

Table of Contents

Hsmar1 Homo sapiens mariner type 1.....	2
1 INTRODUCTION.....	4
1.1 Discovery of transposable elements	4
1.2.1 DNA elements.....	6
1.2.2 RNA elements.....	7
1.5 The Tc1/mariner superfamily of transposable elements: how they live and how they die.....	15
1.5.1.1.2 The transposase.....	19
1.5.1.1.2.1 DNA recognition.....	19
1.5.1.2 The molecular mechanism of Sleeping Beauty transposition.....	22
1.5.1.2.3 The biochemistry of cut-and-paste transposition.....	26
1.7.1 Insertional mutagenesis.....	32
1.7.2 Transgenesis.....	39
1.7.3 Sleeping Beauty: an integrating, nonviral gene delivery vector for gene therapy	40
1.7.4.1.1 Site-selectivity in viral integration.....	45
1.7.4.1.3 Site-specific transposable elements.....	48
1.7.5.1 Targeting through fusion to DNA binding domains.....	52
1.7.5.2 Targeting through interaction with DNA-binding proteins.....	56
5 CONCLUSIONS.....	90
6 SUMMARY.....	90
7 ACKNOWLEDGEMENTS.....	92

Abbreviations

Ac	Activator
ATP	adenosine 5'-triphosphate
BSA	bovine serum albumin
cDNA	complementary DNA
CHO	Chinese hamster ovary
CMV	cytomegalovirus
DBD	DNA-binding domain
DDD	amino acid sequence containing of three aspartic acids
DDE	amino acid sequence containing two aspartic acids and one glutamic acid

DMEM	Dulbecco's modified Eagle medium
DNA	deoxyribonucleic acid
DR	direct repeat
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
EN	endonuclease
ENU	ethylnitrosourea
FACS	fluorescence-activated cell sorting
FCS	foetal calf serum
FP	Frog Prince
gag	group-specific antigen
gal	galactose
GFP	green fluorescent protein
HIV	human immunodeficiency virus
HMG	high-mobility group protein
HTH	helix-turn-helix
<i>Hsmar1</i>	<i>Homo sapiens mariner type 1</i>
IN	integrase
indel	insertion and/or deletion
IR	inverted repeat
IRES	internal ribosome entry site
IR/DR	inverted repeats containing direct repeats
IS	insertion sequence
kbp	kilo base pairs
kDa	kilo Daltons
L1	LINE1
LINE	long interspersed nuclear element
LTR	long terminal repeat
M	molar
Myr	million years
mRNA	messenger ribonucleic acid
Mu	mutator
NLS	nuclear localization signal
OPI	overproduction inhibition
ORF	open reading frame

PBS	phosphate buffer saline
PCR	polymerase chain reaction
polyA	poly-adenylation signal
PTGS	post-transcriptional gene silencing
RACE	rapid amplification of cDNA ends
RAG	recombination-activating gene
RNA	ribonucleic acid
RNAi	RNA interference
rpm	rotation per minute
RT	reverse transcriptase
SA	splice acceptor
SB	Sleeping Beauty
SD	splice donor
sem	standard error of the mean
shRNA	short hairpin RNA
SINE	short interspersed nuclear element
SV40	simian virus 40
Tc1	transposon of <i>Caenorhabditis elegans</i> 1
TE	transposable element
TSD	target site duplication
Ty	yeast transposon
UTR	untranslated region
UV	ultraviolet
V(D)J	variable (diversity) joining
vs.	<i>versus</i>
X-Gal	4-Cl-5-Br-3-indolyl- β -galactosidase
ZF	zinc finger
ZFN	zinc finger nuclease

1 INTRODUCTION

1.1 Discovery of transposable elements

Transposable genetic elements (“jumping genes”) were first discovered by Barbara McClintock in the 1940s. She found that certain spontaneous mutations in enzymes required for the productions of the purple anthocyanin pigment in maize are due to “controlling” elements that could apparently move from site to site in different chromosomes. This idea of jumping genes ran contrarily to the traditional view of the age that genomes are sable and static entities. The possibility that pieces of DNA can “jump around” in a genome was viewed by biologists with much skepticism. Therefore, McClintock’s observations were thought to be rare phenomena and not of general interest.

