1. Introduction

1.1. Transposons or “jumping genes”

Transposable elements are genetic elements capable of moving from place to place between sites on chromosomes or plasmids. They are distributed across the living world, and play a fundamental role as motors of genome plasticity. Transposons were first discovered in maize (*Zea mays*) by Barbara McClintock in the 1940s (McClintock, 1987). McClintock was studying the genetic consequences of chromosome breakage, and she called the element responsible for this breakage "Dissociation" or Ds. By following changes in the phenotype of the grains of maize, she proved that Ds can “jump” from one place to another in chromosomes. Genetic crosses revealed that the mobilization of such element is dependent on a protein product of another element that she named "Activator", or Ac, which itself was able to move by using its own protein product. This transposable element system is called Ac/Ds, and is considered to be the prototype of *trans*-activating transposons, where the activity of one element is conditional upon functions encoded by another element.

Transposons can be viewed as molecular parasites, that make use of functions of the host cell (DNA and RNA synthesis, energy, the replication apparatus etc.), and only provide the factors that are needed for the recognition of the border between themselves and their host DNA, and for the initiation of events that result in their own propagation throughout the host genome (Sherratt, 1995). The fact that these sequences code only for proteins involved in their own mobilization, suggests they are of a parasitic nature (Sherratt, 1995). In general, transposons move through a recombination process called the transposition. Transposons play a very important role in evolution (Lim et al., 1994) where they are a source of many genetic variations. In addition to their ability to insert into or nearby genes, they are associated with chromosomal rearrangements such as deletions, duplications, and inversions.

Transposable elements are widely spread. They constitute about 22% of the *Drosophila melanogaster* genome (Kapitonov and Jurka, 2003) and about 45% of the human genome (I. H. G. S. C., 2001), so they must be tightly regulated to avoid the accumulation of mutations which could be deleterious to the host (Hartl et al., 1997).

1.2. Types of transposable elements

Transposable elements can be divided into two major groups according to the genetic material utilized as an intermediate in the transposition reaction; Class I transposable elements which use RNA as intermediate and Class II elements which use DNA as intermediate genetic material.

1.2.1. Elements that transpose via an RNA intermediate (Class I)

In yeast, *Drosophila*, and vertebrates the most abundant mobile elements appear to be those which move through an RNA intermediate; these elements are called retroelements. A key component of the life-cycle of retroelements is reverse transcription, which copies the single-stranded RNA of the element into a linear double-stranded DNA, dsDNA, which then integrates into the host genome by the integrase enzyme encoded by the same element. Retroelements are divided into two major groups (Xiong and Eickbush, 1990), the first comprising of Long Terminal Repeat- (LTR) containing retrotransposons and retroviruses. The second group does not contain LTRs and thus they are called non-LTR retrotransposons and retroposons.

1.2.1.1. Retroviruses

Retroviruses are composed of the LTRs flanking a region that contains three genes called *gag*, *pol* and *env*, which encode the capsid core proteins (*gag*), reverse transcriptase (RT), integrase (IN) and protease (PR) (*pol*) and the envelope antigens (*env*). They require the RT to transcribe the RNA intermediate into a cDNA copy which is joined to the host genome by the IN (Hindmarsh and Leis, 1999). The integrated copy is called a provirus, and new infectious viral particles are made from the integrated proviruses by the synthesis of new RNA copies and the packaging of these RNAs into nucleocapsid particles prior to the release of mature retroviruses (Polard and Chandler, 1995).

1.2.1.2. Retrotransposons

Like retroviruses, LTR retrotransposons replicate through reverse transcription of their genomic RNA, and they encode proteins with homology to the Gag and Pol proteins. The main difference is that they do not encode an *env* gene, therefore they are not infectious. According to the phylogenetic comparisons of conserved (mostly RT) protein sequences, retrotransposons are further subdivided into two families: the Ty3-*gypsy* family which has the more familiar arrangement of the *pol*-encoded proteins (PR, RT, IN), and the Ty1-*copia* family which has an inversion in the order of the domains encoded within *pol* (PR, IN, RT). In addition to the similarities between the retroviruses and the retrotransposons in their genomic organization, they also share a similar mechanism of transposition (Adams et al., 1987).

1.2.1.3. Non-LTR retrotransposons and retroposons

Non-LTR retrotransposons represent a large group of elements which have been colonizing the genomes of many hosts, including the human genome. These elements are further subdivided into Long INterspersed Elements (LINEs) and Short INterspersed Elements (SINEs) (Weiner, 2002). LINEs or poly(A) retrotransposons (Eickbush, 1992) have two open reading frames, the first open reading frame codes for a protein with RNA binding ability, and the second codes for an RT protein. SINEs are typified by the human *Alu* elements and mouse B2 elements. *Alu* elements are nonautonomous and are therefore dependent in their mobilization on LINEs. About 20% of the human DNA is estimated to consist of LINE1 (or L1) DNA (Smit, 1996). The non-LTR retrotransposition is primed by exposed 3´-hydroxyl group introduced by a nick in the target site. The RT uses the retroposon RNA as a template for the reverse transcription (Ostertag and Kazazian, 2001).

1.2.2. Elements that move via a DNA intermediate (Class II)

A diverse group of transposable elements relies solely on DNA intermediates without an RNA phase. These are called DNA transposons and they vary in size, structure and complexity, from small, simple insertion sequences (ISs) to more complex composite transposable elements (Mahillon and Chandler, 1998), and bacteriophages. This class includes many eukaryotic transposons in *Drosophila* and higher organisms, including vertebrates. This class of transposable element is characterized by the presence of two inverted repeats flanking a DNA sequence encoding a protein that has recombinase activity called the transposase. The transposase must accurately recognize the inverted repeats. Different transposase subunits, accessory proteins and antibiotic resistance genes can as well be encoded between the ends.

1.2.2.1. Transposable elements in bacteria as an example of DNA transposons

Transposons play a special role in bacterial evolution because of their ability to move between the chromosome and various plasmid and integrated phage DNA. They vary in size, structure, and the way they move. Generally, bacterial transposons can be classified into four groups: the first is the IS sequences, they are normal constituents of bacterial chromosomes and plasmids. They consist of a fairly short (700-1500 bp) DNA segment flanked by 10-40 bp inverted repeat sequences. The DNA segment codes for the transposase protein that catalyses the transposition event (Mahillon and Chandler, 1998). Second, simple transposons are similar to IS elements. They contain DNA segments flanked by short inverted repeat sequences. The DNA segments, however, usually code for a number of gene products in addition to a transposase, and they may contain one or more antibiotic resistance genes, like in the case of Tn7 (Peters and Craig, 2001). Third, composite transposons which are DNA segments that are flanked by IS elements at both ends in opposite orientation. For example, Tn5 (Reznikoff et al., 1999) and Tn10 (Kleckner et al., 1996) are such composite transposons. Fourth, bacteriophage elements, such as the *E. coli* bacteriophage Mu is an unusual phage. It can infect *E. coli* as a normal phage, and then integrates randomly into the host genome by a transpositional mechanism (Craigie, 1996).

