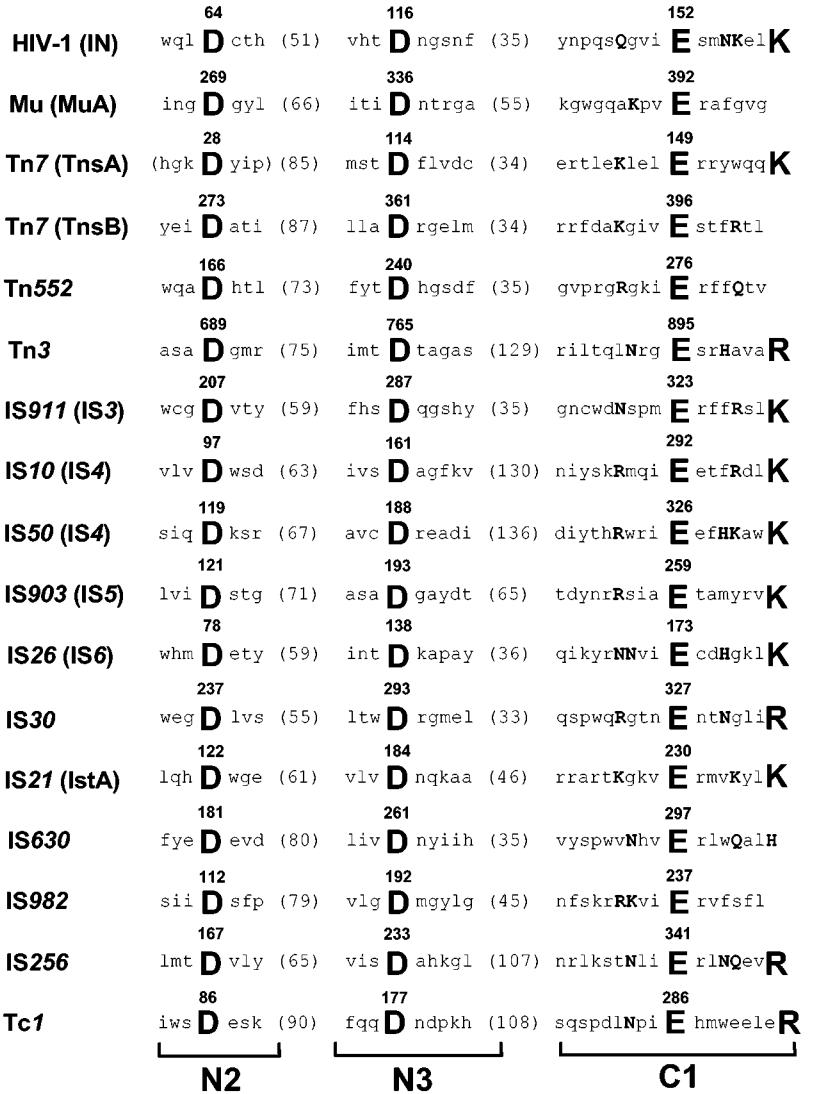
Some of the earliest comparisons between retroviral IN proteins and Tpases revealed a highly conserved triad of acidic amino acids with a characteristic spacing known as the DD(35)E motif and several additional conserved residues, in particular, a K or R residue seven amino acids downstream (54, 71, 80, 87, 125, 132). Subsequent studies generalized this signature to many other bacterial ISs and revealed three relatively well conserved regions (called N2, N3, and C1) centered on the D, D, and E residues, respectively (129) (Figures 3 and 4) . The triad was also detected in bacteriophage Mu [MuA (10)] and Tn*7* [TnsA and B (137)] Tpases, in eukaryotic ISs of the Tc/*mariner* group (38), and in Tpases of the Tn*3* family (182). A recent compilation includes a majority of known ISs (106) (Figure 4). Although both the spacing and conservation is somewhat variable from group to group and not all groups have been analyzed in sufficient detail to confirm the biological significance of these residues, the conservation is nevertheless remarkable. One essential role of this triad is to coordinate the divalent cations necessary for catalysis (see "Catalysis").

The importance of the DDE residues has been demonstrated by site-directed mutagenesis of enzymes of several members of the family. These include different IN proteins and the Tpases of bacteriophage Mu (10), Tn*7* (137), IS*10* (74), Tc*1/3* (167), and IS*911* (65).

Integrases In the case of IN, mutation of the conserved amino acids was found to abolish processing, joining, and disintegration, indicating that all three reactions are catalyzed by a single active site (40, 47, 87, 158), although residual disintegration activity was observed in IN_{HIV} mutants in which the second D (D116)

Figure 4 DDE motifs in transposases (Tpases) and integrases (INs). In addition to those enzymes described in detail in the text, the figure includes representative Tpases from various insertion sequence (IS) families. The source or relevant IS family is indicated to the left. Where the enzymes shown are not those of the founding member of the family, this is included between brackets where appropriate. The DDE triad is shown in large bold letters together with the associated downstream basic lysine or arginine. The number of residues between the first and second D residue and between the second D and the E residue is shown between brackets. The regions N2, N3, and C1 are those defined by (129). Note the presence of additional relatively well conserved residues. These include a W or other hydrophobic amino acid three residues upstream from the first D, a G two residues downstream from the second D, a basic residue four residues upstream, and four residues downstream from the final E. Uppercase bold letters indicate conserved residues shown in the case of IN to cross-link to nucleotides in the viral ends.

and E (E152) residues were exchanged (47). In addition, the observation that the core fragment IN_{HIV}[50–212], which carries the DD(35)E motif, is capable of catalyzing disintegration but not the processing and joining reactions indicated that it contained the active site of the enzyme (25, 88, 165). Mutation of other amino acids in the vicinity of the DD(35)E residues can change the specificity both for the type of divalent cation and for the nucleophile used in the processing reaction (47, 156). These results therefore provide strong evidence that the DD(35)E triad lies at the heart of the catalytic site.



Other studies have suggested that D116 of IN_{HIV} also contributes to stable binding of DNA substrates (39, 164), whereas E152 may provide the specificity for the terminal viral A/T base pair (Figure 5*A*) (56). Furthermore, cross-linking experiments and directed mutagenesis have suggested that the neighboring and highly conserved K159 and the less well conserved K156 are involved in positioning the terminal viral–CA3′ bases (Figure 5*A*) (71), whereas another nearby residue, Q148 (and, less strongly Y143), may be involved in interactions with the 5′ nucleotide at the processed viral end (52, 56). Finally, Q62 was found to cross-link to the penultimate 5′ C nucleotide (Figure 5*A*) (52). Mutation at these and other neighboring residues also increases the salt sensitivity of the enzyme, suggesting that they may disrupt interactions with DNA (56).

Many of these results can now be understood from the known structure of the IN_{HIV} and IN_{ASV} catalytic domains, and mutagenic studies were indeed often directed by a structure-based approach. Although the crystal structure of intact IN has yet to be determined, that of the core domains (IN_{HIV}[50–212] and IN_{ASV} [52–207]) has been resolved (see Figure 5*B*). For IN_{HIV}[50–212], the mutation F185K was introduced to increase solubility. Initial studies (41) showed that the domain crystallizes as a dimer in which each monomer consists of a five-stranded β-sheet surrounded by six α-helices with the two Asp residues, D64 and D116, located close together on β1 and β4, respectively. The crystal structure of an F185H mutant subsequently revealed that E152 lies close to the Asp residues in a region (residues 141–153) that was disordered in the initial study (21). More recently, two additional mutant forms of IN_{HIV-1} were used to generate crystals in which one of the α-helices (α4) was observed to be extended to include E152. This structure showed that the side chains of all three members of the triad face toward each other (Figure 5*B*) (58). The crystal structure of wild-type IN_{ASV} [52–207] showed similar overall topology to that of HIV (23, 24) with some differences (including an ordered active site). The flexible nature of the IN_{HIV} region containing E152 and Q148 and other amino acids that have been shown to influence the choice of nucleophile for the processing reaction (160) suggests that it plays a role in positioning both the nucleophile and viral DNA. In the IN_{HIV} structure, the DDE triad is clustered at the surface close to the important K156 and K159 residues and other residues shown to undergo cross-linking with viral DNA ends (Figure 5*B*).

