Transposases and Integrases

Secondary article

Nancy L Craig, *Johns Hopkins School of Medicine, Baltimore, Maryland, USA*

Transposases and integrases are the proteins that mediate transposition reactions, i.e. the recombination reaction in which discrete DNA segments move between nonhomologous sites.

Article Contents

- Introduction
- Overview of Properties and Mechanisms
- Classes of Transposases and Integrases
- Evolutionary Relationships

Introduction

Transposable elements are discrete pieces of deoxyribonucleic acid (DNA) that can move between many nonhomologous positions in chromosomal DNA and extrachromosomal DNAs, such as plasmids and viruses. These mobile elements are widespread in nature and have been found in virtually every organism examined. The movement of these elements is a major source of genetic plasticity, altering information within genomes or adding new genetic determinants.

The element-encoded recombinase that mediates transposition is called a transposase, or an integrase in the case of an element such as a retrovirus whose life cycle contains a ribonucleic acid (RNA) intermediate. Although retroviral-like elements do involve an RNA intermediate, a DNA copy of the element generated by reverse transcription of the RNA is actually acted upon by the integrase. Thus, transposases and integrases both execute recombination using DNA substrates. The recombinase recognizes and binds to specific sequences at the ends of the transposon, mediates synaptic interactions between the ends to bring them together, interacts with the target DNA, and executes the DNA breakage and joining reactions that underlie recombination. These recombinases may act alone or, in some cases, in concert with element and host-encoded proteins. Typically, transposases contain only a single species of polypeptide; however, the transposase of the bacterial element Tn*7* comprises two different proteins.

Overview of Properties and Mechanisms

At the core of all transposition reactions is an ordered and coordinated series of DNA breakage and joining reactions that take place in elaborate protein–DNA complexes and result in the translocation of DNA from place to place. Despite the varied natures of transposable elements and their hosts, biochemical analysis of transposition mechanisms and the structural characterization of transposases and integrases in the past few years has revealed striking similarities in the transposition systems of elements from a wide variety of hosts including bacteria, yeast, flies, worms, plants and humans.

Pathways of DNA breakage and joining

The nature of the strand breakage and joining events that can be carried out by the transposase determines the recombination pathway. A major mechanism of transposition is ‘cut and paste’ transposition (**Figure 1**). The transposon is excised from the flanking donor DNA backbone by double-strand breaks at each end of the element. Joining to a target DNA then occurs by the direct attack of the exposed 3′ hydroxyl (OH) ends of the transposon at staggered positions on the target DNA. Because the positions at which each transposon end joins to the target DNA are staggered, the newly inserted transposon is flanked by small gaps. These gaps are then repaired by host functions to give the target site duplications that are characteristic features of transposition. This type of recombination is widespread, with elements that

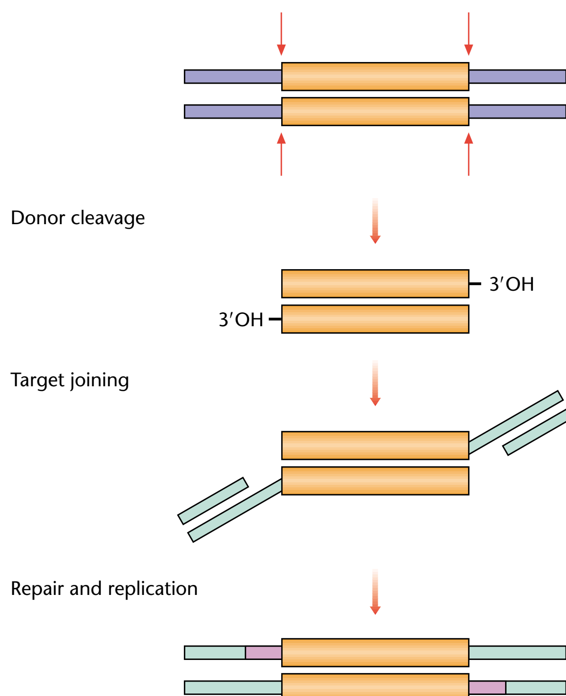

Figure 1 Cut and paste transposition.

are able to execute this type of recombination being present in organisms ranging from bacteria to plants; relics of such elements have also been observed in mammals.