Tn5 is a composite bacterial transposon consisting of two IS50 elements in inverted orientation, IS50R and IS50L, flanking a unique DNA sequence encoding three antibiotic resistances (Dodson and Berg, 1989). IS50R is itself an autonomous transposable element encoding two proteins, the transposase protein that is required for the transposition, and an inhibitor protein that is translated from the same open reading frame as the transposase, but lacks 55 amino acid residues at the N-terminus. IS50R and IS50L can be individually mobilized, their transposition requires one outside end and one inside end. Transposition of Tn5 requires two, 19-bp DNA sequences located at the ends of the transposable element. Tn10 has similarity to the overall structure of Tn5: it is also made up of nearly identical copies of IS elements, called IS10R and IS10L, in a mirrored-image orientation flanking a tetracycline resistance gene. The Tn10 transposase is produced from IS10R and is able to mobilize the IS10 sequences as well as the complete, composite element including the two IS10 elements (Kleckner et al., 1996).

1.2.2.2. The Tc1/*mariner* superfamily is a wide spread family of DNA transposons

The Tc1/*mariner* superfamily is exceptionally widespread in living organisms, ranging from protozoa to vertebrates (Hartl et al., 1997). The Tc1/*mariner* superfamily consists of two families: the Tc1 family, the founding member of which was genetically recognized and identified in *Caenorhabiditis elegans* (Emmons et al., 1983), and the *mariner* family, the first member of which was genetically described in *Drosophila mauritiana* (Jacobson and Hartl, 1985 and Marsh, 1986). Based on sequence similarities, the elements of each family can be further divided into subfamilies. Over the years, a variety of related elements have been discovered in several species, and recently in several fish and other vertebrate species. Tc1/*mariner* elements are possibly less dependent on specific host factors than other elements such as the P element of *Drosophila melanogaster* which is known not to transpose outside of the *Drosophila* genus. These elements might therefore be ideal tools for genetic manipulation of many different species (Plasterk et al., 1999). However, the vast majority of naturally occurring Tc1/*mariner*-like transposons are nonfunctional due to inactivating frame shift mutations, small deletions, and internal translational termination codons. In vertebrates, not a single active element has been found. All Tc1/*mariner* elements are about 1300-2400 bp in length and contain a single gene encoding a transposase enzyme which is flanked by terminal inverted repeats. The sequence and the length of the IRs are not conserved among the elements, except for the terminal nucleotides (5´-CAGT) (Radice et al, 1994). A distinct feature for the Tc1/*mariner* elements is the absolute requirement for a TA dinucleotide at the target site (Plasterk, 1999). Integration of these elements is accompanied by a duplication of the TA target site.

1.3. The integrases of LTR-retrotransposons and class II element transposases have a common ancestor

The catalytic domains of retrovirus and retrotransposon integrases, some IS element transposases (Fayet et al., 1990), Tc1/mariner transposases (Doak et al., 1994) and the RAG1 immunoglobulin gene recombinase (Kim et al., 1999) all contain an evolutionarily conserved amino acid triad, the DDE signature (D=aspartic acid and E=glutamic acid). The DDE residues are thought to coordinate a divalent metal ion (probably Mg^{++} *in vivo*) that is a required cofactor for DNA cleavage (Haren et al., 1999). Mutagenic studies with some of the transposable elements belonging to the Tc1/*mariner* (Vos and Plasterk, 1994), retroviral integrase (Bushman et al., 1993), bacteriophage Mu (Baker and Luo, 1994) and RAG1 (Kim et al., 1999) clearly underline the functional importance of these residues. The Tc1 and *mariner* families differ characteristically in the composition of the DDE motif, the E in this conserved motif is D in the mariner transposases (Robertson, 1995). Remarkably, the conservative change from DDD to DDE completely obliterates mariner transposase activity (Lohe et al., 1997). The initial description of the DDE motif indicated that the number of the amino acids between the first two Ds was variable, while the number of amino acids between the second D and the final E was relatively constant at 35, giving the very conserved motif, "DD(35)E".

1.4. *Sleeping Beauty* is kissed back to life

Tc1-like transposons from teleost fish, including zebrafish (*Danio rerio*) (Izsvak et al., 1995), as well as other, closely related elements from nine aditional fish species (Godier and Davidson, 1994; Ivics et al., 1996) are by far the best characterized DNA transposons in vertebrates. Based on molecular phylogenetic studies, the majority of the fish Tc1-like elements can be classified into three major types: zebrafish-, salmonid-, and *Xenopus* Txr-type elements (Izsvak et al, 1995), of which the salmonid subfamily is probably the youngest and thus most recently active (Ivics et al, 1996).

Based on a comparative phylogenetic approach, a transposase gene of the salmonid subfamily of fish elements has been reconstructed; this transposase is capable of catalyzing transposition of an engineered, nonautonomous salmonid element, called the T element, in fish as well as in mammalian cells. This transposon system was named *Sleeping Beauty* (*SB*) (Ivics et al, 1997). The *SB* transposition system is a two-component system, the *cis*-acting sequences carry the SB transposase binding sites within the inverted repeats, and the *trans*-acting transposase encoded by a synthetic gene catalyzes transposition. A transposition assay was developed to measure the transposition frequency *in vivo* (Ivics et al., 1997). The assay (Fig. 1) involves cotransfection of a substrate transposon containing the IRs flanking an antibiotic resistance gene such as *zeo* together with a transposase-expressing helper plasmid into cultured cells. The integrated transposons confer antibiotic-resistant phenotype to the cells. The cells are therefore placed under zeocin selection, and colonies are allowed to form resistant cells which are stained and counted. The ratio between the numbers obtained in the presence versus the absence of transposase is the readout of the assay, and is the measure of the efficiency of the transposition. An interesting finding of Fischer et al. (2001) was that *SB* jumps ~25-fold more efficiently than the Tc1, Tc3 and *mariner* elements in mammalian cell lines. This finding implies either a higher intrinsic transpositional activity of the SB transposase, or more efficient interaction of the SB transposase with cellular factors in vertebrate cells.