With the great advances of the molecular biology in the 1970s, it turned out that McClintock’s discovery was just the tip of an iceberg. Mobile element were found to be widespread not only in maize but in all kingdoms of living organisms from bacteria to humans. It turned out that transposable elements are indeed so abundant that they form a major fraction of the eukaryotic genome (Kidwell and Lisch 2001). However, most researchers still assumed that repetitive DNA elements do not have any function: they are useless, selfish DNA sequences (Orgel and Crick 1980). The term “junk DNA” coined by Sozumu Ohno repelled mainstream research from studying repetitive elements for many years (Ohno 1972). As Doolittle and Sapienza termed in *Nature* in 1980: transposons’ “only «function» is survival within genomes”...“thus no phenotypic or evolutionary function need to be assigned to them”.

This view started to change in the 1990s, when it became evident that transposons are important integral components of eukaryotic genomes with deep impacts on the host evolution. It turned out that they interact with the surrounding genomic environment, and increase the ability of the organism to evolve (Makalowski 2003).

Since their discovery, transposable elements have been broadening the scope of many fields of modern biology ranging from evolutionary genetics to gene therapy. There are numerous aspects of viewing transposable elements as subjects of scientific investigation. Transposons are of interest for *genome annotators*, for *structural and evolutionary geneticists* who investigate the role of mobile elements in chromosome/genome dynamics and their different contributions to host evolution. The ongoing studies of *molecular biologists* are continuously increasing our understanding of the mechanism transposition. Moreover, *experimental geneticists* use transposons routinely for insertional mutagenesis, gene tagging, germline transformations, gene trapping, and gene therapy. Their experimental model organisms range from bacteria to mammals. Due to the discovery of a variety of different prokaryotic and eukaryotic transposons, they are now routinely used as genetic tools in functional biology. Thus, repetitive elements are relevant to a wide scale of genetic studies, and transposons begin to be viewed as genomic treasure (Brosius 1991; Nowak 1994).

1.2 Classification of transposable elements

Discrete DNA sequences that possess an intrinsic capability to change their genomic locations are called **transposable elements** (TEs). TEs are distinguished whether their movement relies exclusively on DNA intermediates or includes an RNA stage. Transposons that move exclusively through a DNA intermediate are referred to as **DNA elements**.

Mobile elements that move through an RNA intermediate (**RNA elements or retroelements**) are transcribed, reverse transcribed and integrate as double stranded cDNA. These elements include **retroviruses** and the **retrotransposons**. DNA elements can be found in both prokaryotic and eukaryotic organisms, whereas RNA elements are restricted to eukaryotes.

1.2.1 DNA elements

These TEs can loosely be defined as sequences of DNA that can excise and insert into a variety of sites of a target DNA without the need to be reverse transcribed to cDNA. The simplest DNA elements are the **insertion sequences** (ISs) that were first characterized from bacteria in the late 1960s. Since then, approximately 800 ISs were identified (<http://www-is.biotoul.fr>). ISs are short (<2.5 kbps) and carry no genetic information except that necessary for their mobility. Thus, they are composed of a single gene coding for the transposase enzyme responsible for moving the element and of terminal inverted repeats (IRs) flanking it at both ends (Fig. 1D). The IRs contain the recombinationally active nucleotides at the very tips and specific recognition sequences for the transposase enzyme within.