As described below, executing cleavages at the 3′ ends of the transposon is common to all systems that have been studied. An interesting question in cut and paste transposition is how cleavage occurs at both the 3′ and 5′ ends of the transposon. In the case of the bacterial elements insertion sequence (IS) 10 and IS50, a single transposase monomer functioning at each terminus appears to mediate these steps (**Figure 2**). First, cleavage occurs at the 3′ end of the element; the newly exposed 3′ OH end then attacks its complementary 5′ end strand, resulting in a hairpin at the end of the element. This hairpin is then resolved by the transposase, resulting in the generation of a transposon with exposed 3′ and 5′ ends. As described in more detail below, these reactions appear to be executed by multiple phosphoryl transfer reactions performed by the same active site within the transposase.

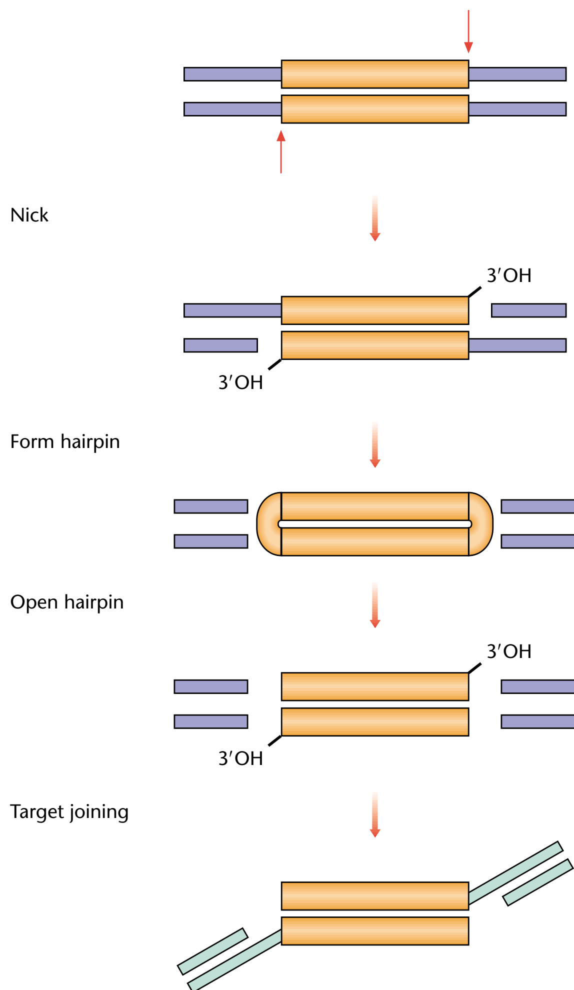

Figure 2 Donor cleavage.

A widespread family of elements in which double-strand breaks occur at the ends of the element, followed by the joining of the 3′ exposed element to the target DNA, is that carried out by the Tc1/mariner elements. These elements were originally observed in worms (Tc1) and flies (mariner) but are now known to be highly dispersed. It is not yet known how this transposase executes the cleavages at both the 3′ and 5′ strands.

The bacterial transposon Tn*7* also translocates by the introduction of double-strand breaks at the ends of the element, followed by the joining of the element to the target DNA. The chemical activities are distributed between two Tn*7*-encoded polypeptides, TnsA and TnsB. TnsB executes cleavages and target joining at the 3′ ends of the element, whereas TnsA executes the cleavages at the 5′ ends of the element. Interestingly, TnsB resembles the transposases and integrases of the retroviral integrase superfamily (see below), whereas TnsA is most similar in structure to a restriction enzyme.