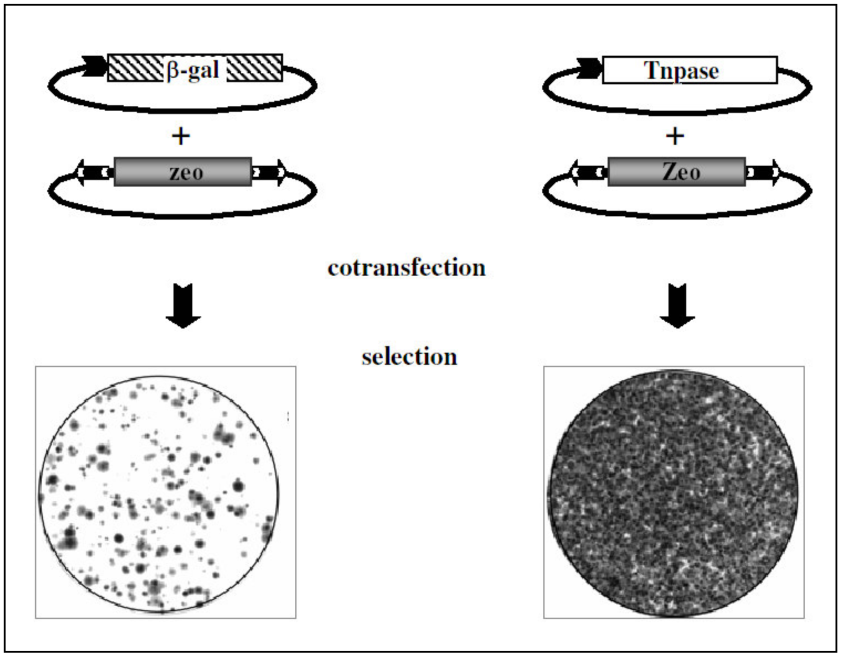
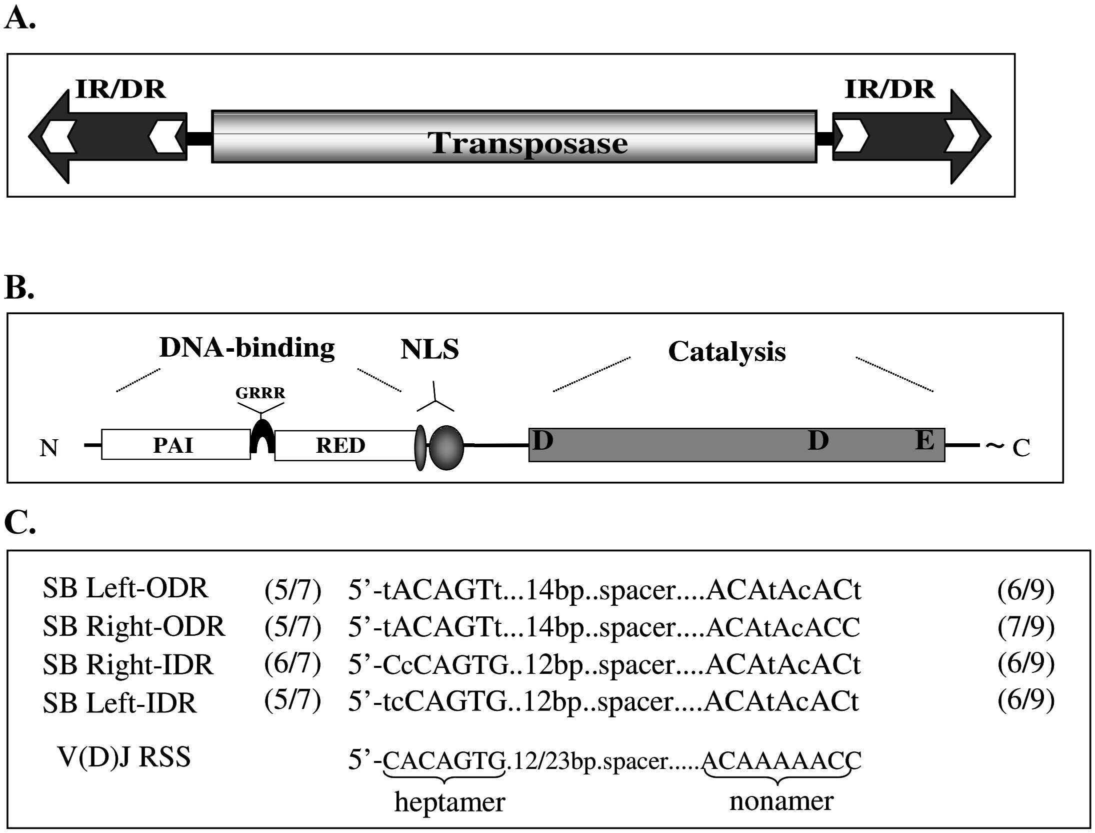


Fig. 1. ***In vivo* transposition assay**. A transposon plasmid carrying a selectable *zeo* gene is cotransfected with a transposase-expressing helper plasmid. In control transfections, a plasmid expressing β-galactosidase is co-transfected. Cells put under zeocin selection. Resistant colonies are stained and counted. Numbers counted in the absence of the transposase is the background of the assay. Numbers counted in the presence of the transposase are higher than the background, due to transposase-mediated transgene integration (transposition). The ratio of the colony numbers in the presence versus in the absence of the transposase is a measure of the transposition efficiency

1.4.1. The structure of the *Sleeping Beauty* (*SB*) transposon system

1.4.1.1. SB transposase structure

The overall structure of transposase domains is predicted to be conserved throughout the Tc1/*mariner* transposon family (Plasterk et al., 1999). Most structure-function analysis was done on the N-terminal region of the transposase, comprising of a DNA-binding domain that is responsible for mediating sequence-specific binding of the transposase to the transposon inverted repeats. Tc1-like transposases, including Sleeping Beauty, contain a bipartite DNA binding domain (Vos and Plasterk, 1994; Izsvak et al., 2002). It has been proposed to consist of two helix-turn-helix (HTH) motifs (Pietrokovski and Henikoff, 1997), similar to the paired domain found in Pax transcription factors (Franz et al., 1994; Ivics et al., 1996). The paired domain consists of two DNA-binding motifs (PAIRED=PAI + RED) each recognizing distinct DNA sequences (Fig. 2B) (Czerny et al., 1993; Izsvak et al., 2002). The PAI sub-domain is the N-terminal HTH of the paired domain, and RED is the C-terminal HTH of the paired domain. Both sub-domains have the ability to bind DNA, but usually the N-terminal PAI sub-domain has more specific DNA recognition.