Tpases Although mutational studies of Tpases have not been as extensive as for IN, they have provided complementary evidence concerning the importance of the DD(35)E motif in catalysis. For MuA, directed mutagenesis based on primary sequence alignments demonstrated that D269 (equivalent to the first of the two Asp residues) and E392 were essential for donor cleavage and strand transfer (Figure 4) (10, 82). The second Asp residue, D336, which could not be predicted from sequence alignments or localized by mutagenesis, was identified from the crystal structure of the catalytic domain (IIα and β; 131). Remarkably, one of the two subdomains (IIα; Figure 3) can be superimposed on the IN_{HIV} structure with the two active-site Asp residues located in identical positions. The third member of the

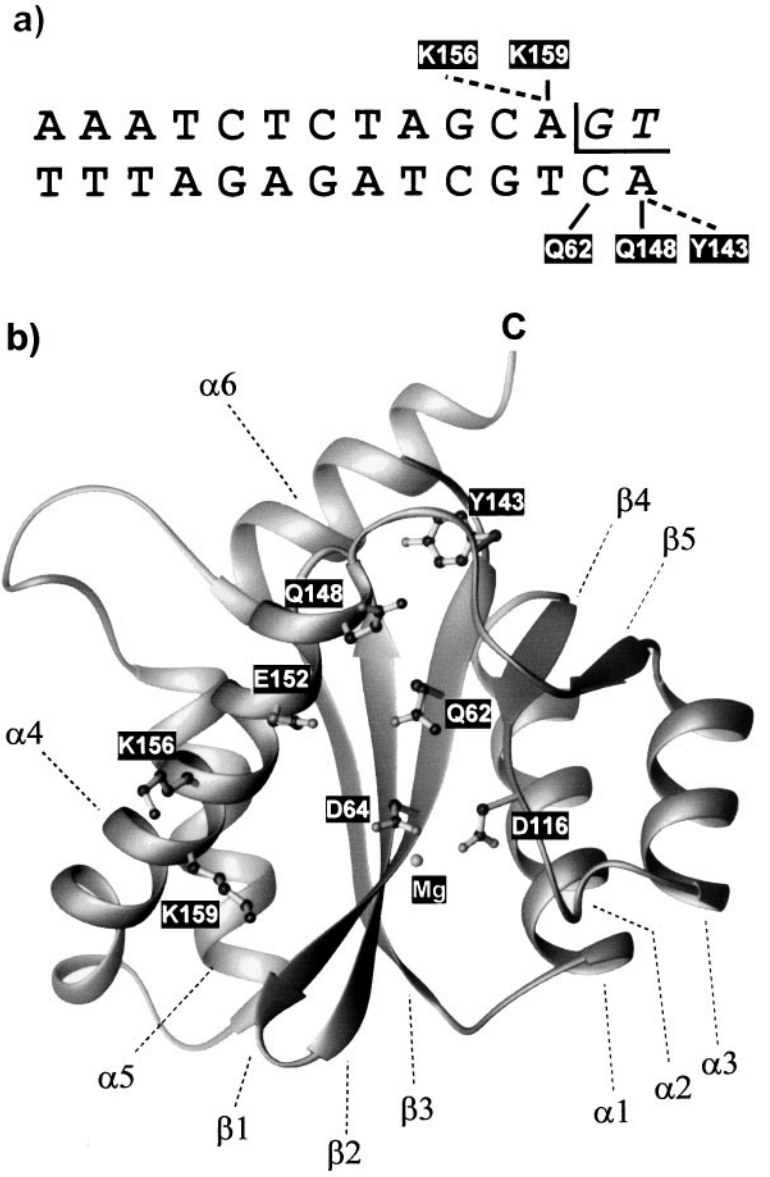


Figure 5 (*A*) The HIV end. The sequence of HIV-1 ends showing the dinucleotide, which is removed during processing (in italics). The amino acids within the catalytic core that have been shown to cross-link to specific bases are indicated. (*B*) The structure of the catalytic domain of IN_{HIV}. The structure shown is taken from (58). Relevant amino acids are indicated as is a single Mg^{2+} ion. We thank F Dyda for providing us with this figure.

triad, E392, was found to extend away from the active site and was proposed to require activation by a conformational change induced by protein-protein or protein-DNA interactions, consistent with the tightly regulated Mu transposition cycle (see "Assembling the Pieces") (131). The second subdomain (IIβ) forms a six-stranded β barrel, and, because of a large area of positive electrostatic potential located on its surface, it has been proposed to be involved in nonspecific DNA binding.

The crystal structure of the Inh protein of IS*50*, a regulatory derivative of the Tpase lacking the first 55 amino acids (128) has recently been determined (37a). Here, too, the DDE triad (D119, D188, E326) forms a distinct catalytic pocket with a similar fold to that found in IN and MuA. In this case, the presence of R322, a residue that is conserved in several Tpases (Figure 4), appears to occupy the pocket. This raises the possibility that the protein may undergo a conformational change to displace R322 before catalytic activation. In the case of the related IS*10* Tpase, the equivalent residue (R288) is essential for catalytic activity (18).

Although the crystal structure of the IS*10* Tpase is not available, sequence alignments coupled with directed mutagenesis of 18 conserved positions identified D97, D161, and E292 as the carboxylate triad. A variety of elegant screening procedures also distinguished several additional pertinent residues. Mutation of this DDE triad (Figure 4) gave an absolute defect in all catalytic reactions, whereas mutation of the highly conserved K299 downstream of E292 was found to compromise target capture and strand transfer (18). Mutants of K299 (K299A) together with those of the partially conserved R102 (R102H) and P167 (P167S) retained their ability to induce the host SOS response, presumably as a result of their wild-type capacity to catalyze strand cleavage, but they were defective for transposition (79). Three other residues, A162, M289, and E263, were also shown to influence catalysis. Mutations A162V(T) and M289I (Figure 4) and E263K generated a hypernicking phenotype (83).

The situation for Tn*7* is more complex than for other transposons because two proteins, TnsA and TnsB, are involved in the double-strand cleavages at each end and are both physically required for all cleavage and strand-transfer reactions (see "Mechanism" above). TnsB binds to the transposon ends, whereas TnsA is directed to the ends by interaction with TnsB. TnsB exhibits a well-defined DD(35)E motif (D237, D361, and E396), whereas in TnsA, the first Asp residue is less obvious (D114, E149) (Figure 4). Mutation of D114 or E149 of TnsA reduced or abolished cleavage at the 5′ end while leaving 3′ cleavage and 3′ strand transfer unaffected. This resulted in conversion of the normal cut-and-paste Tn*7* pathway into a cointegrate mode (108). Similarly, mutations of triad members in TnsB resulted in a block in 3′ cleavage (and strand transfer), leaving 5′ end cleavage unaffected (137).

The limited information available for other systems serves to generalize the importance of the DDE triad. As a member of the IS*5* family, IS*903* also carries the DDE core domain. Site-directed mutants at these residues (D121, D193, and E259) were found to exhibit severely reduced transposition in vivo (151), as were mutants of the partially conserved Y196, Y252, and R255 (Figure 4). No mechanistic analysis has yet been possible. For the Tc*3* element, mutants in the potential DDE

motif (D144, D231, and E266) were also severely compromised for catalysis (162).

One important observation that stems from the structural studies of the catalytic core is that all exhibit similar topologies not only within the Tpase/IN group but also with the catalytic domains of other enzymes that promote phosphoryltransfer reactions, notably RNaseH and the RuvC resolvase. This has led to the notion that such enzymes belong to a superfamily of proteins known as polynucleotidyltransferases (61, 130).

Comparison of the ensemble of structural and mutagenic data is beginning to provide a general picture of the detailed functions of the catalytic core and its interaction with the transposon ends, especially in the C1 region (Figure 4). In IN_{HIV}, this region forms α4 (Figure 5*B*). K159 or K156, which cross-link to the terminal CA3′ dinucleotide, are located on the same side of the α-helix and are strongly conserved in those Tpases shown in Figure 4. Q148 also lies on the same side and appears to interact with the 5′ A on the nonprocessed strand (Figure 5*A*). An amide or basic amino acid is highly conserved at this or the neighboring position (Figure 4), and mutation results in severe impairment of catalysis [IN (56); IS*10* (18); IS*50* (WS Reznikoff, personal communication); IS*903* (151)]. Additional convergences in the structure-function relationships of the catalytic domains and their role in target sequestration and positioning will certainly be forthcoming.

Multimerization

Multimerization is another fundamental property of many Tpases of both prokaryotic and eukaryotic origin (see 5, 19, 27, 66, 90, 101, 173). It is important in the assembly of the ternary synaptic complex, in ensuring collaboration between protomers for catalysis, and in regulating Tpase activity (see “Assembling the Pieces” below).

The self-association properties of Tpases are complex and still poorly understood. This is presumably because the proteins undergo a series of conformational changes during the course of the transposition reaction, and these involve changes in protein-protein interactions. Multimerization has been demonstrated physically in the case of IN (see 5) and for the Mu (95, 110), IS*50* (19, 173), and IS*911* (66) Tpases. It is also implied from complementation experiments in vivo and in vitro (see “Assembling the pieces” below) and from experiments using the yeast two-hybrid system (55, 75, 89, 101). In the case of bacteriophage Mu, MuA is a monomer in solution and oligomerizes only in the presence of its cognate DNA-binding sites (95, 114), but the nature of the interactions and the regions of the protein involved have not been probed. This type of information is available for a very limited number of enzymes.

For IN (Figure 3), which has been shown to form both dimers and tetramers, all three isolated domains (the N-terminal, core, and C-terminal) carry determinants for self-association and dimerize independently. It has been argued (26)

that although the C-terminal and core dimers are compatible (i.e. can both bridge the same pair of full-length protein molecules), the N-terminal dimers are not and must therefore bridge two dimers to generate a tetramer. Dimerization of the N-terminal domain involves an HTH motif similar to that used for DNA recognition by DNA-binding proteins. It is stabilized by coordination of Zn^{2+} with a Cys_2-His_2 motif also located in this domain (26, 43), and Zn^{2+} has been shown to enhance tetramerization of the entire protein (184). Moreover, IN derivatives deleted for the N-terminal domain form both dimers and tetramers, suggesting that a C-terminal domain dimer may bridge a core dimer (4, 23, 41, 42, 70, 100). Although the type of active multimer of IN is still unclear, circumstantial evidence indicates that the ability to form tetramers is important for activity. Zn^{2+} not only stimulates tetramerization but also processing and strand transfer (97, 184), and mutation L241A in the C-terminal domain inhibits tetramerization, processing, and strand-transfer activities (102).

