DNA Transposition

Genomes are highly stable in a global sense. That is the relative organization or order of genes of a bacterium (E. coli, for example) or that within a eukaryotic chromosome remains the same whether one maps them today or tomorrow or a few months from now. However, within this background of global stability, there are local DNA changes going on constantly. Some of these changes have evolutionary advantages, and can ultimately get fixed in the genome. We already discussed the shuffling of genetic composition mediated by homologous recombination, and directed DNA rearrangements mediated by site-specific recombination. We shall now discuss another class of rearrangements mediated by DNA transposition.

DNA transposition is carried out by elements that can move from one locale to another within a genome or between genomes. Such elements are called transposons. Their movement is mediated by enzymes called transposases. Transposases recognize the specific end sequences that define the limits of a transposable element, and facilitate its movements by cutting DNA these DNA ends (donor sites) and joining them to a target (or recipient) DNA site. Although the donor sites are specific, recognized by a particular transposase, the target sites are quite non-specific (except in a few rare instances). Thus, these elements can end up in a variety of locations on the genome.

The discovery of transposons dates back to early cytogenetic analysis on maize carried out by Dr. Barbara Mclintock in the early 1960s. With unusual insight, she proposed that the puzzling observations on particular phenotypes (or gene expression patterns) that she observed could be explained by mobile genetic elements (now transposons) that move into and/or leave from a genetic locale causing local genetic rearrangements which are manifested as 'unstable' phenotypes. Following this pioneering studies, with the advent of genetic and molecular-genetic and biochemical advances, transposable elements were discovered in all sorts of living creatures, bacteria, fungi, plants and animal cells.

With the advent of the genome sequence technology, it is clear that the genomes of higher living systems (including humans) contain an abundance of repeated DNA elements 'Alu repeats, Line elements, Sine elements etc.), roughly 40 to 50% of the genome. These are vestiges of the movement of DNA elements and their spread that have occurred over evolutionary time scales. Many of these elements are inactive because they have been mutated at the ends required for transposition or

their transposases have been rendered non-functional. Furthermore, the rate of movements of active elements is regulates strictly, and brought down to a level that is not destructive to the continued propagation of the genome. Thus transposable elements provide a significant pool of DNA that can contribute to genome evolution over long periods of time.

As you can imagine, when a transposon moves from its original position to a new one, it can cause a variety of effects. For example, if it brings a promoter of its own, it could turn on the expression an otherwise quiescent gene. Alternatively, it could disrupt the function of a normal promoter and turn off the expression of an active gene. When a transposon lands within a coding sequence, it can disrupt a gene, and cause an 'insertional' mutation.

Some of the simplest transposons encountered in bacteria are called insertion elements or IS sequences. They are relatively small elements, and code for their respective transposases. The transposition function is constituted by two proteins. They do not contain additional genes, and do not harbor a drug resistance marker (such as tetracycline resistance). They are genetically identified by the polar mutation they induce when inserted into a bacterial operon. They not only mutate the gene in which they are inserted but also turn off (or significantly turn down) the transcription of the genes downstream from the point of insertion. This is because they carry sequences that act as transcription termination signals. The apparent effect is as if the downstream genes have been mutated.

There three types of more complex transposons in bacteria. : (A) the composite type of transposons; (B) the Tn3 family of transposons; and (3) certain bacterial viruses that replicate by transposition.

A. **The composite type of transposons. These are transposons that are typically bordered by two insertion sequences. The central element harbors a drug resistance marker such as tetracycline resistance or kanamycin resistance. Examples are Tn5 which carries kanamycin resistance and Tn10 which carries tetracycline resistance. Two modes of transposition are possible. The transposon as whole moves from one point to another on or between genomes. Occasionally, the IS elements at the termini can by themselves transpose.**
B. **The Tn3 family of transposons contains well defined short inverted repeats at their ends. Examples are the closely related transposons Tn3 and gamma-delta ($\gamma\delta$). They also encode a drug resistance marker and the transposase**

protein. In addition, they harbor a resolvase genes that codes for a site-specific recombinase. This protein is responsible for resolving the transposition intermediate and completing the process of movement.