SB Left-ODR	(5/7)	5’-tACAGTt...14bp..spacer....ACAtAcACt	(6/9)
SB Right-ODR	(5/7)	5’-tACAGTt...14bp..spacer....ACAtAcACC	(7/9)
SB Right-IDR	(6/7)	5’-CcCAGTG..12bp..spacer....ACAtAcACt	(6/9)
SB Left-IDR	(5/7)	5’-tcCAGTG..12bp..spacer....ACAtAcACt	(6/9)
V(D)J RSS		5’-CACAGTG.12/23bp.spacer.....ACAAAAACC	
		heptamer nonamer	

Fig. 2. Schematic representation of the *Sleeping Beauty* transposable element system. A. In nature, the terminal inverted repeats (black arrows) flank a gene encoding the *SB* transposase. The inverted repeats of SB elements have a characteristic structure (IR/DR), and contain two binding sites for the transposase (white arrows) per repeat. **B.** The transposase has an N-terminal, paired-like DNA-binding domain consisting of two helix-turn-helix motifs, PAI and RED, in between there is a GRRR-motif which is an AT-hook. The RED subdomain overlaps with a nuclear localization signal (NLS), which is followed by the catalytic domain responsible for the DNA cleavage and joining reactions and characterized by the conserved DDE signature. The PAI subdomain recognizes the 3’-, whereas the RED subdomain recognizes the 5’-half of the bipartite recognition sequence. **C.** Comparison of Sleeping Beauty transposase binding sites and the RAG1/2 recognition signal sequences. The degrees of similarities to the heptamer and nonamer motifs are indicated.

Between the HTH motifs there is a characteristic GRPR-like (GRRR) sequence (Fig. 2B) which contributes to DNA binding (Izsvak et al, 2002). The GRPR motif is an AT-hook (Plasterk et al., 1999; Izsvak et al, 2002). Changing the GRRR motif to the canonical GRPR motif drops the efficiency of the transposition to ~ 80% of the wild type (my own unpublished results). As shown in Fig. 2B, a nuclear localization signal (NLS) overlaps partially with RED sub-domain. A single amino acid replacement in the NLS is detrimental to the over all function of the mariner transposase (Lohe et al., 1997). The NLS is flanked by phosphorylation target sites of casein kinase II (Ivics et al, 1996). Phosphorylation of these sites may play a role in the regulation of the transposition. The third major domain of the transposase is the catalytic domain which is responsible for the DNA cleavage and joining reactions. This domain, like in the other Tc1/*mariner* elements, has the DDE signature (Fig. 2B). These three acidic amino acids are crucial for *SB* transpositional activity (Ivics et al., 1997). Within the catalytic domain there is glycine-rich region conserved in Tc1-like transposases; its function is unknown.

1.4.1.2. The structure of *SB* transposon

The *SB* transposon is flanked by approximately 230 bp terminal inverted repeats (IRs), which contain binding sites for the transposase. The transposase binding sites of *SB* elements are repeated twice per IR in a direct orientation (DRs) (Fig. 2A) (Ivics et al., 1997). This special organization of inverted repeat, termed IR/DR, is an evolutionarily conserved feature of a group of Tc1-like transposons, but not that of the Tc1 element itself (Plasterk et al, 1999). Tc1 and *mariner* elements are the simplest and have a single binding site per IR. Tc3 elements contain two binding sites, each is 33 bp, but the internal binding site plays no major role in the transposition reaction, since deleting it does not affect the transposition frequency (Fischer et al., 1999). In contrast, all four binding sites within the IR/DR structure are required for *SB* transposition (Izsvak et al., 2000). The four binding sites are not identical, the outer ones are longer by two base pairs. In addition to the DRs, the left inverted repeat of SB contains a transpositional enhancer-like sequence, termed HDR (Izsvak et al., 2002).

1.4.2. The mechanism of *SB* transposition

1.4.2.1 Unity of the transposition reactions

Transposition of an element is a recombination reaction involving three separate sites: the two transposon ends and the new target locus. The mechanism of transposition appears to be essentially the same for most DDE-containing transposases and integrases (Craig, 1995). The general process of transposition involves two sequential steps, site-specific cleavage of the transposon ends, and strand transfer. These two critical reaction steps are catalyzed by one or more transposase molecules that bind in a sequence specific manner to the transposon ends, and assemble in a stable nucleoprotein complex, called the synaptic complex (Craig, 1995). First, a pair of site-specific cleavages is made at the host-transposon DNA boundary, exposing a 3´-hydroxyl (OH) group. In some reactions, including transposition of the Tn5 and Tn10 elements and RAG-mediated V(D)J recombination, this single-strand nick is converted into a double-strand break by a transesterification reaction. This process generates a hairpin on the transposon DNA (in Tn5 and Tn10 transposition) or on the flanking DNA (in V(D)J recombination), which has to be resolved by a second hydrolysis reaction to expose the 3'-OH again for the subsequent strand transfer. Strand transfer is carried out by covalently joining the 3´-OH ends of the excised element to the 5´-ends of the target DNA. These steps are referred to here as the transposition reaction, although an additional step also takes place in host cells, where the strand transfer product is repaired and/or replicated by host proteins. This repair step results in the duplication of the target site. The length of the target site duplication depends on the staggered cut made during strand transfer. Cellular repair processes the gap left behind the excised transposon, which leads to a characteristic footprint of 2 bp or 3 bp in the Tc1/*mariner* elements (Plasterk et al., 1999).