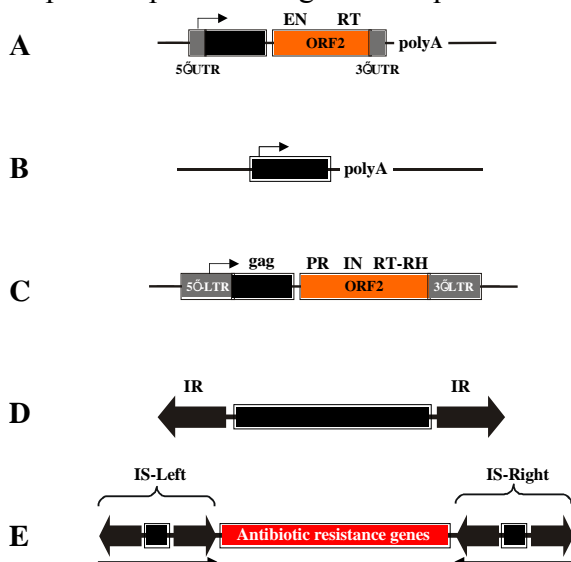


Figure 1. Structures and organization of the main types of transposable elements. (A) **Non-LTR retrotransposon.** The element consists of a 5' untranslated region that has promoter activity (arrow pointing towards the downstream genes), which is required to drive transcription of the element-encoded genes. ORF1 encodes a nucleic acid binding protein. ORF2 encodes an endonuclease (EN) and a reverse transcriptase (RT). The element has a polyA tail. (B) **A typical SINE.** The element is a small, RNA-derived pseudogene, which is transcribed from an RNA polymerase III promoter within the element (arrow). The element has a polyA tail. (C) **LTR-retrotransposon.** The element consists of long terminal repeats (LTRs) similar to those of retroviruses. The LTRs flank two open reading frames. ORF1 encodes the group specific antigen (gag), ORF2 encodes a protease (PR), an integrase (IN), and a reverse transcriptase-RNaseH (RT-RH) function. (D) **DNA transposon.** The central transposase gene (yellow box) is flanked by terminal inverted repeats (IRs, shown as black arrows). The IRs contain the binding sites for the transposase and sequences that are required for transposase-mediated cleavage. (E) **Composite bacterial transposon.** The element consists of antibiotic resistance genes (red box) flanked by two copies of an insertion sequence (IS) element that contains the transposase gene (yellow boxes). The arrows underneath indicate the inverted orientation of the IS elements.

Though most of the ISs are prokaryotic, a significant number of eukaryotic IS has also been documented. The largest and best-known group of these is the Tc1/*mariner* like elements that are structurally the closest to bacterial ISs (Plasterk et al. 1999). Another well-characterized member of the eukaryotic ISs is P element from *Drosophila melanogaster*. Recently, a new family of IS, the *helitrons* has been identified from eukaryotes. These elements lack IRs and move by the rolling circle mechanism, similarly to the replication of plasmids, and together with their descendants they represent 2% of the *Arabidopsis thaliana* and the *Caenorhabditis elegans* genome (Kapitonov

and Jurka 2001). The 300,000 DNA transposon fossils in human add up to around 3% of the genome (Lander et al. 2001).

In the 1960s it became evident that genes responsible for antibiotic resistance in bacteria can move between DNA molecules in a process analogous to the movement of ISs (Hedges and Jacob 1974). It was suggested that mobile elements that carry one or more genes that encode other functions in addition to those related to transposition should be called **transposons** (the term transposon is used in a wider sense, however: authors call all DNA elements, including ISs, transposons). Since these elements carry additional DNA they are usually larger than ISs (approximately 2.5-7 kbp). In some of these elements, called composite transposons, there are two complete ISs flanking a functional gene (Fig. 1E). This element can move as one functional unit, but also one or both of the bordering ISs can mobilize itself independently. There is a characteristic feature that distinguishes eukaryotic TEs from ISs and transposons in bacteria: the presence of a large number of inactive transposon copies that can in many cases be mobilized in *trans* by a limited number of active transposases (Kunze 1996).

1.2.2 RNA elements

Based on their structural properties and evolutionary relationships those transposable elements that can mobilize themselves through an element-derived RNA intermediate are grouped to those with long terminal repeats (LTR-retrotransposons and retroviruses) and those without (non-LTR retrotransposons).

A common feature of ***LTR-retrotransposons*** and ***retroviruses*** is that their coding region is flanked by LTRs (Fig. 1C). These sequences contain important control sequences e.g. promoter, enhancer and polyadenylation (polyA) signals. The coding sequences are divided into at least two open reading frames (ORFs). The first ORF encodes the group-specific antigen (*gag*) protein, required for the assembly of the RNA transcript into

cytoplasmic particles. The second ORF constitutes the *pol* gene, encoding a polyprotein, which consists of a protease (PR), a reverse transcriptase (RT) and an integrase (IN). The difference between retroviruses and LTR-retrotransposons is that retroviruses not only possess the capability to move between DNA molecules like other transposons, but they can leave their host cells too and integrate into new genomes. Nevertheless, retroviruses and LTR-retrotransposons are derived from a common progenitor (McClure 1991).

LTR-retroelements can be subdivided into three families based on homologies within the RT gene. The first two groups are named after their founding members found in yeast and *Drosophila*, Ty1/*copia* and Ty3/*gypsy* (Xiong and Eickbush 1990). The Ty3/*gypsy* elements form two subfamilies based on the presence or absence of a third ORF, *env*, encoding for envelope-like proteins. Retroviruses cluster into the third family of LTR-elements; they always possess a completely functional *env* gene for their viral life cycle. Many retroviruses, for example HIV, contain additional proteins (Trono 1995). Endogenous retroviruses (ERVs) appear to have been recently active in the mammalian genome. LTR-retrotransposons are widely distributed in eukaryotes, and make up about 8% of the human genome. Retroviruses were for long thought to be restricted to vertebrate genomes until it was shown that the *gypsy* retrotransposon is indeed an infectious retrovirus of *Drosophila melanogaster* (Kim et al. 1994). Transposition occurs through reverse transcription of the retrotransposon RNA, and integration of the resultant cDNA into a new location by the integrase protein.