For the IS*911* Tpase, OrfAB, three distinct self-interaction regions have been identified (Figure 3). In addition to the catalytic core (homologous to IN and also able to dimerize), a leucine zipper motif also promotes multimerization. This permits binding to the IRs, whereas an adjacent region appears involved in multimerization leading to synapsis (65, 66). This coiled-coil motif is common to a majority of the members of this IS family (66).

In the IS*50* Tpase, Tnp, partial proteolysis uncovered two distinct regions capable of interacting with full-length Tnp (19). One of these regions, 114–314, overlaps the catalytic domain. The other, 441–476, lies at the extreme C-terminus, and mutations within this region (Figure 3) have significant effects on transposition activity.

ROLE OF DIVALENT METAL IONS

Catalysis

Based on a reaction mechanism proposed for several transesterification reactions (see 180), it was suggested that the acidic DDE triad plays a central role in the catalytic activity of Tpases and IN by coordinating the divalent metal ions essential for activity (80). In this two-metal ion model, one metal ion is proposed to function as a general base. It increases the partial negative charge of the incoming nucleophile by deprotonation. The second is proposed to act as a general acid that assists the leaving 3′ oxyanion and stabilizes a pentacovalent transition state by coordination of the nonbridging oxygen atoms at the cleavage site (Figure 6).

Several lines of evidence support the idea that the DDE triad is responsible for divalent cation coordination. The structure of crystals of IN_{ASV}[52–207] includes a single bound Mg^{2+} or Mn^{2+} ion (24) complexed by the two Asp residues, D64 and D121 (equivalent to IN_{HIV} D64 and D116), and four molecules of water with octahedral coordination. Recent studies with new crystal forms of the IN_{HIV} core have shown that a single Mg^{2+} ion can also be coordinated by D64 and D116

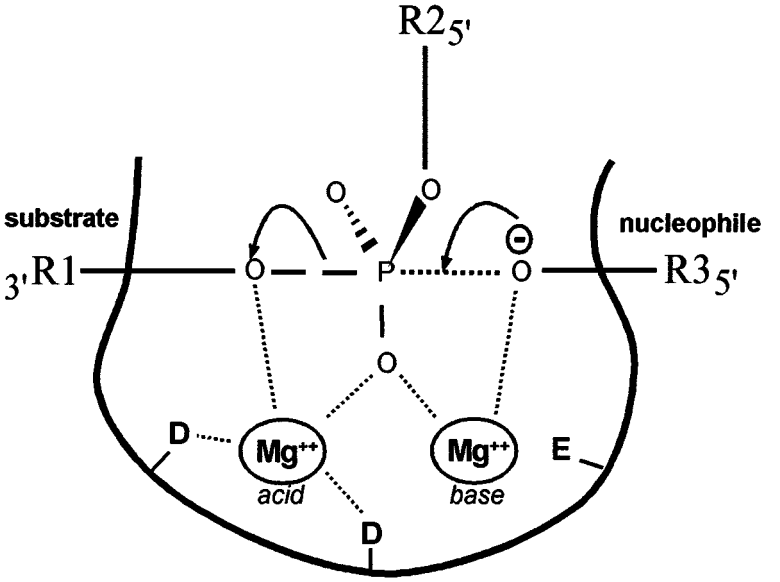


Figure 6 Role of divalent metal ions in catalysis. This figure represents the catalytic pocket containing the pentacoordinated phosphate in the (hypothetical) transition state. One Mg^{2+} ions proposed to act as a Lewis acid is shown to be coordinated with two aspartate residues, the oxyanion of the leaving group (R1 and a nonbridging oxygen). The second Mg^{2+} is also shown to complex with the nonbridging oxygen and to act as a base for deprotonization of the incoming nucleophile. The substrate is 3′R1-R25′, and the product is 3′R3-R25′.

(Figure 5*B*), together with two water molecules in a similar configuration (58). Other metal ions such as Ca^{2+}, Zn^{2+}, and Cd^{2+} also bind the active site of the ASV core (22). The involvement of the DDE triad in coordinating metal ions has also received support from studies on Tn*7*. These have exploited the fact that, compared to Mg^{2+}, Mn^{2+} shows a preference for coordination with sulfur-substituted components. Substitution of an oxygen ligand by sulfur can change the metal-ion specificity of reactions. This approach was first used in demonstrating interaction between metal ion and substrate in RNA self-splicing (37, 121), and in RNaseH activity (155). A D114C mutation in TnsA, which abolished 5′ strand cleavage in the presence of Mg^{2+}, was found to be rescued if the reaction was carried out in Mn^{2+} (137).

Although the biologically relevant cation is generally thought to be Mg^{2+}, other cations can be used in the in vitro reactions. Indeed, in the case of IN from several sources, Mn^{2+} was initially found to be the preferred cation (143, 158). However, Mn^{2+} is known to reduce the specificity of some reactions. Under optimized assay conditions, Mg^{2+} also supports a robust activity (52). In other systems, Mg^{2+} or Mn^{2+} are essential for cleavage and strand transfer and cannot be substituted by Ca^{2+} [IS*10* (73, 135), Tn*7* (137), IS*911* (126)]. Although Zn^{2+} can support the

IN_{ASV} processing reaction, which uses an external nucleophile, it is inefficient for strand transfer, which uses the viral DNA 3′OH end (22). In addition to Mg^{2+} and Mn^{2+}, MuA can also use Zn^{2+} or Co^{2+} for donor cleavage (170). Moreover, although Ca^{2+} does not support cleavage, it is able to promote strand transfer in this case (139). That certain cations can support cleavage but not strand transfer whereas others support strand transfer but not cleavage in these systems is consistent with a two-metal ion mechanism for catalysis.

Structural Role

Besides their fundamental role as cofactors in the catalytic activities of Tpases and IN, divalent metal ions also appear to intervene in the assembly of the transpososome. For retroviruses, Mg^{2+}, Mn^{2+}, or Ca^{2+} is necessary for the formation of stable interactions between IN_{HIV} and the viral LTRs (44, 164). Mn^{2+}-induced conformational changes correlated with an enhancement of catalytic activity have been revealed by a variety of methods. These changes affect both the self-assembly of IN (45, 176) and the organization of its structural domains within the multimer, as judged by the use of monoclonal antibodies directed against each of the three domains and by changes in proteolysis patterns of the full-length protein (7).

For Mu, divalent cations are also required for assembly of an active Tpase tetramer on the ends of the element (see "Assembling the Pieces" below; 11, 114). Ca^{2+} stabilizes the Mu transpososome but does not promote first-strand cleavage. Conversely, although Zn^{2+} and Co^{2+} permit catalysis of first-strand cleavage, they are inefficient in assembly of the synaptic complex (170). Furthermore, a mutant derivative of MuA has been obtained that specifically uses Ca^{2+} for assembly of the synaptic complex and Mn^{2+} for first-strand cleavage (82). This mutation, G348D, lies within the catalytic core. These results reflect a differential action of divalent metal ions in transpososome assembly and in catalysis. For IS*10*, Ca^{2+}, which does not support catalysis, stabilizes both the synaptic complex (135) and interaction with the target (136). As for IN, the stabilizing effect of these cations may result from changes in protein conformation.

ASSEMBLING THE PIECES

We have so far discussed the different steps in the transposition process and described the different functions of IN and the various Tpases associated with these transposons. In the following text, we attempt to show how these different elements are integrated into the overall transposition process. This is understood in some detail from studies on Mu and IS*10*, in which transposition is initiated by the formation of a highly organized nucleoprotein synaptic complex (83, 114) and proceeds through a series of ordered steps showing consecutive increases in complex stability to temperature and protein-denaturing agents (135, 148). Several other systems are being actively investigated, and although certain steps

may be well defined, the overall transposition pathway is generally less well understood.

End Synapsis

Tpase-mediated synapsis has been formally demonstrated for Mu (11, 139) and for IS*10* (135). In both cases, synapsis also involves accessory factors and implies a two-step mechanism leading to activation of the Tpase multimer. In a first step, a transitory constrained complex is formed. This then undergoes conformational changes during which the accessory factors are ejected and the active sites of the enzyme monomers are presumably brought closer to the reactive phosphodiester bonds (28, 110, 171).

Bacteriophage Mu has three Tpase-binding sites at each end arranged in different configurations. In an initial stage, a multimeric MuA complex is assembled on these sites using domain I$\beta\gamma$ (Figure 3). Three of the six synapsed sites (including the two terminal sites) together with a MuA tetramer compose the transpososome core (see 94). MuA bound to the remaining three sites can be removed without compromising activity. Assembly is facilitated by transitory interaction with the enhancer (internal activation sequence) by way of domain Iα of MuA (Figure 3). The enhancer DNA sequence, located >1 kilobase from the left end of the phage genome, binds MuA and presents it to the nascent complex (left end–enhancer–right end complex). Formation of this complex (LER) requires divalent metal ions and two small host proteins, IHF and HU, in addition to a supercoiled donor DNA molecule carrying both ends in their natural configuration. IHF acts indirectly at the level of the enhancer by influencing protein binding at this site, whereas HU binds at the Mu left end and directly facilitates transpososome formation on the Mu ends. Assembly can be arrested in vitro at a stage called the stable synaptic complex (SSC or type 0 complex) by use of Ca^{2+} instead of Mg^{2+} in the reaction (110). The transition from LER to SSC is blocked either by mutation of the terminal base pair or by mutation in domain IIIα of MuA (see "DNA Recognition" above). This implies a conformational change involving domain IIIα in which the Mu ends are engaged for catalysis. In the SSC, the two Mu ends are held together by a tetramer of MuA, no strand cleavage has yet occurred, and the Mu-host junction appears unwound or otherwise distorted (95, 139, 170).