1.4.2.2. The mechanism of *SB* transposition

SB follows the same cut-and-paste mechanism of DNA transposition as described above. The transposition process can arbitrarily be divided into at least four major steps: 1) binding of the transposase to its sites within the transposon IRs; 2) formation of a synaptic complex in which the two ends of the elements are paired and held together by transposase subunits; 3) excision from the donor site; 4) reintegration at a target site. *SB* transposition is initiated by staggered cleavages at each end of the transposable element. The entire element is thereby released from the donor molecule leaving behind 3-nucleotide-long 3´-overhangs at the excision locus (Fig. 3) (Luo et al., 1998). The liberated transposon fragment carries 3-nucleotide-long 3´-overhangs with reactive OH radicals. The excised element integrates exclusively into TA dinucleotide sites at the target DNA (Ivics et al., 1999), with a preference for certain sequences flanking the TA dinucleotide (Vigdal et al., 2002). During integration, another staggered double-stranded DNA break is introduced by nucleophilic attack of the exposed 3´-OH groups at the transposon ends which, through a one-step phosophryl transfer reaction, creates a covalent bond between the 3´-end of the transposon and the 5´-end of the target DNA. After integration of the *SB* transposon in a TA sequence, the four-nucleotide-long single-strand gaps flanking each end of the transposon will be repaired by the cellular machinery to produce a complete *SB* transposon flanked by a duplicated TA sequence (Fig. 3). Repair at the excision site generates 3-bp footprints (Fig. 3) (Luo et al., 1998).

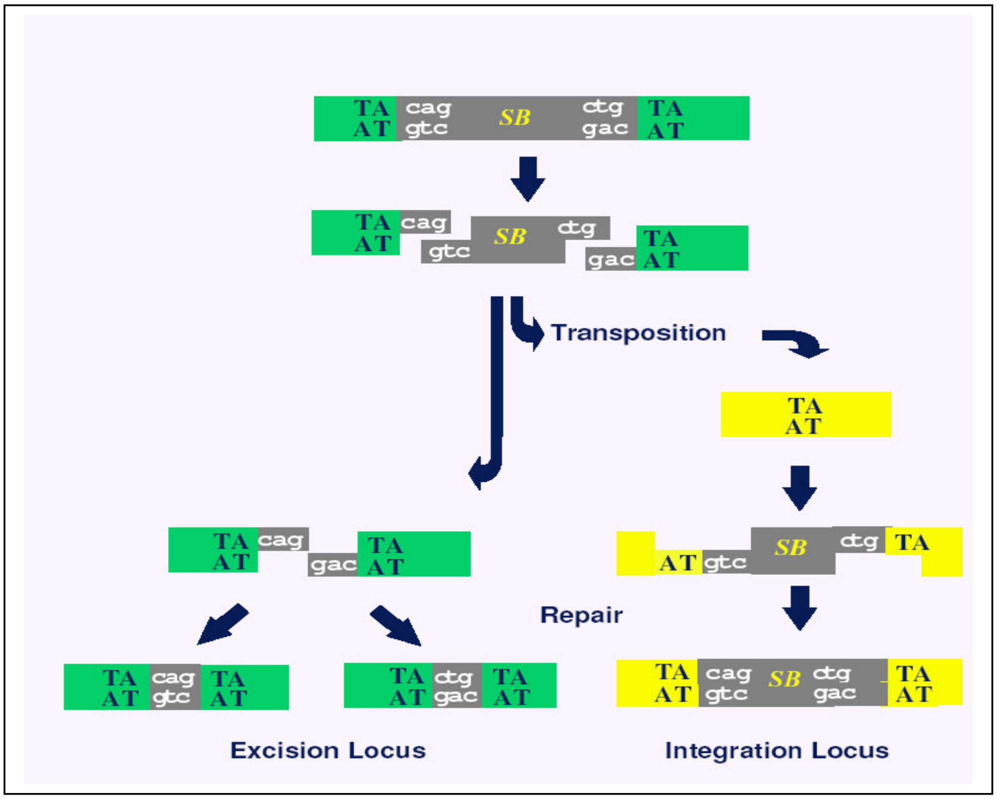


Fig.3. The mechanism of *Sleeping Beauty* transposition. *SB* is excised by double strand breaks at both transposon ends introduced by the transposase. The DNA cut is staggered with three overhanging nucleotides. Three nucleotides are left behind of the transposon at the site of excision. The excised transposon integrates into a TA dinucleotide sequence in the target DNA. During integration the incoming transposon acts as a nucleophile leading to the introduction of a staggered double strand break at the TA target site. DNA after integration is repaired by the cellular repair machinery resulting in the duplication of the target site. The excision site is also repaired forming a distinct, 3-bp footprint (Luo et al., 1998).

1.4.3. Genetic applications of the *SB* transposon system

Transposons can be used to bring new phenotypes into genomes by disrupting functional genes by insertional mutagenesis (loss-of-function mutation), as well as transgenesis (gain-of-function mutation). DNA transposons are used frequently for gene transfer and insertional mutagenesis in non-vertebrate model systems such as *D. melanogaster* (Spradling et al., 1995), and a large number of plant species (Haring et al., 1991). Bacterial transposons, such as Tn10 (Yazgan et al., 2001) and phage Mu (Lamberg et al., 2002) have also been widely used for the identification of genes. However, transposons have not been used for the investigation of vertebrate genomes for two reasons: firstly, until now, there have not been any well-defined, DNA-based mobile elements in these species. Secondly, elements from invertebrates are either inactive or show only limited activity in vertebrate cells (Rio et al., 1988). On the contrary, the *SB* transposable element, like the other Tc1/*mariner* transposable elements, is a promiscuous transposable element without severe host restriction (Plasterk et al., 1999).

1.4.3.1. *SB* as a potential tool for insertional mutagenesis

Insertional mutagenesis is an important tool for studies of gene identity. After completing the sequence of the human genome, which will be soon followed by the zebrafish and *Xenopus* genome sequences, the need for the development of tools for genome-wide insertional mutagenesis has become a demanding issue for functional genomic analysis. Any transposable element that can be mobilized at significant frequencies under laboratory conditions can be a useful mutagenic agent (Sherratt, 1995). Insertional mutagenesis and gene trapping with plasmid or retroviral vectors has been successfully applied in mammalian cultured cells (Zambrowicz *et al.*, 1998; Wiles *et al.*, 2000). The efficiency of plasmid vectors is limited by low insertion rates while retroviral vectors have a preference to integrate into the 5´-regions of actively transcribed genes (Scherdin et al., 1990). Transposable elements present an attractive alternative.

SB could be an ideal element as an insertional mutagen over the currently used retroviral vectors (Luo et al., 1998). It has relatively high transposition frequency, and shows fairly random integration in genomic DNA (Vigdal et al., 2002). TA dinucleotide sequences, the target sites for *SB* transposition, occur approximately once every 20 bp in vertebrate genomes. Thus, a large fraction of genes are expected to serve as suitable targets of *SB* transposition, and can therefore be inactivated by transposon-mediated insertional mutagenesis.