The most abundant transposable elements in mammals are ***non-LTR retrotransposons*** represented by the long interspersed nuclear elements (LINEs) and the short interspersed nuclear elements (SINEs). Although LINEs are especially abundant in mammals (they make up 26% of the human X chromosome alone) (Lander et al. 2001), they have also been found in protozoan, insects, reptiles and plants (Malik et al. 1999). The major LINEs in humans (LINE1 or L1) are 6 kbp long and contain two ORFs (Fig. 1A). These encode for a nucleic acid binding protein and an enzyme with endonuclease (EN) and reverse transcriptase

(RT) activity, respectively (Fanning and Singer 1987). EN generates a single-stranded nick in the target DNA, and RT uses the nicked DNA to prime reverse transcription from the 3'-end of the L1 RNA (Moran and Gilbert 2002). Because reverse transcription is frequently incomplete, the majority of L1s is truncated, and thus nonfunctional. Consequently, even though L1 has about 5×10^5 copies in the human genome, thereby making up about 17% of human genomic DNA (Lander et al. 2001), the vast majority of these elements are inactive: in humans there are only 30-100 potentially active copies of L1 (Sassaman et al. 1997).

SINEs are short (about 100–400 bp) retrotransposable elements that encode no proteins; therefore, all of them are non-autonomous (Fig. 1B), and thought to use the enzymatic machinery of LINEs for transposition (Dewannieux et al. 2003; Dewannieux and Heidmann 2005). The vast majority of known SINEs are derived from tRNA sequences, with the exception of the human *Alu* element, which is derived from the 7SL component of the signal recognition particle (Ullu and Tschudi 1984). *Alu* elements were originally identified as repetitive DNA elements in human DNA renaturation curves, and contain a recognition site for the restriction enzyme *AluI*. *Alu* elements are represented in the human genome with $>1 \times 10^6$ copies which make up about 11% of the total genome. *Alu* is the only active SINE in humans. Full-length *Alus* are 280bp long, contain promoter sequences for RNA polymerase III (Weichenrieder et al. 2000) and a polyA tail (Fig. 1B). The transcripts of RNA polymerase III transcribed *Alus* terminate at RNA polymerase III termination signals fortuitously present in the 3' flanking DNA. Rarely, RNA polymerase II-derived host gene transcripts can also be *trans*-mobilized by functional LINE proteins. These transposition products are named processed pseudogenes. They lack promoters, introns and end in a polyA tail. Only short target site duplications flanking these sequences provide evidence that these integrants are in fact transposition products.

1.3 Modes of transposition

The sum of molecular events involved in the movement of a transposable element from one chromosomal location to another is defined as **transposition**. There are two types of transposition reaction distinguished by whether the TE is replicated during the process or not (Fig. 2).

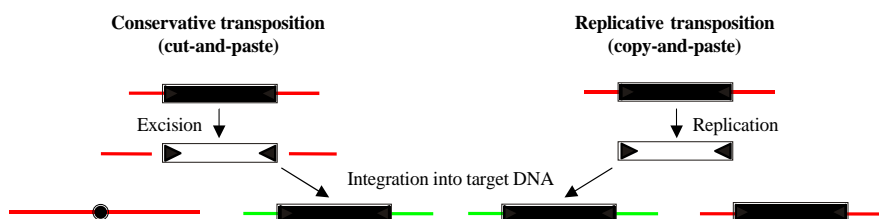


Figure 2. Schematic representation of the two major mechanisms of transposition. During conservative transposition, the element is excised from the donor DNA (red line), and integrates into a new target DNA (green line). The broken donor DNA has to be repaired by host factors, and this process can result in a small “footprint” (black dot) that marks the former presence of the element in that site. Replicative transposition requires amplification of the element either by replication or by copying of the element through transcription followed by reverse transcription. The amplified element gets inserted elsewhere in the genome.

During the vast majority of **replicative** (copy-and-paste) transposition events, the transposon does not get excised from its donor locus,

but instead a copy of it is produced that subsequently inserts elsewhere in the genome (Fig. 2).