For IS*10*, the short terminal IRs, which are not strictly equivalent, are defined as outside and inside ends (OE and IE) with respect to their relationship within the parental transposon Tn*10*. They are distinguished by different dispositions of *dam* methylation sites and the presence of an IHF-binding site proximal to OE. A precleavage synaptic complex has been identified using short linear OE carrying fragments including the IHF site (135). Formation of this complex (paired ends complex), composed of two ends bridged by Tpase, is dependent on IHF but not on divalent metal ions. It is thought that this is a transitory complex because titration of IHF from this complex changes its conformation and renders it competent for subsequent target capture and strand transfer (83). It is worthwhile noting that in

addition to IS*10*, several elements also carry IHF-binding sites within or proximal to their terminal IRs. In the case of Tn*1000*, IHF binding to an IR proximal site has been shown to enhance Tpase binding (174).

The types of nucleoprotein complexes formed with other elements during synapsis are not well characterized. However, for retroviruses, analysis of the preintegration complex of murine leukemia virus has implicated both IN and host proteins in a higher-order structure that also includes an extensive region of the retroviral LTRs (172). For HIV-1, the host HMG I(Y) protein has been identified in such structures and appears to be required for integration (53). These host factors may play a similar architectural role to IHF and HU in IS*10* and Mu transposition.

Target Capture and the Catalytic Steps

The strand cleavage and transfer steps must be orchestrated in a precise manner. For Mu, progression through these steps is well characterized. Following the formation of the SSC, strand cleavage occurs at both ends in the presence of the appropriate metal ion to generate a cleaved donor complex (CDC). This may be facilitated by local unwinding of the DNA at the Mu termini, which occurs in the SSC (170), probably driven by the supercoiling in the donor DNA molecule (169). In addition to the terminal binding sites, MuA was shown to protect ~10 bp of neighboring host DNA (113). This complex gives rise to a strand transfer complex (STC) in the presence of target DNA. At this stage, the 3′ transposon ends have been inserted into the target DNA, and MuA protection is extended by 20 bp into the target (113). The MuA IIβ domain is implicated in the transition between the cleaved donor complex and strand transfer complex because mutation of exposed lysine residues (K506, 515, 529, and 530) to alanine in this domain (see "Catalytic Domain" above) blocks this step (84).

Capture of a target molecule can occur at any step in the assembly process (116) and is stimulated by a second Mu protein, MuB. MuB binds target DNA in an ATP-dependent manner and delivers it to the transposome by interaction with domain IIIβ of MuA (Figure 3). ATP hydrolysis is not required for this step. However, interaction with MuA stimulates hydrolysis, which releases MuB from the target DNA (1, 178). These properties give rise to a phenomenon known as target immunity, which prevents insertion into a target molecule already carrying a Mu copy. Here, binding of MuA to Mu ends in the target stimulates the MuB ATPase activity and purges MuB from the immune target.

Strand cleavage and transfer by MuA is also directly stimulated by MuB (12, 107, 149). The in vitro reaction results in concerted integration of both ends. Recent complementation studies using a supercoiled substrate and a mixture of wild-type MuA and a catalytically inactive mutant have suggested not only that the same catalytic core is used for both cleavage and strand transfer but that both reactions are carried out by the MuA monomer located on the terminal binding site of the opposite end (118). In other words, cleavage and strand transfer occurs in *trans* (3, 138).

Several additional observations reinforce the notion of an intimately interwound Mu transpososome architecture. Complementation data suggest that a competent active site is assembled from region IIα of one MuA monomer and region IIIα of a second (179). In addition to its nonspecific DNA binding, domain IIIα exhibits weak endonuclease activity, which suggests that, in binding, it may render the appropriate phophodiester bond more susceptible to core-catalyzed cleavage (177). It has also been established that domain IIβ is not part of the same complementation group as IIα (119) but like IIIα, is supplied by a different Tpase subunit (84). Moreover, the nucleoprotein complexes become increasingly resistant to denaturation as they progress through the cycle. Finally, a mutation of the terminal nucleotides at one end abolishes cleavage of both ends (149). The interwound transpososome organization thus imposes concerted reactions at both extremities, an important constraint if the transposon is to avoid nonproductive events involving a single end.

After strand transfer, MuA is liberated from the complex and exchanged for the replication machinery necessary to complete formation of the cointegrate transposition product. This process involves several host proteins, namely the chaperone ClpX and components of the replication apparatus (PriA, PriB, DnaT, DnaBC) and occurs in a highly ordered way (72, 85, 99).

Transposition of IS*10*, like that of Mu, proceeds through a series of well-defined steps, with a progressive increase in stability. In the presence of Mg^{2+} (or Mn^{2+} but not Ca^{2+}), the precleavage synaptic complex (PEC) assembled in vitro undergoes double-strand cleavage to generate complexes with single- (SEB) or double- (DEB) end breaks. Kinetic data are consistent with the notion that the DEB is generated from the SEB. Cleavage of each strand at the ends occurs in a sequential manner, with cleavage of the transferred strand taking place before that of the complementary strand (17). This allows the liberated 3′OH of the first strand to cleave the second strand (see "Mechanism" above). DEB complexes are competent for strand transfer which occurs in a coordinated way (135). The DEB complexes are equivalent to a characteristic early intermediate in IS*10* transposition, the excised transposon fragment or ETF, observed both in vivo and in vitro, in which the entire element is excised from the donor backbone and in which the two transposon ends are held together by a Tpase bridge (29, 63). Cleavage requires both ends but does not necessarily occur simultaneously at each end (64). Moreover, a single catalytic site appears to be repeatedly used for the consecutive hydrolysis and transesterification reactions. In contrast to MuA, a Tpase dimer may catalyze the entire transposition reaction (both 5′ and 3′ cleavages and strand transfer), each acting on a distinct end without conspicuous domain sharing between monomers (18). This suggests that the transpososome architecture in the case of IS*10* is less constrained than that of Mu. This notion is supported by the observation that, in contrast to Mu, a mutation of the terminal nucleotides at one end does not significantly affect cleavage of the wild-type end (64). For IS*10*, the absence of a mechanism to ensure concerted cleavage at both ends, as in the case of Mu, is compensated by the late entry of the target into the synaptic complex subsequent

to IR cleavage (64). Stable synaptic complexes (SSCs) including target DNA have been detected only with the DEB complex, indicating that prior cleavage at both ends is a requirement for target capture (136).

Although neither the stoichiometric composition nor the temporal evolution of the retroviral synaptic complex is known in detail, some information is available concerning its activity and structure. IN activity requires a physical DNA end. Extending the ends of the viral DNA substrate severely reduces the processing reaction, whereas mismatches in the terminal region can increase activity (141, 166). These observations suggest that unwinding or DNA distortion assists IN activity. Other data clearly indicate that the complex is multimeric and that monomers collaborate with each other within the assembled structure. This is implied from complementation assays of mutant retroviral IN derivatives. Although the C-terminal domain seems to be able to act together with an active site on the same (in *cis*) or another IN monomer in the complex (in *trans*), the N-terminal domain is required to act in *trans*, and part of the C-terminal domain is required in *cis* (45, 46, 161). Moreover, a combination of cross-linking and complementation experiments has suggested that the active site of an IN monomer may act in *trans* on substrates bound to a partner monomer in the complex, further substantiating this view. Finally, like Mu, mutation at one end abolishes cleavage of both ends, reinforcing the idea of close collaboration between monomers (86, 115).

Modeling the known structures of the separated domains with suitable DNA molecules has led to the proposal that the active complex of IN may have an octameric architecture, with a tetramer acting on each LTR of the viral DNA (67). In this model, two of the four core domains assembled on each end are directly involved in catalysis and recognize the reactive terminal CA3′. The remaining two intervene indirectly by interacting with target DNA (68).

For other systems, knowledge of the steps involved in progression through the transposition cycle is sketchy. For Tn*7*, the nature of the synaptic and catalytic complexes has yet to be determined. However, like Mu, target capture for Tn*7* is relatively elaborate. Tn*7* transposition occurs in two alternative conservative modes: site-specific integration at a single chromosomal site, *att*Tn*7*, and random insertion. The site-specific integration pathway has been extensively studied and involves four Tn*7*-encoded proteins (34). In addition to the strand-specific Tpases, TnsA and B, these include TnsD, which binds directly to the target DNA site, and TnsC, which is recruited to the target site by TnsD. Interestingly, *Escherichia coli* ribosomal protein L29 and acyl carrier protein (ACP) have also been implicated in Tn*7* transposition both in vivo and in vitro, and L29 stimulates binding of TnsD to its target attTn*7* site (142). TnsC appears to assume a similar function to that of the MuB protein in both target capture and immunity, and, like MuB, it exhibits Tpase- (TnsB) dependent ATPase activity (147). However, unlike bacteriophage Mu, transposition in vitro into attTn*7* not only necessitates the presence of the target DNA before donor strand cleavage but also requires the complete set of transposition proteins. This ensures tight control in the commitment of the transpososome to successful transposition (34).