1.4.3.2. Potential application of *SB* in gene therapy

Gene therapy can be broadly defined as the transfer of genetic material to cure a disease or at least to improve the clinical status of a patient. Two basic goals of gene therapy, firstly, is to introduce a transgene with high efficiency into specific cells of a patient whose corresponding gene is defective or missing, secondly, the transgene must have a sustained and regulated expression. A variety of recombinant viruses have been explored for the application of gene therapy, none of them is providing a perfect solution. Of these, retroviral vectors were among the first and the most extensively utilized vector system. The major problems of retroviral vectors are their inability to infect non-dividing cells and difficulties of preparing high titer virus stocks (Romano et al., 2000).

The last decade has witnessed a substantial progress in the development and application of non-viral vectors in gene therapy. The simplest form of non-viral vectors is naked DNA which can be used either alone (Yang et al., 2001) or in conjunction with a variety of molecular conjugates, such as liposomes, polymers, and polypeptides (Li and Huang, 2000). A limitation of this approach is the absence of life-long gene expression. The Tc1/*mariner* elements are promiscuous and have been successfully used as transgene vectors in species other than their original hosts (Plasterk et al., 1999). Studies have shown that *SB* can insert foreign genes in cultured vertebrate cell lines including mouse embryonic stem (ES) cells (Luo et al., 1998), human cells (Ivics et al., 1997) and cells of many different mammalian, frog and fish species (Izsvak et al., 2000). Yant et al. (2000) has provided evidence that the *SB* transposon system could be a useful tool for gene therapy applications. They showed that the SB transposase can efficiently insert transposon DNA into the mouse liver genome with an approximately 5-6% efficiency of transformation. This transformation efficiency was similar to that obtained with integrating viral vectors such as lentiviruses (Kafri et al., 1997) and recombinant adeno-associated viruses (Xiao et al., 1998). Moreover, the *SB* transposon can efficiently insert transgenes into the mammalian chromosomes which can be expressed for a prolonged period of time, since transgenic mice generated with *SB* transposon containing the human alpha-1-antitrypsin (hAAT) cDNA expressed hAAT in their blood for more than 6 months (Yant et al., 2000). Stable genomic integration and expression of an *SB* vector carrying the human factor IX (FIX) expression cassette resulted in a partial correction of the bleeding disorder of the haemophilic mice (Yant et al., 2000), and sustained production of biologically active FIX at levels which would convert a severely affected patient with haemophilia B to one with a much milder phenotype. Another example of *in vivo* gene therapy application of *SB* transposon system is the phenotypic correction of murine hereditary tyrosinemia type 1 (HT1). Deficiency in fumarylacetoacetate hydrolase (FAH), the enzyme that catalyzes the last step in tyrosine degradation and its deficiency is responsible for the HT1 phenotype, was corrected in 62% of FAH-deficient mice when they received the FAH-expression *SB* transposon construct (Montini et al., 2002).

In conclusion, *SB* has some advantages over the currently used viral and non-viral vectors, it mediates stable, single copy integration with the use of simple plasmid DNA and the integrated transgenes show long-term expression through many generations of transgenic cells and organisms.

1.4.3.3 The potential of mutational analysis in the enhancement of transpositional activity

In vitro evolution of genes can create versions with a higher functional capacity than the wild type protein product (e.g. enzymes). The newly evolved function could be due to a major change in the protein structure or a slight selective modification. The beta-lactamase (ampicillin resistance) gene is a good example of an enhancement of the ampicillin-resistant activity (Yano and kagamiyama, 2001). This gene was enhanced by about 50 rounds of directed evolution by using DNA shuffling, which eliminates the negative mutations and leads to the accumulation of beneficial mutations. The evolved genes gave more active products that not only resist ampicillin, but also other drugs that inhibit bacterial cell-wall synthesis. Transposases can also be selected *in vitro* for hyperactivity. In Tn5, hyperactive phenotypes are due to either the reduction of the self-inhibitory activity of intact Tn5 transposase (Wiegand et al., 1992), or to reduced affinity of the co-translated inhibitor protein to the transposase (Weinreich et al., 1994), or to increasing the binding affinity of the transposase to its binding sites within the transposon inverted repeats (Zhou and Reznikoff, 1997). The combination of these three hyperactive mutants yields a synergistic effect, leading to an extraordinary active transposase (Goryshin et al., 1998).

Some of the hyperactive Tn5 mutants were due to amino acid replacements that change glutamic acid (E) residues to lysine (K) (Zhou and Reznikoff, 1997). This shift from acidic to basic amino acid could make a more favorable interaction with the negatively charged DNA backbone. E to K mutations also led to hyperactive transposase versions of Himar 1 (Lampe et al., 1999). This hyperactive Himar 1 transposase showed hyperactivity in *E. coli* (Lampe et al., 1999) as well as in mammalian cell lines (Fischer et al., 2001). Mutation of Tn10 transposase led to versions of the Tn10 transposase which have a hypernicking activity *in vitro* (Kleckner et al., 1996). Other Tn10 transposase mutations increase the binding affinity to mutant termini of its own substrate (Sakai and Kleckner, 1996). Altering the secondary structure of the Tn5 transposase by introducing a proline residue, a secondary structure breaker, at a defined site in the transposase resulted in a hyperactive version of the transposase. Hyperactivity in this case was the result of elimination or reduction of the interference between the C-terminal domain and the N-terminal domain of the transposase (Davies et al., 2000). Changing a serine amino acid residue in the P-element transposase to alanine led to a hyperactive version (Beall et al., 2002). In this version, P-element transposase is thought to escape negative regulation by phosphorylation, which normally would occur at the serine. Thus, there are numerous examples for specific amino acid changes that lead to hyperactive phenotypes of transposases. Presumably, the SB transposase can also be mutated to higher activity.

Mutational analysis of the transposon ends showed the absolute requirement for several base-pairs for transposition, probably because these sequences are in direct contact with the transposase (Vos et al., 1993; Van Luenen et al., 1993; Izsvak et al., 2000). In Tn5, mutated end sequence of the transposon substrate led to a hyperactive phenotype of the transposon which significantly enhances the transposition efficiency (Steiniger-White et al., 2002). The combination of four mutations in the *SB* transposon ends was shown to elevate the transposition frequency about 4-fold compared to wild-type (Cui et al., 2002).