Thus, replicative transposition leads to an increase in the copy number of the transposon within a genome. If the new copy is produced by transcription and subsequent reverse transcription of transposon sequences, the process is referred to as retrotransposition. The movement of retroviruses and retrotransposons is always of the replicative type, because it is the cDNA copy, not the original transposon, which is transposed. However, replicative transposition is not restricted to retroelements. For example, the *IS6* family and *Tn3* (Mahillon and Chandler 1998), and the complex DNA transposon, bacteriophage Mu (Lavoie and Chaconas 1996), can also follow the replicative mode of transposition.

In **non-replicative** (also called conservative) transposition, the element is excised from a genomic locus and integrates to another through a so-called “cut-and-paste” mechanism (Fig. 2). In non-replicative transposition, the genetic information of the element is carried by DNA. The bacterial *IS10* (Kleckner et al. 1996), *Tn7* (Craig 1996) and eukaryotic transposons including the P element (Kaufman and Rio 1992), members of the Tc1/*mariner*

family and the maize transposon *Ac/Ds* discovered by McClintock all use the cut-and-paste mechanism for their transposition (Plasterk et al. 1999).

In cut-and-paste transposition, amplification is not inherent to the transposition process itself; nevertheless, the copy numbers of DNA transposons also increase over time. Transposon amplification can occur when transposition takes place in the S-phase of the cell cycle. If a transposon is excised from an already replicated segment of the DNA, and reintegrates into a chromosome that has not been replicated, the process results in an increase by one copy of the transposon. If this event is followed by meiosis, two of the four germ cells have one more transposon copies compared to its parental cell (Lodish et al. 2004). Another way of increasing in copy number of non-replicative transposons was described for the P element (Engels et al. 1990) and the *Mos1 mariner* transposon (Lohe et al. 2000). After the excision of these elements the resulting gap in the donor chromosome can be sealed by a process called template-directed gap repair. This host repair mechanism uses the sister chromatid, the homologous chromosome or an ectopic site for refilling the gap created by the excised element.

1.4 DNA elements in natural hosts

1.4.1 Impact of transposons on host genomes: Mutations, genome size and the evolution of novel gene functions

When first colonizing new genomes, TEs are only parasitic sequences, however, over evolutionary times they can become integral components of genomes. Their effects can result directly from transpositional activity and TE-induced mutations, or because TEs represent a rich enzymatic and regulatory diversity that can result in the co-option of their sequences and enzymatic activities by the host (Kidwell and Lisch 2002).

A general feature of transposable elements is that they can replicate independently of the cellular replication cycle, and new copies can emerge at new locations in the genome.

Thus, mobile elements can cause insertional mutagenesis if they land within a gene (Kazazian 1999), but they can also lead to altered gene expression and genetic recombination. Insertion of a transposable element into the protein-coding region of a gene (exon) can disrupt gene function. For example, bacterial IS elements were identified as DNA insertions in the *E. coli gal* operon which cause highly polar mutations. In maize, the *Mutator* system can increase the mutation frequency by 50-fold over background. In *Drosophila*, it is estimated that 50-70% of all mutations are due to transposition, and the DNA transposon Tc1 is the main cause of mutation in the nematode *Caenorhabditis elegans*. In addition to direct insertion into exon sequences of a gene, TEs can effect gene expression and regulation by integrating into non-transcribed or non-translated regions of genes (Rubin et al. 1982; Bradley et al. 1993; Selinger and Chandler 1999). For example, insertion between the core promoter and adjacent enhancer regions would increase the distance between these regions, and thus negatively affect promoter activity. In addition to insertional mutagenesis, cut-and-paste TEs can alter gene function by excising. It is because after the DNA break the host repair can rarely reproduce the sequence, as it existed before the integration. The excision can result in addition of new sequences or deletion of host sequences (Colot et al. 1998). Another damaging aspect of transposable elements is that repeated, dispersed copies of homologous sequences can promote secondary rearrangements (Deininger and Batzer 1999), which can result in deletions, duplications and inversions. This potential of dispersed transposon copies to promote homologous recombination can be even more damaging to the genome than a *de novo* insertion.

TEs can also induce large-scale changes in the whole genome size. There is greater than 80,000-fold difference in size of the smallest and the largest eukaryotic genomes, however, the genome size is not correlated with organism complexity, which gives rise to the C-value paradox. In many plant and animal species and also in humans, abundant TEs account for the C-value paradox (Casa et al. 2000; Lander et al. 2001). A striking example of

TE-induced genome restructuring is the programmed somatic excision of interstitial DNA segments in ciliates. These cells contain a macro- and a micronucleus. The genome of the transcriptionally active macronucleus consists of segments of the micronucleus, which is rearranged during development. The process involves extensive DNA excision and rejoining. TEs are major components of the eliminated DNA, and it has been proposed that invasion of these TEs contributed to the evolution of the nuclear excision process (Klobutcher and Herrick 1995).