IS*911* transposition probably involves two sets of cleavage and strand transfer reactions: one that generates the single-strand transfer product and the other that leads to integration of the transposon circles. This implies the formation of two distinct kinds of synaptic complex. Current evidence from both in vivo and in vitro studies indicates that the Tpase, OrfAB, alone is sufficient for the first set of reactions that are essential for subsequent circle formation (124, 126, 153). The first synaptic complex that comes into play presumably involves only OrfAB and must show an asymmetric functional organization because only one IS end undergoes cleavage, whereas the other functions as a target (see "Transposition Cycle" above). After evolution of this complex for accomplishing strand cleavage and transfer, it must presumably be disassembled to allow second-strand processing and to generate the transposon circle. The single-strand transfer product forms a branched structure resembling that formed in the Mu strand transfer product, and this step, like the transition of the Mu (STC) to the replication complex, may occur in a highly ordered way.

A second type of synaptic complex must then be assembled on the circle junction. For the second set of reactions, a second IS*911*-encoded protein, OrfA, is required (153, 154). The IS*911* Tpase, OrfAB, like that of other members of the IS*3* family, is a fusion protein produced by translational frame shifting between two consecutive reading frames (see 106). OrfA is the product of the first reading frame and therefore carries the N-terminal domain of OrfAB, including the leucine zipper motif (Figure 3). This motif is required for the formation of OrfAB and OrfA homomultimers and for generating heteromultimers between the two proteins essential for circle integration (65). Although circle integration requires both proteins, OrfAB alone is sufficient for cleavage of both ends in the circle junction, suggesting that OrfA is involved in target capture.

CONCLUSIONS AND PERSPECTIVES

Over the past several years, it has become clear that many transposable elements have adopted an identical chemistry for their displacement. This is reflected in similarities in the primary sequence of their Tpases and INs. The accumulating data concerning the structure of the catalytic domains of these enzymes demonstrate an even higher degree of similarity in their topological organization. As yet, the number of known structures is quite limited, especially in view of the range and diversity of these elements. It will, therefore, be of considerable interest to determine and compare the structures of Tpases from the major groups of transposons.

Of the structures so far determined, not a single example of a full-length protein has been obtained. In the most complete case, Inh of IS*50* (37a), 55 N-terminal amino acids that determine sequence-specific binding to the transposon ends are not present. For IN and MuA, the structures of all the individual domains are available, although the manner in which these are connected is a matter of conjecture. A major challenge will therefore be to determine how the different domains of these

enzymes are articulated. Moreover, because transposition invariably necessitates the assembly of higher-order nucleoprotein complexes, it will be of importance to ascertain the way in which the individual components are integrated into such complexes and to define both protein-protein and protein-DNA interactions involved.

Transposition is a dynamic process. Lessons from phage Mu and Tn*10* have shown that transposition proceeds through a series of well-defined steps. Not only do these need to be defined for the other important transposition systems, but the structural details of these transitions need to be documented, as is currently being undertaken in the case of certain site-specific recombinases (59). In regard to the chemistry of the reactions, although a two-metal ion mechanism is currently favored, this has yet to receive direct experimental proof. In particular, knowledge of the transition states would obviously provide invaluable information.

Finally, although not addressed in this review and often overlooked in a purely in vitro approach, the way in which transposition activity is controlled both by the element itself and by the host represents an important aspect of transposon biology.

ACKNOWLEDGMENTS

We thank the members of the Mobile Genetic Elements Group (R Alazard, C Turlan, C Normand, G Duval-Valentin, and P Rousseau), A Diaz, A Lopez and J-F Tocanne for discussions and AJ Carpousis, A Diaz, O Fayet, and C Normand for reading the manuscript. Fred Dyda kindly supplied Figure 5*B*. This work was supported by grants from the Centre National pour la Recherche Scientifique (CNRS, France), Région Midi-Pyrenées, and l'Association pour la Recherche sur le Cancer (ARC).

LITERATURE CITED

1. Adzuma K, Mizuuchi K. 1991. Steady-state kinetic analysis of ATP hydrolysis by the B protein of bacteriophage mu. Involvement of protein oligomerization in the ATPase cycle. *J. Biol. Chem.* 266:6159–67
2. Agrawal A, Eastman QM, Schatz DG. 1998. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature.* 394:744–51
3. Aldaz H, Schuster E, Baker TA. 1996. The interwoven architecture of the Mu transposase couples DNA synapsis to catalysis. *Cell* 85:257–69
4. Andrake MD, Skalka AM. 1995. Multimerization determinants reside in both the catalytic core and C terminus of avian sarcoma virus integrase. *J. Biol. Chem.* 270:29299–306
5. Andrake MD, Skalka AM. 1996. Retroviral integrase, putting the pieces together. *J. Biol. Chem.* 271:19633–36
6. Arciszewska LK, McKown RL, Craig NL. 1991. Purification of TnsB, a transposition protein that binds to the ends of Tn7. *J. Biol. Chem.* 266:21736–44
7. Asante-Appiah E, Skalka AM. 1997. A metal-induced conformational change and

activation of HIV-1 integrase. *J. Biol. Chem.* 272:16196–205
8. Atkinson PW, Warren WD, O'Brochta DA. 1993. The hobo transposable element of Drosophila can be cross-mobilized in houseflies and excises like the Ac element of maize. *Proc. Natl. Acad. Sci. USA* 90:9693–97
9. Bainton R, Gamas P, Craig NL. 1991. Tn7 transposition in vitro proceeds through an excised transposon intermediate generated by staggered breaks in DNA. *Cell* 65:805–16
10. Baker TA, Luo L. 1994. Identification of residues in the Mu transposase essential for catalysis. *Proc. Natl. Acad. Sci. USA* 91:6654–58
11. Baker TA, Mizuuchi K. 1992. DNA-promoted assembly of the active tetramer of the Mu transposase. *Genes Dev.* 6:2221–32
12. Baker TA, Mizuuchi M, Mizuuchi K. 1991. MuB protein allosterically activates strand transfer by the transposase of phage Mu. *Cell* 65:1003–13
13. Beall EL, Rio DC. 1997. Drosophila P-element transposase is a novel site-specific endonuclease. *Genes Dev.* 11:2137–51
14. Beall EL, Rio DC. 1998. Transposase makes critical contacts with, and is stimulated by, single-stranded DNA at the P element termini in vitro. *EMBO J.* 17:2122–36
15. Becker HA, Kunze R. 1997. Maize activator transposase has a bipartite DNA binding domain that recognizes subterminal sequences and the terminal inverted repeats. *Mol. Gen. Genet.* 254:219–30
16. Berg DE, Howe MM. 1989. *Mobile DNA.* Washington, DC: Am. Soc. Microbiol.
17. Bolland S, Kleckner N. 1995. The two single-strand cleavages at each end of Tn10 occur in a specific order during transposition. *Proc. Natl. Acad. Sci. USA* 92:7814–18
18. Bolland S, Kleckner N. 1996. The three chemical steps of Tn10/IS10 transposition involve repeated utilization of a single active site. *Cell* 84:223–33
19. Braam LA, Reznikoff WS. 1998. Functional characterization of the Tn5 transposase by limited proteolysis. *J. Biol. Chem.* 273:10908–13
20. Brown PO. 1997. *Integration, Retroviruses*, ed. JM Coffin, SH Hughes, HE Varmus, pp. 161–203. Plainview, NY: Cold Spring Harbor Lab. Press
21. Bujacz G, Alexandratos J, Qing ZL, Clement-Mella C, Wlodawer A. 1996. The catalytic domain of human immunodeficiency virus integrase: ordered active site in the F185H mutant. *FEBS Lett.* 398:175–78
22. Bujacz G, Alexandratos J, Wlodawer A, Merkel G, Andrake M, et al. 1997. Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity. *J. Biol. Chem.* 272:18161–68
23. Bujacz G, Jaskolski M, Alexandratos J, Wlodawer A, Merkel G, et al. 1995. High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J. Mol. Biol.* 253:333–46
24. Bujacz G, Jaskolski M, Alexandratos J, Wlodawer A, Merkel G, et al. 1996. The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations. *Structure* 4:89–96
25. Bushman FD, Engelman A, Palmer I, Wingfield P, Craigie R. 1993. Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. *Proc. Natl. Acad. Sci. USA* 90:3428–32
26. Cai M, Zheng R, Caffrey M, Craigie R, Clore GM, Gronenborn AM. 1997. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat. Struct. Biol.* 4:567–77
27. Chaconas G, Lavoie BD, Watson MA. 1996. DNA transposition: jumping gene