This same fundamental process of exposure of reactive 3′ OH ends of a mobile element, which subsequently attack the target DNA, also underlies retroviral integration (**Figure 3**). In this case, an RNA copy of the element is first generated by the host transcription machinery. This RNA is then used as a template for reverse transcription to generate a double-stranded DNA copy of the element that is the actual substrate for integration. Recombination initiates by cleavage reactions that remove a few nucleotides at the termini of the element to expose the actual 3′ OH ends, which attack the target DNA, again at staggered positions, so that a target site duplication results from host repair.

In both the cut and paste pathway and the retroviral integration pathway, the newly inserted transposable element is connected only to the target DNA and is no longer joined to the donor DNA that originally flanked the mobile element. In some transposition reactions in bacteria, however, the newly inserted transposon remains joined to the old donor site upon insertion into the new insertion site. This pathway is, for example, carried out by bacteriophage Mu, which uses transposition to replicate its DNA during lytic phage growth (**Figure 4**). In this case, the transposon sitting in its donor site undergoes cleavage reactions only at the 3′ ends of the element, leaving the 5′ ends attached to the donor DNA. As with the other elements described above, these exposed 3′ ends attack the target DNA at staggered positions. Following this recombination event, host DNA repair and replication across the element generate a structure called a cointegrate, containing two copies of the transposable element joined by the flanking donor replicon and the target replicon. Some other elements, notably those of the Tn*3* family, also likely use this recombination pathway.

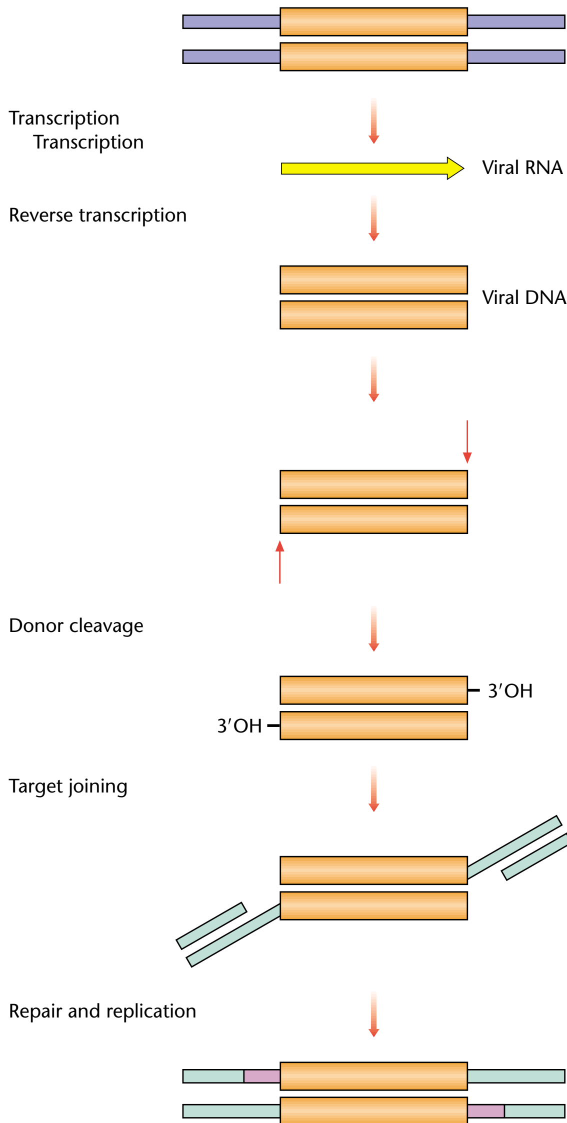

Figure 3 Retroviral integration.