C. **Viral transposons: Bacterial viruses phage Mu and its close relative D108 are in one sense giant transposons. They encode the structural and functional genes that are typical of bacterial viruses. In addition, they harbor a transposition machinery. The virus replicates as it transposes. Thus transposition is the mechanism by which the viral genome is duplicated multiple times during a cycle of phage infection.**

The mechanism of phage Mu transposition is fundamentally similar to the mechanism by which retroviruses integrate into a eukaryotic genome. These viruses are therefore retro-transposons. They include the human immune deficiency virus HIV. The RNA genome of a retrovirus is transcribed inside a cell into a complementary DNA strand, which is then copied into a double stranded form. This cDNA copy enters the nucleus and with the help of a viral integrase protein integrates into the host DNA. The integration sites are quite non-specific. The chemistry of integration follows the same basic chemical steps as the 'transposition' reaction.

We shall discuss the different types of transposition mechanism below. We will deal with the individual steps more thoroughly in class.

DNA TRANSPOSITION DNA cleavage and transfer by partially conservative mechanisms

DNA transposases, retroviral integrases and the RAG1/RAG2 recombinase (that triggers rearrangements to create functional immunoglobin genes) first break specific phosphodiester bonds hydrolytically (the non-conservative step). The resulting 3'hydroxyl groups then attack their target phosphodiester bonds in the conservative step to form joint molecules, which may be further processed in a number of ways. Overall, these reactions are only partially conservative. Comparison with conservative site-specific recombination mechanisms that we already discussed provides a more balanced perspective of the different ways in which enzyme systems have evolved to cut and paste DNA.