machine, some assembly required. *Curr. Biol.* 6:817–20
28. Chalmers R, Guhathakurta A, Benjamin H, Kleckner N. 1998. IHF modulation of Tn10 transposition: sensory transduction of supercoiling status via a proposed protein/DNA molecular spring. *Cell* 93:897–908
29. Chalmers RM, Kleckner N. 1994. Tn10/IS10 transposase purification, activation, and in vitro reaction. *J. Biol. Chem.* 269: 8029–35
30. Chow SA, Vincent KA, Ellison V, Brown PO. 1992. Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science* 255:723–26
31. Clubb RT, Schumacher S, Mizuuchi K, Gronenborn AM, Clore GM. 1997. Solution structure of the Iγ subdomain of the Mu end DNA-binding domain of phage Mu transposase. *J. Mol. Biol.* 273:19–25
32. Coen ES, Robbins TP, Almeida J, Hudson A, Carpenter R. 1989. Consequences and mechanisms of transposition in Antirrhinum majus. In *Mobile DNA*, ed. D Berg, M Howe, pp. 413–36. Washington, DC: Am. Soc. Microbiol.
33. Colot V, Haedens V, Rossignol JL. 1998. Extensive, nonrandom diversity of excision footprints generated by Ds-like transposon Ascot-1 suggests new parallels with V(D)J recombination. *Mol. Cell Biol.* 18:4337–46
34. Craig NL. 1996. Transposon Tn7. *Curr. Top. Microbiol. Immunol.* 204:27–48
35. Craigie R, Mizuuchi K. 1987. Transposition of Mu DNA: joining of Mu to target DNA can be uncoupled from cleavage at the ends of Mu. *Cell* 51:493–501
36. Craigie R, Mizuuchi M, Mizuuchi K. 1984. Site-specific recognition of the bacteriophage Mu ends by the Mu A protein. *Cell* 39:387–94
37. Dahm SC, Uhlenbeck OC. 1991. Role of divalent metal ions in the hammerhead RNA cleavage reaction. *Biochemistry* 30:9464–69
37a. Davies DR, Braam LM, Reznikoff WS, Rayment I. 1999. Three dimensional structure of a Tu5-related protein determined to 2.9 Å resolution. *J. Biol. Chem.* 274: 11904–13
38. Doak TG, Doerder FP, Jahn CL, Herrick G. 1994. A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. *Proc. Natl. Acad. Sci. USA* 91:942–46
39. Drelich M, Haenggi M, Mous J. 1993. Conserved residues Pro-109 and Asp-116 are required for interaction of the human immunodeficiency virus type 1 integrase protein with its viral DNA substrate. *J. Virol.* 67:5041–44
40. Drelich M, Wilhelm R, Mous J. 1992. Identification of amino acid residues critical for endonuclease and integration activities of HIV-1 IN protein in vitro. *Virology* 188:459–68
41. Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR. 1994. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* 266:1981–86
42. Eijkelenboom AP, Lutzke RA, Boelens R, Plasterk RH, Kaptein R, Hard K. 1995. The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nat. Struct. Biol.* 2:807–10
43. Eijkelenboom AP, van den Ent FM, Vos A, Doreleijers JF, Hard K, et al. 1997. The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: a three-helix bundle stabilized by zinc. *Curr. Biol.* 7:739–46
44. Ellison V, Brown PO. 1994. A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration in vitro. *Proc. Natl. Acad. Sci. USA* 91:7316–20
45. Ellison V, Gerton J, Vincent KA, Brown PO. 1995. An essential interaction between distinct domains of HIV-1 integrase me-

diates assembly of the active multimer. *J. Biol. Chem.* 270:3320–26
46. Engelman A, Bushman FD, Craigie R. 1993. Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *EMBO J.* 12:3269–75
47. Engelman A, Craigie R. 1992. Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. *J. Virol.* 66:6361–69
48. Engelman A, Hickman AB, Craigie R. 1994. The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. *J. Virol.* 68:5911–17
49. Engelman A, Mizuuchi K, Craigie R. 1991. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* 67:1211–21
50. Engels WR. 1996. P elements in Drosophila. *Curr. Top. Microbiol. Immunol.* 204: 103–23
51. Engels WR, Johnson-Schlitz DM, Eggleston WB, Sved J. 1990. High-frequency P element loss in Drosophila is homolog dependent. *Cell* 62:515–25
52. Esposito D, Craigie R. 1998. Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J.* 17: 5832–43
53. Farnet CM, Bushman FD. 1997. HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro. *Cell* 88:483–92
54. Fayet O, Ramond P, Polard P, Prere MF, Chandler M. 1990. Functional similarities between retroviruses and the IS3 family of bacterial insertion sequences? *Mol. Microbiol.* 4:1771–77
55. Fletcher TM, Soares MA, McPhearson S, Hui H, Wiskerchen M, et al. 1997. Complementation of integrase function in HIV-1 virions. *EMBO J.* 16:5123–38
56. Gerton JL, Ohgi S, Olsen M, DeRisi J, Brown PO. 1998. Effects of mutations in residues near the active site of human immunodeficiency virus type 1 integrase on specific enzyme-substrate interactions. *J. Virol.* 72:5046–55
57. Gierl A. 1996. The En/Spm transposable element of maize. *Curr. Top. Microbiol. Immunol.* 204:145–59
58. Goldgur Y, Dyda F, Hickman AB, Jenkins TM, Craigie R, Davies DR. 1998. Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc. Natl. Acad. Sci. USA* 95:9150–54
59. Gopaul DN, Guo F, Van DG. 1998. Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination. *EMBO J.* 17:4175–87
60. Goryshin IY, Reznikoff WS. 1998. Tn5 in vitro transposition. *J. Biol. Chem.* 273: 7367–74
61. Grindley ND, Leschziner AE. 1995. DNA transposition: from a black box to a color monitor. *Cell* 83:1063–66
62. Hagemann AT, Craig NL. 1993. Tn7 transposition creates a hotspot for homologous recombination at the transposon donor site. *Genetics* 133:9–16
63. Haniford DB, Benjamin HW, Kleckner N. 1991. Kinetic and structural analysis of a cleaved donor intermediate and a strand transfer intermediate in Tn10 transposition. *Cell* 64:171–79
64. Haniford D, Kleckner N. 1994. Tn 10 transposition in vivo: temporal separation of cleavages at the two transposon ends and roles of terminal basepairs subsequent to interaction of ends. *EMBO J.* 13:3401–11
65. Haren L. 1998. *Relations structure/fonction des facteurs protéiques exprimés par l'elément transposable bacterien IS911*. PhD thesis. Univ. Paul Sabatier, Toulouse, France
66. Haren L, Polard P, Ton-Hoang B, Chandler M. 1998. Multiple oligomerisation domains in the IS911 transposase: a leucine

zipper motif is essential for activity. *J. Mol. Biol.* 283:29–41
67. Heuer TS, Brown PO. 1997. Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photocross-linking. *Biochemistry* 36:10655–65
68. Heuer TS, Brown PO. 1998. Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex. *Biochemistry* 37:6667–78
69. Hiom K, Melek M, Gellert M. 1998. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* 94:463–70
70. Jenkins TM, Engelman A, Ghirlando R, Craigie R. 1996. A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization. *J. Biol. Chem.* 271:7712–18
71. Jenkins TM, Esposito D, Engelman A, Craigie R. 1997. Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photocrosslinking. *EMBO J.* 16:6849–59
72. Jones JM, Nakai H. 1997. The phiX174-type primosome promotes replisome assembly at the site of recombination in bacteriophage Mu transposition. *EMBO J.* 16:6886–95
73. Junop MS, Haniford DB. 1996. Multiple roles for divalent metal ions in DNA transposition: distinct stages of Tn10 transposition have different Mg2+ requirements. *EMBO J.* 15:2547–55
74. Junop MS, Haniford DB. 1997. Factors responsible for target site selection in Tn10 transposition: a role for the DDE motif in target DNA capture. *EMBO J.* 16:2646–55
75. Kalpana GV, Goff SP. 1993. Genetic analysis of homomeric interactions of human immunodeficiency virus type 1 integrase using the yeast two-hybrid system. *Proc. Natl. Acad. Sci. USA* 90:10593–97
76. Katz RA, Merkel G, Kulkosky J, Leis J, Skalka AM. 1990. The avian retroviral IN protein is both necessary and sufficient for integrative recombination in vitro. *Cell* 63:87–95
77. Kaufman PD, Rio DC. 1992. P element transposition in vitro proceeds by a cut-and-paste mechanism and uses GTP as a cofactor. *Cell* 69:27–39
78. Kennedy AK, Guhathakurta A, Kleckner N, Haniford DB. 1998. Tn10 transposition via a DNA hairpin intermediate. *Cell* 95:125–34
79. Kennedy AK, Haniford DB. 1996. Isolation and characterization of IS10 transposase separation of function mutants: identification of amino acid residues in transposase that are important for active site function and the stability of transposition intermediates. *J. Mol. Biol.* 256:533–47
80. Khan E, Mack JP, Katz RA, Kulkosky J, Skalka AM. 1991. Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Res.* 19:851–60 (Erratum. 1991. *Nucleic Acids Res.* 19:1358)
81. Kim K, Harshey RM. 1995. Mutational analysis of the att DNA-binding domain of phage Mu transposase. *Nucleic Acids Res.* 23:3937–43
82. Kim K, Namgoong SY, Jayaram M, Harshey RM. 1995. Step-arrest mutants of phage Mu transposase. Implications in DNA-protein assembly, Mu end cleavage, and strand transfer. *J. Biol. Chem.* 270:1472–79
83. Kleckner N, Chalmers RM, Kwon D, Sakai J, Bolland S. 1996. Tn10 and IS10 transposition and chromosome rearrangements: mechanisms and regulation in vivo and in vitro. *Curr. Top. Microbiol. Immunol.* 204:49–82
84. Krementsova E, Giffin MJ, Pincus D, Baker TA. 1998. Mutational analysis of the Mu transposase. Contributions of two distinct regions of domain II to recombination. *J. Biol. Chem.* 273:31358–65