About 1 in 600 mutations in humans is estimated to arise from retrotransposon-mediated insertion. The major causative agent of endogenous genomic insertions is L1 (Kazazian 1999). An average human being has 80-100 retrotransposition-competent L1s, which belong to a particular subfamily of these elements in the human genome. Results also suggest that a relatively small number of very active L1s comprise the bulk of L1 activity (Kazazian 1999). A current estimate for transpositional frequencies in humans is that about 1 in 8 individuals harbor a new L1 insertion (Moran and Gilbert 2002). New, disease-causing insertions of L1 in humans were in fact the first retrotransposition events detected in mammals. These insertions occurred in the blood clotting protein Factor VIII, dystrophin, APC and β -globin genes.

Alu elements continue to amplify at a rate of about one insertion every 200 new births. New insertion events can lead to genetic disorders including hemophilia, neurofibromatosis, cholinesterase deficiency, breast cancer and leukemia (Deininger and Batzer 1999). *Alu* element insertion is estimated to contribute to about 0.1% of human genetic diseases. The large number of *Alu* elements within the human genome also provides ample opportunity for homologous recombination events between dispersed *Alu* repeats. These events can result in deletion or duplication of exons in a gene, and other chromosomal abnormalities. This mode of mutagenesis is estimated to account for 0.3% of human genetic diseases, including Fabry disease, Duchenne's muscular dystrophy, ADA deficiency and a variety of cancers.

One of the properties that distinguish TEs from other mutagens is that they are regulated both by themselves and by the host. Self-restraint of transposons has probably evolved to decrease the extent of damage to the host. However, the costs and benefits of TE movement can change during evolutionary times. Indeed, increased rates of transposition can even be selected when the host population is under stress (Hartl 2000). If the tight regulation of TEs breaks down due to stress, their activity can produce potentially fitter host variants. There are many examples of this phenomenon in plant evolution (Wessler 1996).

1.4.1.1 Transposons as a creative force

Transposable elements cannot only do harm, but also represent a creative force. In *Drosophila*, telomere maintenance is not brought about by telomerase, but by repeated transposition of two non-LTR retrotransposons, HeT-A and TART, into chromosome ends. The acquisition of new transposon insertions can donate regulatory elements to genes, or even lead to the evolution of new genes. L1 elements can carry non-transposon sequences into new places, a process that can contribute to “exon shuffling” and thus to gene evolution (Moran and Gilbert 2002). This is because L1 transcription can read through the native transcription termination site of the element into flanking genomic sequences. It is estimated that about 0.5-1% of the human genome may have been generated by L1-mediated transduction of 3'-flanking sequences (Pickeral et al. 2000). The L1 retrotransposition machinery can also mediate reverse transcription and genomic insertion of host gene mRNAs, resulting in processed preusogenes. Some of these insertions can give rise to functional processed genes. Approximately 50 human genes evolved from transposable elements, mostly from DNA transposons (Lander et al. 2001). These include the RAG1 and RAG2 immunoglobulin gene recombinases (Jones and Gellert 2004) and the centromere-binding protein CENPB (Casola et al. 2008). Thus, although transposable elements have not been selected for conferring selective advantage to the host, they can contribute useful functions to genomes.

1.5 The Tc1/*mariner* superfamily of transposable elements: how they live and how they die

Members of the Tc1/*mariner* family are probably the most widespread DNA transposons in nature, represented in ciliates, plants, fungi and animals (Plasterk et al. 1999). This monophyletic family is defined on the basis of transposase sequence homologies, and the similar molecular mechanism of transposition (Robertson 1995). These transposons are generally 1.3-2.4 kbp long, and encode a transposase gene flanked by terminal inverted repeats (IRs) (Fig. 1D). Tc1/*mariner* elements follow the cut-and-paste mechanism of transposition without overt target preference, except that they always integrate into TA target dinucleotides of host chromosomes (Plasterk et al. 1999).