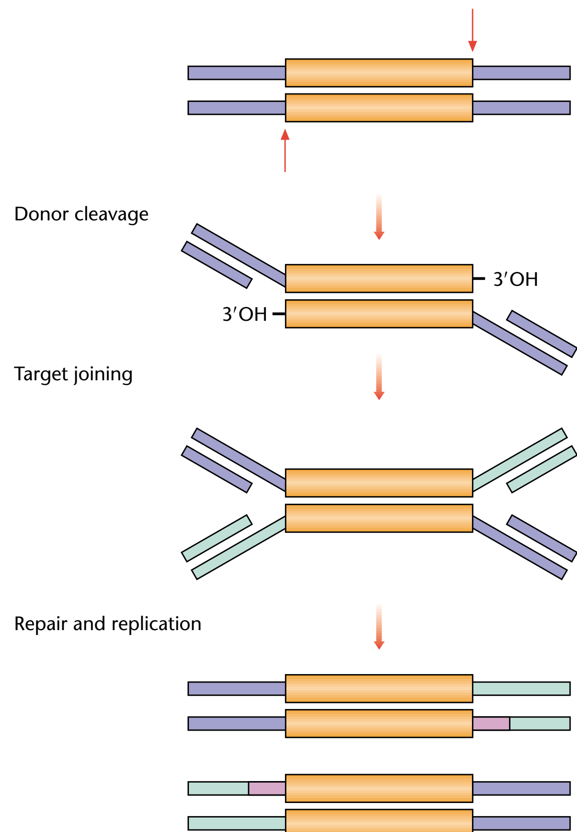

Figure 4 Bacteriophage Mu.

Transposases act at the ends of transposable elements

The ends of transposable elements are organized as inverted repeats. Thus the binding of the transposase to these sequences specifically and symmetrically positions the recombinase with respect to its DNA substrate, the transposon end flanked by donor DNA. Often these terminal transposase binding sites are tens of base pairs in length. However, in some cases there may be multiple transposase binding sites so that the transposon end sequence information required for transposition may be hundreds of base pairs long. In a number of elements, the sequences at the extreme termini of the element are critical for specifying information necessary for DNA breakage and joining, although the specific transposase binding site lies just internal to these terminal residues.

In addition to transposase binding sites, the ends of some transposable elements include binding sites for other proteins that assist the transposase. In the case of several bacterial elements, binding sites for the sequence-specific DNA binding and bending protein IHF (integration host factor) are located in the ends of the element, and these sites are necessary for transposition. Another example of the involvement of a DNA bending protein in transposition is the nonsequence-specific DNA binding and bending protein HU, which assists Mu transposition. In these cases, the role of these host-encoded proteins is an architectural one to promote the assembly of the transposase and substrate DNA into the highly organized protein–DNA complexes that actually execute transposition.

Another example of a nontransposon-encoded protein interacting with the ends of the transposon is found in the case of the P element of *Drosophila*. In this case, the Ku protein that is involved in double-strand break repair also binds specifically to the ends of the element. Although the exact role of Ku remains to be determined, it does play a key role in transposition as Ku^{-} mutants show altered

transposition frequency. The binding of the transposase to the P element ends is also distinctive in that its specific recognition determinants are located about 100 base pairs internal from the termini of the element.

The binding of the plant Ac transposase is also complex, binding to both the terminal inverted repeats and subterminal repeats. It is not yet known how the actual steps in recognition and end-cleavage occur for this element.

The transposases of many elements are organized such that the end-specific DNA binding region is located in a discrete domain at the N-terminus of the protein. Multiple distinct motifs for DNA binding can also occur within transposases; for example, the MuA transposase contains different specific DNA binding domains for recognition of the ends of the element and an internal enhancer sequence. The crystal structures of the DNA binding domains of several elements have been determined, but no single specific motif or structure has been associated with transposition.