85. Kruklitis R, Welty DJ, Nakai H. 1996. ClpX protein of *Escherichia coli* activates bacteriophage Mu transposase in the strand transfer complex for initiation of Mu DNA synthesis. *EMBO J.* 15:935–44
86. Kukolj G, Skalka AM. 1995. Enhanced and coordinated processing of synapsed viral DNA ends by retroviral integrases in vitro. *Genes Dev.* 9:2556–67
87. Kulkosky J, Jones KS, Katz RA, Mack JP, Skalka AM. 1992. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell Biol.* 12:2331–38
88. Kulkosky J, Katz RA, Merkel G, Skalka AM. 1995. Activities and substrate specificity of the evolutionarily conserved central domain of retroviral integrase. *Virology* 206:448–56
89. Kunze R. 1996. The maize transposable element activator (Ac). *Curr. Top. Microbiol. Immunol.* 204:161–94
90. Kunze R, Behrens U, Courage-Franzkowiak U, Feldmar S, Kuhn S, Lutticke R. 1993. Dominant transposition-deficient mutants of maize activator (Ac) transposase. *Proc. Natl. Acad. Sci. USA* 90: 7094–98
91. Kwon D, Chalmers RM, Kleckner N. 1995. Structural domains of IS10 transposase and reconstitution of transposition activity from proteolytic fragments lacking an interdomain linker. *Proc. Natl. Acad. Sci. USA* 92:8234–38
92. LaFemina RL, Callahan PL, Cordingley MG. 1991. Substrate specificity of recombinant human immunodeficiency virus integrase protein. *J. Virol.* 65:5624–30
93. Lampe DJ, Churchill ME, Robertson HM. 1997. A purified mariner transposase is sufficient to mediate transposition in vitro (erratum). *EMBO J.* 16:4153
94. Lavoie BD, Chaconas G. 1996. Transposition of phage Mu DNA. *Curr. Top. Microbiol. Immunol.* 204:83–102
95. Lavoie BD, Chan BS, Allison RG, Chaconas G. 1991. Structural aspects of a higher order nucleoprotein complex: induction of an altered DNA structure at the Mu-host junction of the Mu type 1 transpososome. *EMBO J.* 10:3051–59
96. Lee CC, Beall EL, Rio DC. 1998. DNA binding by the KP repressor protein inhibits P-element transposase activity in vitro. *EMBO J.* 17:4166–74
97. Lee SP, Xiao J, Knutson JR, Lewis MS, Han MK. 1997. Zn2+ promotes the self-association of human immunodeficiency virus type-1 integrase in vitro. *Biochemistry* 36:173–80
98. Leschziner AE, Griffin TI, Grindley NF. 1998. Tn552 transposase catalyzes concerted strand transfer in vitro. *Proc. Natl. Acad. Sci. USA* 95:7345–50
99. Levchenko I, Luo L, Baker TA. 1995. Disassembly of the Mu transposase tetramer by the ClpX chaperone. *Genes Dev.* 9: 2399–408
100. Lodi PJ, Ernst JA, Kuszewski J, Hickman AB, Engelman A, et al. 1995. Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry* 34:9826–33
101. Lohe AR, Sullivan DT, Hartl DL. 1996. Subunit interactions in the mariner transposase. *Genetics* 144:1087–95
102. Lutzke RP, Plasterk RA. 1998. Structure-based mutational analysis of the C-terminal DNA-binding domain of human immunodeficiency virus type 1 integrase: critical residues for protein oligomerization and DNA binding. *J. Virol.* 72:4841–48
103. Lutzke RA, Vink C, Plasterk RH. 1994. Characterization of the minimal DNA-binding domain of the HIV integrase protein. *Nucleic Acids Res.* 22:4125–31
104. Machida C, Machida Y. 1989. Regulation of IS1 transposition by the insA gene product. *J. Mol. Biol.* 208:567–74
105. Maekawa T, Amemura-Maekawa J, Ohtsubo E. 1993. DNA binding domains in Tn3 transposase. *Mol. Gen. Genet.* 236: 267–74

106. Mahillon J, Chandler M. 1998. Insertion sequences. *Microbiol. Mol. Biol. Rev.* 62:725–74
107. Maxwell A, Craigie R, Mizuuchi K. 1987. B protein of bacteriophage mu is an ATPase that preferentially stimulates intermolecular DNA strand transfer. *Proc. Natl. Acad. Sci. USA* 84:699–703
108. May EW, Craig NL. 1996. Switching from cut-and-paste to replicative Tn7 transposition. *Science* 272:401–4
109. Mizuuchi K. 1992. Polynucleotidyl transfer reactions in transpositional DNA recombination. *J. Biol. Chem.* 267:21273–76
110. Mizuuchi K. 1992. Transpositional recombination: mechanistic insights from studies of Mu and other elements. *Annu. Rev. Biochem.* 61:1011–51
111. Mizuuchi K. 1997. Polynucleotidyl transfer reactions in site-specific DNA recombination. *Genes Cells* 2:1–12
112. Mizuuchi K, Adzuma K. 1991. Inversion of the phosphate chirality at the target site of Mu DNA strand transfer: evidence for a one-step transesterification mechanism. *Cell* 66:129–140
113. Mizuuchi M, Baker TA, Mizuuchi K. 1991. DNase protection analysis of the stable synaptic complexes involved in Mu transposition. *Proc. Natl. Acad. Sci. USA* 88:9031–35
114. Mizuuchi M, Baker TA, Mizuuchi K. 1992. Assembly of the active form of the transposase-Mu DNA complex: a critical control point in Mu transposition. *Cell* 70:303–11
115. Murphy JE, Goff SP. 1992. A mutation at one end of Moloney murine leukemia virus DNA blocks cleavage of both ends by the viral integrase in vivo. *J. Virol.* 66: 5092–95
116. Naigamwalla DZ, Chaconas G. 1997. A new set of Mu DNA transposition intermediates: alternate pathways of target capture preceding strand transfer. *EMBO J.* 16:5227–34
117. Nakayama C, Teplow DB, Harshey RM. 1987. Structural domains in phage Mu transposase: identification of the site-specific DNA-binding domain. *Proc. Natl. Acad. Sci. USA* 84:1809–13
118. Namgoong SY, Harshey RM. 1998. The same two monomers within a MuA tetramer provide the DDE domains for the strand cleavage and strand transfer steps of transposition. *EMBO J.* 17:3775–85
119. Namgoong SY, Kim K, Saxena P, Yang JY, Jayaram M, et al. 1998. Mutational analysis of domain IIβ of bacteriophage Mu transposase: domains IIα and IIβ belongs to different catalytic complementation groups. *J. Mol. Biol.* 275:221–32
120. Olasz F, Stalder R, Arber W. 1993. Formation of the tandem repeat (IS30)2 and its role in IS30-mediated transpositional DNA rearrangements. *Mol. Gen. Genet.* 239:177–87
121. Piccirilli JA, Vyle JS, Caruthers MH, Cech TR. 1993. Metal ion catalysis in the Tetrahymena ribozyme reaction. *Nature* 361:85–88
122. Pietrokovski S, Henikoff S. 1997. A helix-turn-helix DNA-binding motif predicted for transposases of DNA transposons. *Mol. Gen. Genet.* 254:689–95
123. Plasterk RH. 1996. The Tc1/mariner transposon family. *Curr. Top. Microbiol. Immunol.* 204:125–43
124. Polard P, Chandler M. 1995. An in vivo transposase-catalyzed single-stranded DNA circularization reaction. *Genes Dev.* 9:2846–58
125. Polard P, Chandler M. 1995. Bacterial transposases and retroviral integrases. *Mol. Microbiol.* 15:13–23
126. Polard P, Ton-Hoang B, Haren L, Betermier M, Walczak R, Chandler M. 1996. IS911-mediated transpositional recombination in vitro. *J. Mol. Biol.* 264:68–81
127. Reimmann C, Haas D. 1987. Mode of replicon fusion mediated by the dupli-