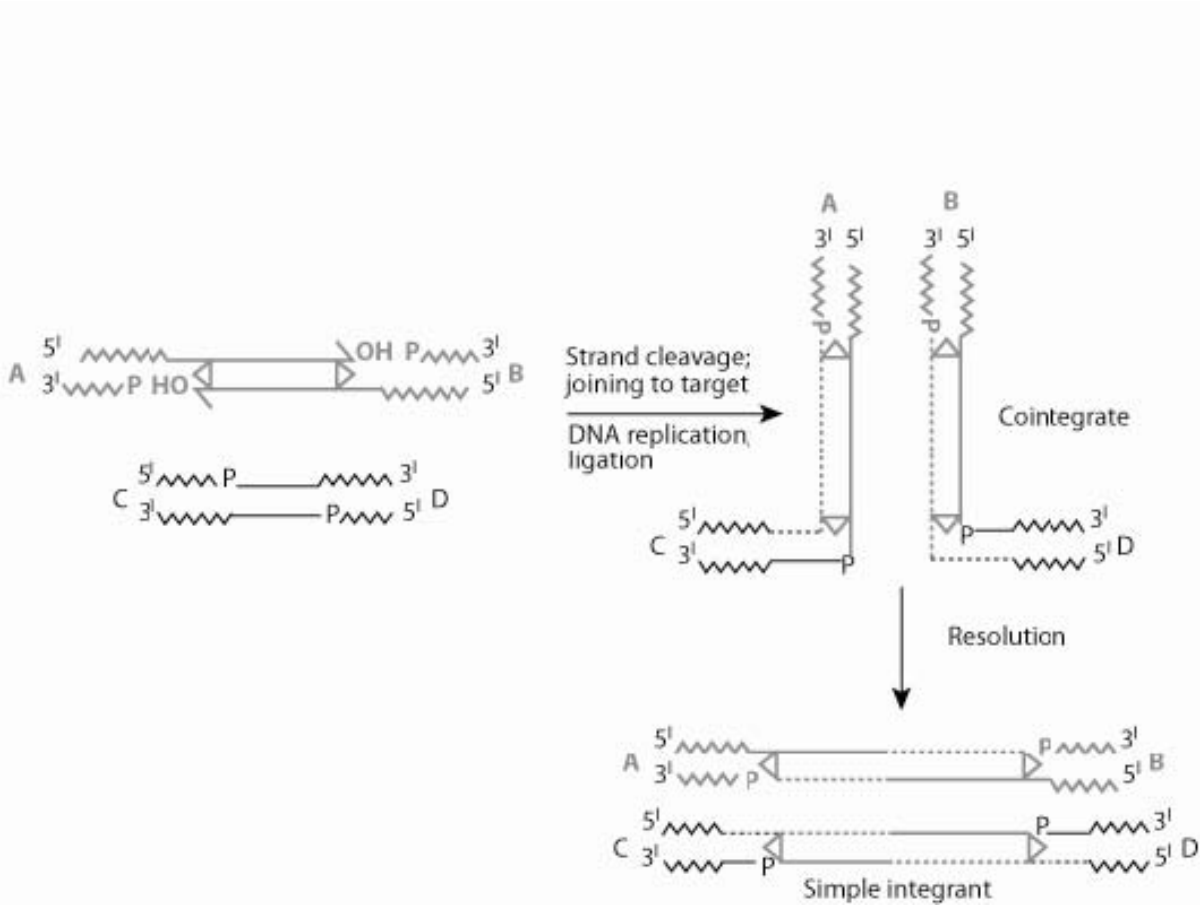

Fig. 1. Strand cleavage and transfer during replicative DNA transposition. Hydrolytic single strand cleavages at the transposon ends (triangles), transfer of the cleaved ends to the target by transesterification and replicative repair and ligation result in the cointegrate. In this transposition intermediate, the transposon and the target sites are duplicated. Recombination within the transposon copies releases the donor DNA, leaving a simple integrant as the end product. DNA sequences flanking the transposon and the target are labeled A, B and C, D, respectively. When the donor and target DNA molecules are circular, the cointegrate is a fusion between the two circles. This can be seen by joining the A and B ends and the C and D ends.

For a more balanced mechanistic perspective on conservative site-specific recombination that we already discussed, a brief comparison with strand transfer reactions that are only partly conservative is helpful. DNA transposases, retroviral integrases and the RAG1/RAG2 recombinase (that triggers rearrangements to create functional immunoglobin genes) first break specific phosphodiester bonds hydrolytically (the non-conservative step). The resulting 3'hydroxyl groups then attack their target phosphodiester bonds in the conservative step to form joint molecules, which may be further processed in a number of ways.

Replicative transposition: cointegrate intermediate

During replicative transposition of a transposon such as phage Mu, the 3'hydroxyls formed at the two ends of the transposon are transferred to two phosphodiesters in the recipient DNA that are staggered by five base pairs (Fig, 1). Extensive DNA synthesis primed by the 3'-hydroxyls formed adjacent to the donor-recipient junction followed by ligation duplicates the transposon and the short target sequence. Assume that this reaction occurs between a circular donor (a plasmid containing the transposon) and a circular recipient (a bacterial chromosome). The net result is the formation of a fusion circle between the donor and recipient called the cointegrate. This is readily seen by joining the A and B ends (as in the donor circle) and the C and D ends (as in the recipient circle). Recombination between the duplicated copies of the transposon, homologous or site-specific, can resolve the cointegrate into a simple integrant plus the donor that still retains a copy of the transposon. The former product is the recipient DNA that has acquired a copy of the transposon with a duplication of the target site (5 bp in the case of Mu) flanking it.