The synapsis of transposon ends is a key step

A key step in recombination is bringing together the ends of the element in the step called synapsis. A requirement for synapsis prior to DNA strand breakage and joining ensures that cleavage and target joining occurs only in the presence of an appropriate substrate: the intact transposable element bounded by its two transposase recognition sites. Multimerization of the recombinase is a key step in synapsis. In some cases this multimerization is shown to involve the specific binding of the transposase DNA binding domain to one end of the element while the catalytic core of this transposase molecule acts on the other end of the transposable element. Such heteromeric interactions through the intertwined assembly of protein–DNA complexes are a key feature in ensuring that recombination acts only at particular sites.

Recognition of the target site

In addition to recognizing the ends of the transposon, the transposase must also recognize the proper target site. The transposase may recognize specific features of the target DNA itself or, in some cases, recognize proteins bound to the target DNA. For example, human immunodeficiency virus (HIV) integrase preferentially recognizes (and directs insertion into) bent DNA while the transposase of the bacterial IS10 element preferentially recognizes a particular consensus sequence at favoured sites of insertion. In other cases, multiple transposon-encoded proteins mediate target site selection. A particularly elaborate use of varying target site selection strategies is that used by Tn7. This element inserts at high frequency into a single specific site called attTn7 in the chromosomes of many bacteria. The transposase is recruited to this site by the specific interaction of a transposon-encoded protein which recognizes a particular sequence in attTn7; this interaction in turn recruits and activates a regulator of transposition, which then interacts with and activates the transposase. Alternatively, the yeast Ty1 and Ty3 elements appear to be recruited to preferential sites of insertion upstream of RNA polymerase III initiation through the interaction with certain host transcription factors. Another example of recognition of a specific kind of target occurs with the Ty5 element, which preferentially inserts near regions of heterochromatin and telomeres. The molecular basis of these interactions between the host target-selecting proteins and transposases is not yet understood.

Chemical steps of transposition

All reactions that have been established *in vitro* have been found to involve metal (Mg^{2+})-dependent breakage and joining steps. Each of these steps can be described as a phosphoryl transfer reaction, the strand breakage step involving water (H_2O) as the attacking nucleophile (i.e. a hydrolysis reaction), and the joining step using the newly exposed 3′ OH at the end of the transposon as the attacking nucleophile (**Figure 5**). All of these phosphoryl transfer reactions appear to involve one-step nucleophilic attacks, i.e. the nucleophile water molecule directly attacks the junctions between the transposon ends and the donor DNA, and the nucleophile 3′ OH end of the transposon directly attacks the target DNA. These recombination reactions do not appear to involve the covalent protein–DNA intermediates used in conservative site-specific recombination reactions such as the integration–excision cycle of bacteriophage lambda. Another interesting feature of at least some transposases is that a single active site of one transposase molecule can actually promote multiple phosphoryl transfer steps. This strategy is probably key to ensuring a high degree of coordination between the breakage and joining steps.

It is now also clear that, in addition to a similar chemistry underlying many transposition reactions (i.e. the central importance of breakage and joining at the 3′ ends of the transposon), these reactions are mediated by structurally related catalytic domains that may share little primary amino acid sequence similarity. Crystallographic studies of the HIV and avian sarcoma virus (ASV) integrases, the phage Mu transposase and the transposase of the bacterial element IS50 reveal that they share a similarly folded metal binding domain in which acidic amino acid residues from dispersed positions along their linear polypeptide chains are clustered together. These signature acidic amino acids are called the DDE motif (D, aspartic acid; E, glutamic acid).

Mutational experiments have demonstrated the critical role of these amino acids in recombination: changing these