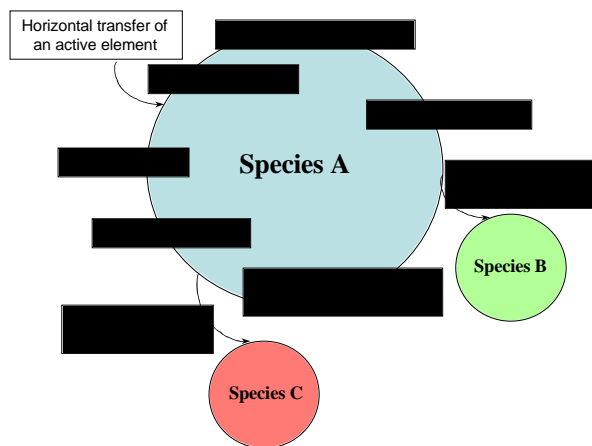


Figure 3. Evolutionary life-cycle of Tc1/*mariner* elements in natural hosts. The main events of the life-cycle are depicted (for details see text). The cycle was proposed to describe the evolution of *mariner* elements (Hartl et al., 1997), but is probably also valid for other DNA elements. Horizontal transfer of active transposons into new species can occur before or after functional diversification. Modified after Hartl et al., 1997 and Lampe et al., 2001.

Phylogenetic relationships between very closely related Tc1/*mariner* elements are often inconsistent with those of their hosts (Robertson 1993; Ivics et al. 1996). For instance, the closest relatives of a *mariner* subfamily in humans can be found in insects, worms and in a hydra species (Robertson and Zumpano 1997; Lampe et al. 2001). It has been suggested that “horizontal

transfer” accounts for the spreading of elements across distantly related phyla (Garcia-Fernandez et al. 1995; Lam et al. 1996) (Fig. 3). Because TEs are not infectious, it is not exactly known how they can invade new genomes. Potential vectors of horizontal transmission include viruses, external and intracellular parasites (Houck et al. 1991; Kidwell 1992; Kidwell 1992). Once a transposon is transferred to a new host, it has to colonize its germline to persist in a population or, ultimately, in the entire species. At this initial phase, transposons can explosively amplify themselves (Engels 1989) (Fig. 3). However,

transposons are not selected for function, and thus mutations may accumulate in them in a time-proportional manner (neutral evolution), resulting in partially or completely inactive transposon copies. This process is termed vertical inactivation (Lohe et al. 1995) (Fig. 3). In parallel, mutated transposase copies might become dominant negative regulators of transposition. Thus, with time, the rate of propagation slows down and finally, due to random genetic drift, transposons start to be extinct from their host genomes. The phenomenon is known as “stochastic loss” (Hartl et al. 1997) (Fig. 3). Therefore, in order to survive, transposons have to be horizontally transferred to new germ lines and start their life cycle over again (Fig. 3). DNA transposons are believed to be transferred horizontally more often than retroelements, possibly because the endurance of DNA intermediates of transposition within cells offers a better chance for hitchhiking transfer vectors (Silva et al. 2004). Indeed, in some retrotransposition reactions the RNA intermediate is directly reverse transcribed into the integration site (Luan et al. 1993), thereby offering little chance to be horizontally transferred.

Due to the above mechanisms, *Tc1/mariner* transposons are extraordinarily widespread in nature, but the vast majority of these elements are defective in all eukaryotic genomes. The active invertebrate *Tc1/mariners* were isolated from *Caenorhabditis elegans* [*Tc1*, (Emmons et al. 1983) and *Tc3* (Collins et al. 1989)], from the *Drosophila* genus [*Mos1*, (Medhora et al. 1991) and *Minos* (Franz and Savakis 1991)] and from the earwig *Forficula auricularia* [*Famar1*, (Barry et al. 2004)]. The active *Himar1* element is a majority rule consensus of cloned genomic copies obtained from the horn fly *Haematobia irritans* (Lampe et al. 1996). However, extensive search for active vertebrate transposons has so far failed to yield an active vertebrate *Tc1/mariner*-like transposon.

1.5.1 *Sleeping Beauty* kissed back to life

As discussed above, despite their wide distribution, all Tc1/*mariner* transposons isolated to date from vertebrates are transpositionally inactive. To address this problem, an ancestral Tc1-like element was reactivated from fish genomes. The molecular resurrection procedure involved the systematic removal of inactivating mutations by mutagenesis of an inactive transposase sequence. The active element is a majority rule consensus sequence of several