Cut and paste transposition

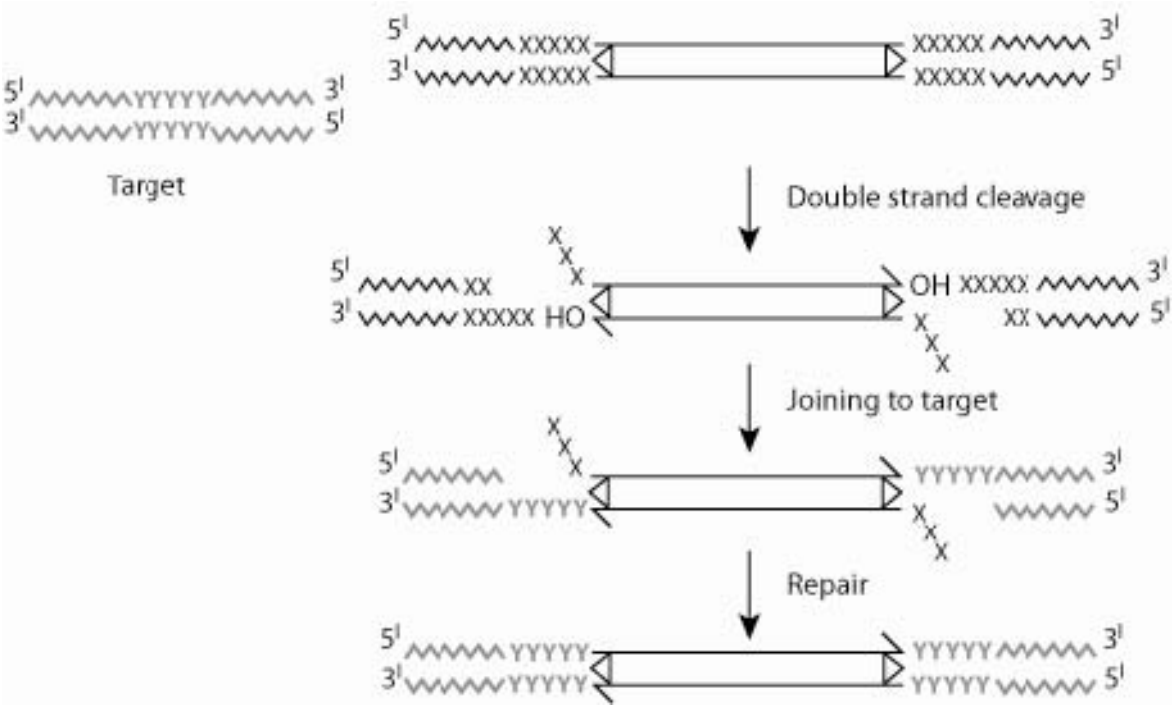

Fig. 2. Double stranded hydrolytic cleavages release the transposon with 3'OH ends and 5' flaps. The 3'-hydroxyl ends are joined to the target DNA by transesterification. The X/X's are base pairs flanking the transposon in the donor molecule; the Y/Y's are base pairs that comprise the target site.

By contrast to phage Mu, transposition of certain mobile DNA elements follows the 'cut and paste' mechanism. For example, Tn7 is excised from the donor site by double stranded cuts, generating 3'-hydroxyls at the transposon ends (the result of transposase cleavage) and 3 nt long 5' flaps (the result of cleavage by a transposon encoded restriction enzyme like nuclease) (XXX in Fig. 2). After strand joining, the short gaps resulting from the stagger of the target phosphodiesters are filled (YYYYY/YYYYY in Fig. 2), the 5' overhanging nucleotides are removed, and free ends are ligated. A variation of this mechanism is employed during retroviral integration. The integrase clips off two terminal nucleotides from the double stranded DNA formed by reverse transcription to produce 3'-hydroxyls at either end flanked by 5' overhangs on the opposite strand. Subsequent events, strand transfer, gap filling and removal of overhangs are analogous to those of Tn7 transposition.

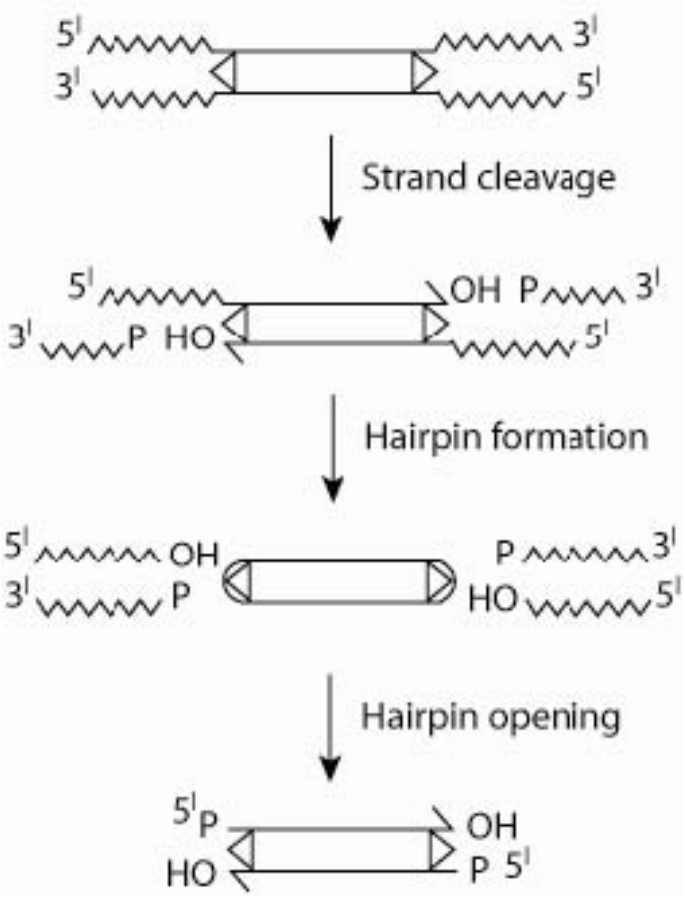

Fig. 3. The transposon is released in three steps. Single strand cleavages (hydrolysis) generate 3'-OH groups. They attack the phosphodiester bond on the opposite strand causing hairpin formation (transesterification). The hairpins are opened by hydrolysis to yield the free transposon.