The transposition frequency of *SB* decreases with increasing the length of the transposon: each kb increase in transposon length decreases the transposition efficiency by approximately 30% (Izsvak et al., 2000). However, small *mariner* elements can be used successfully in transformation of large pieces (13.2 kb) of DNA into the germ line of *D. melanogaster* (Lidholm et al., 1993). An unusual transposon consisting of two identical copies of *Paris* elements flanking a relatively large piece of DNA (longer than 10kb) has been mobilized in *Drosophila virilis.* (Petrov et al., 1995). *Paris* is a Tc1/*mariner*-type transposon, indicating that mimicking such construct by arranging two, complete *Sleeping Beauty* transposons flanking a relatively large piece of DNA in an inverted orientation could possibly extend the capacity of *SB* vectors to transpose large DNAs.

In summary, the *SB* transposon system could be improved either by enhancing the recominational activity of the transposase or by altering transposon composition. A further possibility is the identification (and manipulation) of host factors that play a positive role in the transposition process, which is described in the next section.

1.4.3.4. Effect of host factors on transposition

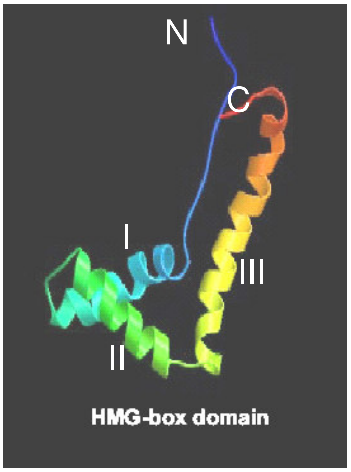
Most transposons require accessory host factors for optimal activity (Sherrat, 1995). Host factors can affect transposition in at least six ways: 1) by modulating the activity of the transposase through post-translational modifications, or through protein-protein interactions; 2) by modulating the structure of the transposon DNA; 3) by modulating synaptic complex assembly; 4) by modulating the structure and accessibility of the target DNA site, 5) by controlling DNA conformation at the donor site, and 6) by affecting other factors which may be directly involved in the control of the transposition reaction. These effects are generally specific for each element. Some of the best documented examples include factors with DNA chaperone activity. The DNA chaperones may play roles in ensuring the correct three-dimensional architecture of various nucleoprotein complexes necessary for effective transposition (Mahillon and Chandler, 1998). They may also be involved in regulating transposase expression. Integration host factor (IHF), and HU are closely related histone-like DNA bending proteins that are widespread in prokaryotes. IHF binds to a defined consensus DNA sequence while HU binds to DNA non-specifically (Gossen and Van de Putte, 1995). In the cell, they serve as architectural factors in many cellular processes, such as transcription, replication, and site-specific recombination helping to bring distant DNA sites closer to each other by virtue of their ability to introduce sharp bends into the DNA. For example, IHF was found to enhance the Tn10 synaptic complex assembly (Sakai et al., 1995). The transposase binding sites of bacteriophage Mu are brought together by the bending action of the *E. coli* HU protein (Lavoie et al., 1990). Hin recombinase-mediated recombination and bacteriophage-lambda integration are strongly stimulated by HU (Haykinson and Johnson, 1993) and IHF (Goodman and Nash, 1989), respectively. The eukaryotic high mobility group

(HMG) proteins can functionally replace HU and IHF in some recombination reactions, indicating some level of exchangeability between these DNA-bending proteins (Segall et al., 1994). All of these DNA-bending proteins are believed to assist recombinational mechanisms by facilitating the formation of active recombinase-DNA complexes (Paull et al., 1993; Bustin, 1999).

1.4.3.4.1. High Mobility Group proteins

High Mobility Group (HMG) proteins are a class of non-histone nuclear proteins, which share the common property of rapid migration in PAGE gels. They are represented in all vertebrates, and ubiquitous in mammalian cells. The first isolated members were HMG-1/-2. Additional members were subsequently isolated and numbered according to their locations on an electrophoretic gel, the HMG-14/-17 proteins were initially distinguished from the HMG-1/-2 proteins by their lower molecular weights of 10-12 kd, compared to the 28 and 27 kd weights of HMG-1/-2, respectively (Einck and Bustin, 1985). Two additional low molecular weight (10-12 kd) "HMG-like" proteins were isolated from human cells (Lund,et al., 1983), and named HMG-I/-Y, considered to be isoforms of the same protein (Lehn et al., 1988). In the narrowest traditional sense, the HMG protein family consists of six proteins, and can be classified into three subfamilies: HMG1/2, HMGI/Y, and HMG-14/-17 (Bustin 1999), these proteins have recently been renamed (Bustin, 2001); HMGB1/2, HMGA1a/b, and HMGN1/2; respectively. The three HMG subfamilies share many physical characteristics (Bustin and Reeves, 1996), but differ in their main functional motifs (Thomas and Travers, 2001). These functional sequence motifs are the main site of interaction between the HMG proteins and DNA or chromatin targets. The binding domain of HMGB1 and its close relative HMGB2 is referred to as the HMG-box. The HMG-box has a globally conserved fold (Bustin, 1999; Thomas, 2001) taking a characteristic, twisted, L-shape made of three alpha-helices (Fig. 4). The long arm comprises the N-terminal extended strand and helix III, and the short arm comprises helices I and II.

Fig. 4. ***HMG-box domain.*** The three-dimentional structure of the canonical HMG-box is formed from three α-helices, helix I,II, and III. This fold is globally conserved between the HMG-box proteins taking an L-shaped structure.



HMGB1/2 are composed of two HMG-boxes which are adopting homologous folded domains of about ~80 amino acid residues which mediate DNA binding (Bustin, 1999; Thomas, 2001), and a long polyacidic C-terminal tail, separated by short linker sequences. The tail is the longest in HMGB1 (30 aspartic and glutamic acid residues), and shorter in HMGB2 (20 acidic residues) (Bianchi et al., 1992). The C-terminal tail domain is not required for DNA binding. The HMG-box is a very conserved functional domain for many transcription factors, as well as for many other sequence-specific DNA-binding proteins (Bustin, 1999) throughout the four eukaryotic kingdoms. The HMG-box domain binds DNA through its concave face (the region encompassing the arms of the L), and partially intercalates into the minor groove through a hydrophobic wedge close to the N-terminus of helix I. This opens up the minor groove, thus increasing the protein-DNA interaction face, and widens the minor groove on the account of the opposing major groove which is thereby compressed (Thomas and Travers, 2001). HMGA1 bind DNA through the minor groove with a special motif termed AT-hook motif, this motif binds DNA through the minor groove of AT-rich DNA sequences (Aravind and Landsman, 1998).