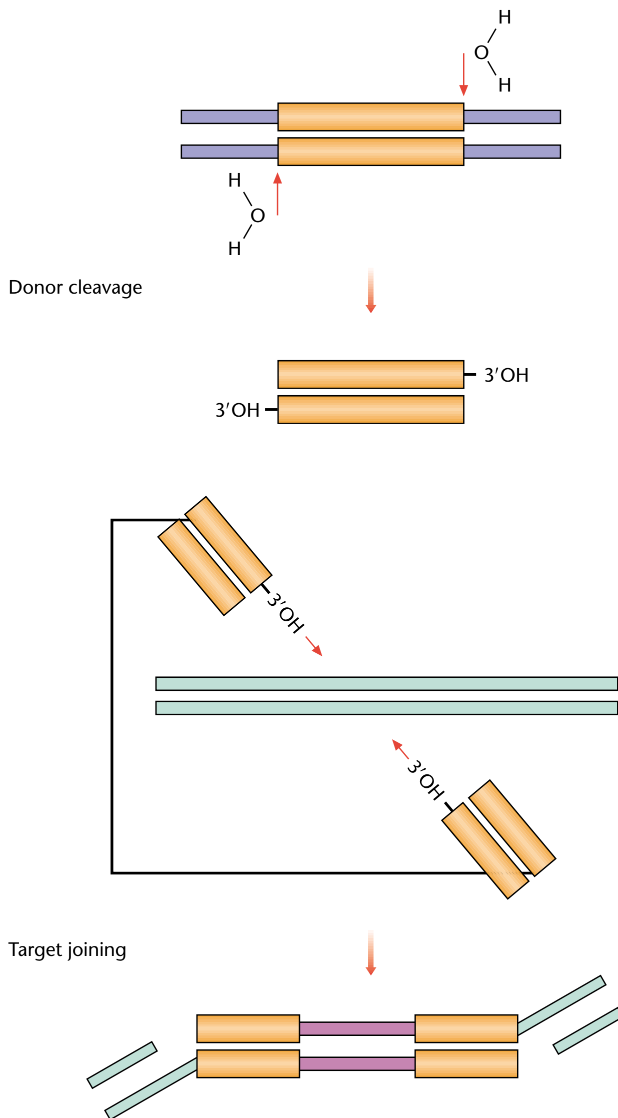

Figure 5 Phosphoryl transfers in transposition.

residues results in catalytic defects. The recombinases of this class are called members of the retroviral integrase superfamily. It is, however, also important to understand that the DDE structure is present in recombinases that share very little amino acid sequence homology. Thus, until it was apparent from crystallographic analysis, structures of the retroviral integrases and the phage Mu transposase were not expected to be related.

It remains to be seen whether all transposases will use related catalytic domains. The sequences of many putative transposases that lack an obvious DDE motif have been reported. The fact that the HIV and phage MuA transposase structures are virtually superimposable in the absence of readily observed sequence homology indicates that protein sequence alone is insufficient to determine whether a particular transposase belongs to this class. More structural studies are needed to elucidate relationships among recombinases.

It is also interesting to note that the structural features that form the metal binding site in transposases and integrases are used in other enzymes involved in a variety of phosphoryl transfer reactions. For example, the same metal binding fold is seen in RuvC, a nuclease that cuts Holliday junctions in homologous DNA recombination, and in ribonuclease H, an enzyme that degrades RNA in RNA–DNA hybrids. Thus, all these enzymes are part of a larger polynucleotidyl transferase family containing similar metal binding motifs.

Repair is required to regenerate intact duplex DNA

Once the chemical steps of transposition have occurred, the short gaps at the newly formed junctions between the transposon ends and the target DNA must be repaired to regenerate intact duplex DNA. This repair is mediated by a host DNA polymerase and ligase. The first step in repair must be to remove the transposition proteins from the recombination product. In the case of Mu, the conformation of the MuA transposase is changed by a host chaperone, and the altered transposase then helps to recruit the repair machinery. It will be interesting to determine whether this strategy for disassembly and repair is used by other elements.