cated insertion sequence IS21 in *Escherichia coli*. *Genetics* 115:619–25
128. Reznikoff WS. 1993. The Tn5 transposon. *Annu. Rev. Microbiol.* 47:945–63
129. Rezsohazy R, Hallet B, Delcour J, Mahillon J. 1993. The IS4 family of insertion sequences: evidence for a conserved transposase motif. *Mol. Microbiol.* 9:1283–95
130. Rice P, Craigie R, Davies DR. 1996. Retroviral integrases and their cousins. *Curr. Opin. Struct. Biol.* 6:76–83
131. Rice P, Mizuuchi K. 1995. Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell* 82:209–20
132. Rowland SJ, Dyke KG. 1990. Tn552, a novel transposable element from Staphylococcus aureus. *Mol. Microbiol.* 4:961–75
133. Rowland SJ, Sherratt DJ, Stark WM, Boocock MR. 1995. Tn552 transposase purification and in vitro activities. *EMBO J.* 14:196–205
134. Saedler H, Gierl A. 1996. *Transposable Elements*. Berlin: Springer-Verlag
135. Sakai J, Chalmers RM, Kleckner N. 1995. Identification and characterization of a pre-cleavage synaptic complex that is an early intermediate in Tn10 transposition. *EMBO J.* 14:4374–83
136. Sakai J, Kleckner N. 1997. The Tn10 synaptic complex can capture a target DNA only after transposon excision. *Cell* 89:205–14
137. Sarnovsky RJ, May EW, Craig NL. 1996. The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *EMBO J.* 15:6348–61
138. Savilahti H, Mizuuchi K. 1996. Mu transpositional recombination: donor DNA cleavage and strand transfer in trans by the Mu transposase. *Cell* 85:271–80
139. Savilahti H, Rice PA, Mizuuchi K. 1995. The phage Mu transpososome core: DNA requirements for assembly and function. *EMBO J.* 14:4893–903
140. Schumacher S, Clubb RT, Cai M, Mizuuchi K, Clore GM, Gronenborn AM. 1997. Solution structure of the Mu end DNA-binding Iβ subdomain of phage Mu transposase: modular DNA recognition by two tethered domains. *EMBO J.* 16:7532–41
141. Scottoline BP, Chow S, Ellison V, Brown PO. 1997. Disruption of the terminal base pairs of retroviral DNA during integration. *Genes Dev.* 11:371–82
142. Sharpe PL, Craig NL. 1998. Host proteins can stimulate Tn7 transposition: a novel role for the ribosomal protein L29 and the acyl carrier protein. *EMBO J.* 17:5822–31
143. Sherman PA, Fyfe JA. 1990. Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. *Proc. Natl. Acad. Sci. USA* 87:5119–23
144. Sherratt D. 1989. Tn3 and related transposable elements: site-specific recombination and transposition. In *Mobile DNA*, ed. D. Berg, M. Howe, pp. 163–84. Washington, DC: Am. Soc. Microbiol.
145. Shibagaki Y, Chow SA. 1997. Central core domain of retroviral integrase is responsible for target site selection. *J. Biol. Chem.* 272:8361–69
146. Stalder R, Caspers P, Olasz F, Arber W. 1990. The N-terminal domain of the insertion sequence 30 transposase interacts specifically with the terminal inverted repeats of the element. *J. Biol. Chem.* 265:3757–62
147. Stellwagen AE, Craig NL. 1997. Avoiding self: two Tn7-encoded proteins mediate target immunity in Tn7 transposition. *EMBO J.* 16:6823–34
148. Surette MG, Buch SJ, Chaconas G. 1987. Transpososomes: stable protein-DNA complexes involved in the in vitro transposition of bacteriophage Mu DNA. *Cell* 49:253–62

149. Surette MG, Harkness T, Chaconas G. 1991. Stimulation of the Mu A protein-mediated strand cleavage reaction by the Mu B protein, and the requirement of DNA nicking for stable type 1 transpososome formation. In vitro transposition characteristics of mini-Mu plasmids carrying terminal base pair mutations. *J. Biol. Chem.* 266:3118–24
150. Szeverenyi I, Bodoky T, Olasz F. 1996. Isolation, characterization and transposition of an (IS2)2 intermediate. *Mol. Gen. Genet.* 251:281–89
151. Tavakoli NP, DeVost J, Derbyshire KM. 1997. Defining functional regions of the IS903 transposase. *J. Mol. Biol.* 274:491–504
152. Ton-Hoang B. 1998. *Etude du mécanisme de transposition de la séquence d'insertion bactérienne IS911: role d'une forme circulaire dans sa mobilité*. PhD thesis. Univ. Paul Sabatier, Toulouse, France
153. Ton-Hoang B, Betermier M, Polard P, Chandler M. 1997. Assembly of a strong promoter following IS911 circularization and the role of circles in transposition. *EMBO J.* 16:3357–71
154. Ton-Hoang B, Polard P, Chandler M. 1998. Efficient transposition of IS911 circles in vitro. *EMBO J.* 17:1169–81
155. Uchiyama Y, Iwai S, Ueno Y, Ikehara M, Ohtsuka E. 1994. Role of the Mg2+ ion in the *Escherichia coli* ribonuclease HI reaction. *J. Biochem. (Tokyo).* 116:1322–29
156. van den Ent FM, Vos A, Plasterk RH. 1998. Mutational scan of the human immunodeficiency virus type 2 integrase protein. *J. Virol.* 72:3916–24
157. van Gent DC, Elgersma Y, Bolk MW, Vink C, Plasterk RH. 1991. DNA binding properties of the integrase proteins of human immunodeficiency viruses types 1 and 2. *Nucleic Acids Res.* 19:3821–27
158. van Gent DC, Groeneger AA, Plasterk RH. 1992. Mutational analysis of the integrase protein of human immunodeficiency virus type 2. *Proc. Natl. Acad. Sci. USA* 89:9598–602
159. van Gent DC, Mizuuchi K, Gellert M. 1996. Similarities between initiation of V(D)J recombination and retroviral integration. *Science* 271:1592–94
160. van Gent DC, Oude Groeneger AA, Plasterk RH. 1993. Identification of amino acids in HIV-2 integrase involved in site-specific hydrolysis and alcoholysis of viral DNA termini. *Nucleic Acids Res.* 21:3373–77
161. van Gent DC, Vink C, Groeneger AA, Plasterk RH. 1993. Complementation between HIV integrase proteins mutated in different domains. *EMBO J.* 12:3261–67
162. van Luenen HG, Colloms SD, Plasterk RH. 1994. The mechanism of transposition of Tc3 in *C. elegans*. *Cell* 7984:293–301
163. van Pouderoyen G, Ketting RF, Perrakis A, Plasterk RHA, Sixma TK. 1997. Crystal structure of the specific DNA-binding domain of Tc3 transposase of C. elegans in complex with transposon DNA. *EMBO J.* 16:6044–54
164. Vink C, Lutzke RA, Plasterk RH. 1994. Formation of a stable complex between the human immunodeficiency virus integrase protein and viral DNA. *Nucleic Acids Res.* 22:4103–10
165. Vink C, Oude Groeneger AM, Plasterk RH. 1993. Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein. *Nucleic Acids Res.* 21:1419–25
166. Vink C, van Gent DC, Elgersma Y, Plasterk RH. 1991. Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNA recognition sequence for efficient cleavage. *J. Virol.* 65:4636–44
167. Vos JC, Plasterk RH. 1994. Tc1 transposase of Caenorhabditis elegans is an endonuclease with a bipartite DNA binding domain. *EMBO J.* 13:6125–32
168. Vos JC, van Luenen HG, Plasterk RH.

1993. Characterization of the Caenorhabditis elegans Tc1 transposase in vivo and in vitro. *Genes Dev.* 7:1244–53

169. Wang Z, Harshey RM. 1994. Crucial role for DNA supercoiling in Mu transposition: a kinetic study. *Proc. Natl. Acad. Sci. USA* 91:699–703
170. Wang Z, Namgoong SY, Zhang X, Harshey RM. 1996. Kinetic and structural probing of the precleavage synaptic complex (type 0) formed during phage Mu transposition. Action of metal ions and reagents specific to single-stranded DNA. *J. Biol. Chem.* 271:9619–26
171. Watson MA, Chaconas G. 1996. Three-site synapsis during Mu DNA transposition: a critical intermediate preceding engagement of the active site. *Cell* 85:435–45
172. Wei SQ, Mizuuchi K, Craigie R. 1997. A large nucleoprotein assembly at the ends of the viral DNA mediates retroviral DNA integration. *EMBO J.* 16:7511–20
173. Weinreich MD, Mahnke-Braam L, Reznikoff WS. 1994. A functional analysis of the Tn5 transposase. Identification of domains required for DNA binding and multimerization. *J. Mol. Biol.* 241:166–77
174. Wiater LA, Grindley ND. 1991. Gamma delta transposase. Purification and analysis of its interaction with a transposon end. *J. Biol. Chem.* 266:1841–49
175. Wiegand TW, Reznikoff WS. 1994. Interaction of Tn5 transposase with the transposon termini. *J. Mol. Biol.* 235:486–95
176. Wolfe AL, Felock PJ, Hastings JC, Blau CU, Hazuda DJ. 1996. The role of manganese in promoting multimerization and assembly of human immunodeficiency virus type 1 integrase as a catalytically active complex on immobilized long terminal repeat substrates. *J. Virol.* 70:1424–32
177. Wu Z, Chaconas G. 1995. A novel DNA binding and nuclease activity in domain III of Mu transposase: evidence for a catalytic region involved in donor cleavage. *EMBO J.* 14:3835–43
178. Yamauchi M, Baker TA. 1998. An ATP-ADP switch in MuB controls progression of the Mu transposition pathway. *EMBO J.* 17:5509–18
179. Yang JY, Kim K, Jayaram M, Harshey RM. 1995. A domain sharing model for active site assembly within the Mu A tetramer during transposition: the enhancer may specify domain contributions. *EMBO J.* 14:2374–84
180. Yang W, Steitz TA. 1995. Recombining the structures of HIV integrase, RuvC and RNase H. *Structure* 3:131–34
181. Yin JC, Krebs MP, Reznikoff WS. 1988. Effect of dam methylation on Tn5 transposition. *J. Mol. Biol.* 199:35–45
182. Yurieva O, Nikiforov V. 1996. Catalytic center quest: comparison of transposases belonging to the Tn3 family reveals an invariant triad of acidic amino acid residues. *Biochem. Mol. Biol. Int.* 38:15–20
183. Zerbib D, Jakowec M, Prentki P, Galas DJ, Chandler M. 1987. Expression of proteins essential for IS1 transposition: specific binding of InsA to the ends of IS1. *EMBO J.* 6:3163–69
184. Zheng R, Jenkins TM, Craigie R. 1996. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc. Natl. Acad. Sci. USA* 93:13659–64

Annual Review of Microbiology
Volume 53, 1999

CONTENTS