1.4.3.4.2. Functional roles of HMG proteins

Interactions between DNA and proteins are fundamental in biological processes, including DNA packaging, recombination, repair, transcription, and replication. Assembly of nucleoprotein complexes is needed to overcome the axial rigidity of DNA that could be achieved by bending the DNA to facilitate nucleoprotein complex formation. HMG proteins are known to be involved in such processes. HMGA1 proteins are of special interest, because they exhibit specific binding to the minor groove of B-form DNA, thereby stabilizing the B-form and facilitating the binding of transcription factors in the opposing major groove

(Bustin, 1999, Aravind and Landsman, 1998). They function as architectural proteins to modulate the expression of tumor necrosis factor beta, Interleukin 2 receptor alpha, Interleukin 4 and other genes (Bustin and Reeves, 1996). HMGB1 is an abundant (~10^6 molecules/cell), non-histone, nuclear protein associated with eukaryotic chromatin (Bustin, 1999). Through the HMG-box, HMGB1 binds DNA in a sequence-independent manner, but with preference for certain DNA structures including four-way junctions and severely undertwisted DNA (Thomas and Travers, 2001; Bianchi et al., 1989; Bianchi et al., 1992; Pohler et al., 1998). HMGB1 has low affinity to B-form DNA, and is thought to be recruited by other DNA-binding proteins through protein-protein interactions, and to induce a local distortion of the DNA upon binding. The ability of HMGB1/2 proteins to bend DNA was demonstrated *in vitro* (Thomas and Travers, 2001). These proteins facilitate self-ligation of short DNA fragments (Pil et al, 1993; Stros, 1998), and can bridge linear DNA fragments thereby enhancing multimerization of longer DNAs (Stros et al., 2000). Together with the closely related HMGB2 protein, HMGB1 has been implicated in a number of eukaryotic cellular processes including gene regulation, DNA replication, and recombination (Bustin, 1999; Muller et al., 2001). HMGB1/2 directly interact with a number of proteins, including some HOX (Zappavigna et al., 1996) and POU domain (Zwilling et al., 1995) transcription factors, and the TATA-binding protein (Sutrias-Grau et al., 1999), and facilitate their binding through protein-protein interactions. Transient overexpression of HMGB1 enhances the transcriptional activity of Hox protein and steroid hormone receptors, as well as V(D)J recombination in transfection assays, even in cell lines which express abundant amounts of HMGB1 (Zappavigna et al. 1996; Boonyaratanakornkit et al. 1998; Aidinis et al., 1999).

V(D)J recombination is a site-specific recombination process through which the variable antigen receptor gene segments are assembled (Fugmann et al., 2000; Gellert, 2002). Each antigen receptor segment is flanked by highly conserved recombination signal sequences (RSS). The RSS is made of seven nucleotides (the heptamer) which is always contiguous to the coding sequence, followed by a spacer which is either 12 or 23 base pairs, followed by a second conserved sequence of nine nucleotides (the nonamer) (Fig. 2C). Depending on the length of the spacer, these sequences are called 12RSS and 23RSS signals. Efficient recombination occurs only between 12RSS and 23RSS signals, a restriction termed the 12/23 rule (Gellert, 2002). The RAG1/2 recombinase complex binds specifically to the heptamer and nonamer of the RSS, and this binding is enhanced in the presence of HMGB1/2. RAG1/2 probably recruits HMGB1/2 to the RSS through interactions with homeodomain region of

RAG1 (Aidinis et al., 1999). The enhancement of binding is more pronounced in the 23RSS signal (Van Gent et al., 1997), probably due to the bending ability of HMGB1 of 23RSS signal but not of the 12RSS signal (Van Gent et al., 1997). Subsequently, the two RSS signals are bridged in an active synaptic complex (Van Gent et al., 1996) where RAG1/2 cleaves the DNA at the border of coding DNA and the heptamer in a process enhanced in the presence of HMGB1/2 (van Gent et al., 1997).

It is worth noting that the RAG1/2 recombinase can act as a transposase *in vitro* by the help of either HMGB1 or HMGB2 (Agrawal et al., 1998; Hiom et al., 1998). HMGB1 was found to promote Rep protein-mediated site-specific cleavage of adeno associated virus (AAV) DNA (Costello et al., 1997). The production of retroviral cDNA does not require an excision step, but the downstream events of retroviral integration are highly similar to other transpositional reactions (Costello et al., 1997). Interestingly, HMGA1 family members, but not HMGB1/2, are required for retroviral cDNA integration (Hindmarsh et al., 1999; Li et al, 2000). Both V(D)J recombination and retroviral integration have common features with *SB* transposition. RAG-mediated cleavage at the ends of recombination signal sequences (RSS) in V(D)J recombination is probably analogous to the excision step of transposition, whereas the biochemical steps leading to insertion of signal molecules, retrovirus integration and DNA transposition are essentially the same (Craig, 1995).

SB mediates transposition in a variety of vertebrate species (Plasterk et al., 1999), and is more active than other members of the Tc1/*mariner* family (Fischer et al., 2001). Because there is substantial interest in developing transposon technology for gene therapy (Yant et al., 2000) and gene discovery (Fischer et al., 2001), it is of importance to dissect the molecular mechanisms involved in transposition and its regulation. In this work, HMG proteins were evaluated as cellular host factors of *Sleeping Beauty* transposition in mammalian cells. HMGB1 was found to be required for efficient *SB* transposition. *SB* transposition was significantly reduced in HMGB1-deficient mouse cells. This effect was fully complemented by expressing HMGB1, partially by expressing HMGB2, but not with HMGA1 protein. Interestingly, transient overexpression of HMGB1 in wild-type mouse cells enhanced transposition, indicating that HMGB1 is a limiting factor of transposition. *SB* transposase was found to interact with HMGB1 *in vivo*, suggesting that the transposase may actively recruit HMGB1 to transposon DNA via protein-protein interactions. HMGB1 enhanced preferential binding of the SB transposase to the internal transposase binding sites within the transposon inverted repeats, and promoted bending of DNA fragments comprising the transposon inverted repeats. These data are consistent with a role of HMGB1 in synaptic complex formation in transposition.

I also present in this thesis experimental data showing that hyperactive versions of the SB transposase can be selected, and show over 3-fold increase in transpositional activity. The transposon DNA has been modified by constructing a sandwich transposon, mimicking the structure of a naturally occurring *Paris* element. This transposon is able to jump 3-fold more efficiently than similar size marker genes in wild-type *SB* vectors.