Control of transposition

Transposition is a potentially fatal event to the host. For example, excision of an element from a donor chromosome results in a double-stranded gap in DNA that may not be repaired. Alternatively, the transposon may insert into a host gene encoding an essential function. Thus, the frequency of transposition is generally tightly controlled. In the case of the well-studied IS10 and IS50 transposases, the level of transposase (and hence of transposition) is kept at very low levels by downregulation at virtually every step of gene regulation. Several proteins related to the transposase can affect transposition. In the cases of IS50, the element encodes both a transposase and a colinear transposase inhibitor that is about 50 amino acids shorter than the transposase. The presence of this shorter species inhibits transposition, probably by forming a transposition-inactive heterodimer with transposase. Another example of a shorter form of the transposase acting as a regulator of transposition occurs with the P element of *Drosophila*.

Classes of Transposases and Integrases

Transposable elements and their cognate recombinases can be organized in many different ways based on different determinants, including their origin (bacterial or eukaryotic), the kind of nucleic acid that forms the element (DNA only elements versus elements that have an RNA intermediate), the order and types of strand breakages that occur, etc. The scheme presented in **Table 1** focuses on the nature of the recombinase for some of the elements that have been studied in most detail, and emphasizes the similarities among elements found in different hosts.

Evolutionary Relationships

Although it is clear that transposable elements have profound effects on their hosts' genomes and gene expression, relatively little is know about the evolution of these elements and their recombinases. Have they all derived from a common ancestor? How has horizontal gene transfer influenced the evolution of these elements?

As described above and in **Table 1**, elements from many different sources are related in the chemical steps they promote (i.e. breakage and joining at the 3′ ends of transposons by one-step phosphoryl transfer reactions) and the structure of the catalytic core of the recombinases (i.e. they share the DDE motif).

Phylogenetic analysis of the Tc1/mariner family, which includes some bacterial IS elements and the retroviral-like elements, suggests a common (or several) ancestors. In the case of the retroviral elements, the phylogenetic relationships among the DDE motifs is consistent with the relationships suggested among these elements based on their reverse transcriptase domains. It should be noted, however, that many bacterial elements containing the catalytic DDE signatures do not fit easily within the

Table 1 Classes of transposases and integrases

Retroviral integrase superfamily elements[a]
1. Bacterial IS3 class elements including IS911
2. Bacterial IS4 class elements including IS10 and IS50
3. Bacteriophage Mu
4. Bacterial Tn*552* and Tn*7*
5. Eukaryotic Tc1, Tc3 and Tn*7*
6. Retroviral-like transposable elements that proceed through a DNA→RNA→DNA pathway but lack an extracellular phase
Yeast Ty1
Yeast Ty3
Drosophila copia
Drosophila gypsy
7. Retroviruses that proceed through a DNA→RNA→DNA pathway and have an extracellular phase
HIV
ASV
MoMuLV
Other element families
1. Bacterial Tn*3*
2. hAT elements (plant Ac, *Drosophila* hobo and plant Tam elements)
3. Plant En-like elements
4. Plant Mutator
Retroviral integrase superfamily cousins?
1. VDJ recombinase
2. Ciliate excisionases

[a]All elements in these classes are known to have a DDE catalytic domain.

ASV, avian sarcoma virus; HIV, human immunodeficiency virus; IS, insertion sequence; MoMLV, Moloney murine leukaemia virus; Tn, transposon; VDJ, V(D)J.

 ENCYCLOPEDIA OF LIFE SCIENCES / © 2002 Macmillan Publishers Ltd, Nature Publishing Group / www.els.net

families described in **Table 1**, so that the evolutionary relationships between these elements are unclear.

It has also recently been found that another recombination enzyme, the RAG1 protein of the RAG 1/ RAG 2 recombinase that mediates the deletion reactions underlying the assembly of immunoglobulin and T-cell receptor genes, may be closely related to transposases and integrases. The RAG (recombination activating gene) proteins promote double-strand cleavages in a manner related to that described above for IS10 and IS50: the first step in the Mg^{2+}-dependent reaction is the generation of a nick, and the resulting 3′ OH then attacks the complementary strand resulting in a double-stranded break. These reactions occur by a series of one-step phosphoryl transactions. Moreover, mutation of certain acidic residues blocks recombination, and this recombinase can also promote transposition reactions in the test tube. It will be interesting to see whether the structure of the RAG recombinase is related to transposases and integrases.