In transposons Tn5 and Tn10, the 3'-hydroxyls formed at the transposon termini attack the opposing phosphodiester bonds on the uncleaved strand to release the transposon with hairpin termini (Fig. 3). The hairpins are opened hydrolytically to create flush ends prior to strand transfer. During the initiation of V(D)J joining by the RAG1/RAG2 recombinase, exposure of 3'-hydroxyls by cleavage followed by hairpinning occurs at the junction of 'coding' and 'signal' sequences, analogous to the Tn10 reaction (Fig. 4). The hairpins, though, are formed on the coding sides, and the intervening DNA with the signal ends is detached. Here also, hairpins are opened, often in an off-center mode to produce staggered ends, filled in, and joined in a manner that adds diversity at the junction of the rearranged gene segments.

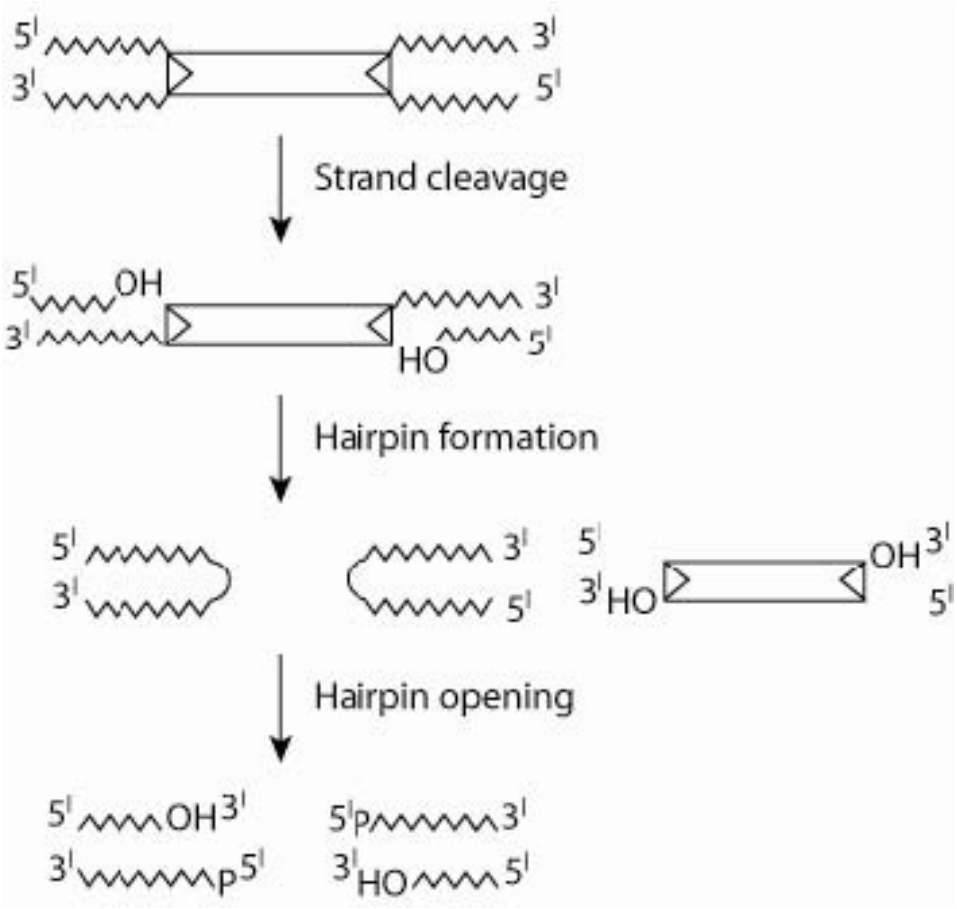

Fig. 4. During V(D)J recombination, similar sequence of reactions as in Fig. 3 releases the internal signal sequence (bordered by the triangles), and forms hairpins at the outside coding ends. The opening of the hairpins and the joining of the coding ends are done in an a manner that generates diversity.