Other recombination reactions may also use DDE-like recombinases. The processing of DNA in the somatic macronucleus of ciliates requires many DNA breakage and joining (excision and ligation) reactions to remove DNA within the germline macronucleus. DDE signatures within candidate ciliate recombinase proteins exist. It will be interesting to see whether the recombinases that execute these transposition-like reactions are indeed related to actual transposases.

Learning more about already described mobile DNAs, and the discovery and characterization of new mobile elements, is sure to provide interesting similarities and differences between recombinases and insights into the evolution and biological roles of mobile DNAs.

Further Reading

Beall EL and Rio DC (1996) *Drosphilia* IRBP/Ku p70 corresponds to the mutagen-sensitive mus309 gene and is involved in P-element excision *in vivo*. *Genes and Development* **10**: 921–933.

Beall EL and Rio DC (1998) Transposase makes critical contacts with, and is stimulated by, single-stranded DNA at the P element termini *in vitro*. *EMBO Journal* **17**: 2122–2136.

Becker HA and Kunze R (1997) Maize Activator transposase has a bipartite DNA binding domain that recognizes subterminal sequences and the terminal inverted repeats. *Molecular and General Genetics* **254**: 219–230.

Capy P, Langin T, Higuet D, Maurer P and Bazin C (1997) Do the integrases of LTR-retrotransposons and class II element transposases have a common ancestor? *Genetica* **100**: 63–72.

Craig NL (1995) Unity in transposition reactions. *Science* **270**: 253–254.

Craig NL (1997) Target site selection in transposition. *Annual Review of Biochemistry* **66**: 437–474.

Craig NL, Craigie R, Gellert M and Lambowitz A (eds) (in press) *Mobile DNA II*. Washington, DC: American Society for Microbiology.

Grindley NDF and Leschziner AE (1995) DNA transposition: from a black box to a color monitor. *Cell* **83**: 1063–1066.

Haren L, Ton-Hoang B and Chandler M (1999) Integrating DNA: transposases and retroviral integrases. *Annual Review of Microbiology* **53**: 245–281.

Hickman AB, Li Y, Mathew SV *et al*. (2000) Unexpected structural diversity in DNA recombination: the restriction endonuclease connection. *Molecular Cell* **5**: 1025–1034.

Kennedy AK, Hanniford DB and Mizuuchi K (2000) Single active site catalysis of the successive phosphoryl transfer steps by DNA transposases: insights form phosphorothioate stereoselectivity. *Cell* **101**: 295–305.

Klobutcher LA and Herrick G (1997) Developmental genome reorganization in ciliated protozoa: the transposon link. *Progress in Nucleic Acid Research and Molecular Biology* **56**: 1–62.

Lavoie BD, Shaw GS, Millner A and Chaconas G (1996) Anatomy of a flexer DNA complex inside a higher-order transposition intermediate. *Cell* **85**: 761–771.

Mahillon J and Chandler M (1998) Insertion sequences. *Microbiology and Molecular Biology Review* **62**: 725–774.

Plasterk RH, Izsvak Z and Ivics Z (1999) Resident aliens: the Tc1/mariner superfamily of transposable elements. *Trends in Genetics* **15**: 326–332.

Reznikoff WS, Bhasin A, Davies DR *et al*. (1999) Tn5: a molecular window on transposition. *Biochemical and Biophysical Research Communications* **266**: 729–734.

Roth DB and Craig NL (1998) VDJ recombination: a transposase goes to work. *Cell* **94**: 1–20.

Sakai JS, Kleckner N, Yang X and Guhathakurta A (2000) Tn10 transpososome assembly involves a folded intermediate that must be unfolded for target capture and strand transfer. *EMBO Journal* **19**: 776–785.