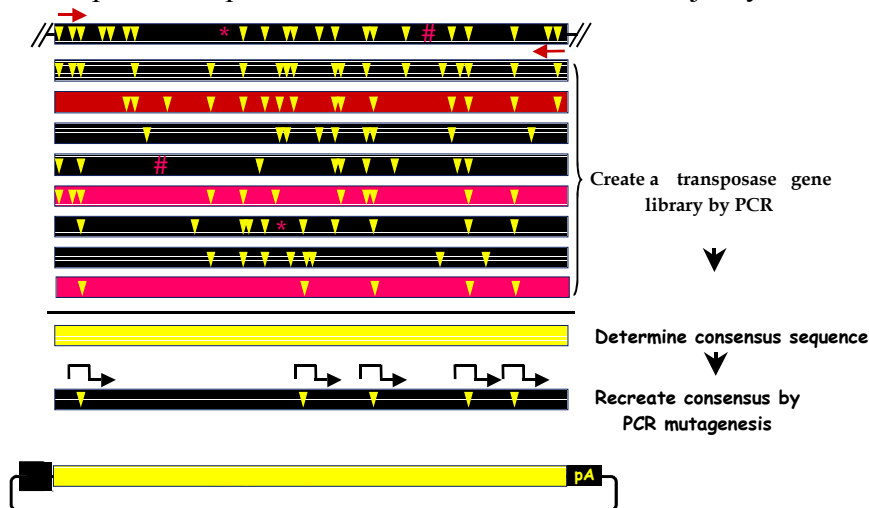


Figure 4. Molecular reconstruction of the *Sleeping Beauty* transposase gene. The strategy of first constructing an open reading frame for a salmonid transposase and then systematically introducing amino acid replacements into this gene is illustrated. Amino acid residues are typed black when different from the consensus, and their positions within the transposase polypeptide are indicated with arrows. Translational termination codons appear as asterisks, frame shift mutations are shown as #.

dead genomic copies of transposons from different fish species. Therefore, the engineered element, which was called *Sleeping Beauty* (SB), represents an archetypical sequence

that was presumably active 10-15 million years ago (Ivics et al. 1997).

1.5.1.1 Structural and functional components of *Sleeping Beauty*

1.5.1.1.1 The transposon inverted repeats

Transposons are bracketed by terminal inverted repeats that contain binding sites for the transposase. Tc1/*mariner* elements have a roughly uniform size of approximately 1.6-1.7 kb, indicating a natural selection in genomes for this particular size. *Sleeping Beauty* has a pair of transposase-binding sites at the ends of the 200-250 bp long inverted repeats (IRs). Within each IR of SB, there are two transposase binding sites that contain short, 15-20 bp direct repeats (DRs). This special organization of inverted repeat is termed IR/DR (Izsvák et al. 1995; Plasterk et al. 1999) (Fig. 5), and can be found in numerous elements in the Tc1

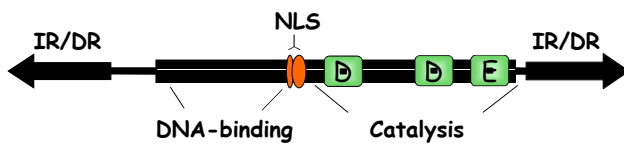


Figure 5. Schematic representation of *Sleeping Beauty*, a Tc1/*mariner* transposon. The terminal inverted repeats (IR/DR, black arrows) contain one or two binding sites for the transposase (white arrows). The element contains a single gene encoding the transposase (purple box). The N-terminal part of the transposase contains a DNA binding domain, followed by a nuclear localization signal (NLS). The C-terminal part of the protein is responsible for catalysis, including the DNA cleavage and rejoining reactions. The DDE amino acid triad is a characteristic signature of the Tc1-like transposases, *mariners* have DDD.

transposon family, including *Frog Prince* from *Rana pipiens*, the *Minos*, *S. Paris* and *Bari* elements in various *Drosophila* species (Franz and Savakis 1991; Merriman et al. 1995; Moschetti et al. 1998; Plasterk et al. 1999), *Quetzal*

elements in mosquitos (Ke et al. 1996), at least three Tc1-like transposon subfamilies in fish (Ivics et al. 1996) and *Txr*, *Eagle*, *Froggy* and *Jumpy* transposons in *Xenopus* (Lam et al. 1996; Sinzelle et al. 2005). The spacing of about 200 bp between the outer and inner binding sites is conserved in all elements within the IR/DR group, but the actual DNA sequences are not similar, suggesting convergent evolution of the IR/DR-type repeats. The IR/DR group significantly differs from Tc1 or the *mariner* elements that are more simple and have repeats of less than 100 bp and a single transposase binding site per repeat. All four binding sites within the IR/DR structure are required for SB transposition (Izsvák et al. 2000). The four binding sites are not identical, the outer ones are longer by two base pairs. The inner DRs are more strongly bound by the transposase than the outer DRs (Cui et al. 2002; Zayed et al. 2003), and replacement of the outer DRs with inner DR sequences was found to abolish transposition (Cui et al. 2002). This suggests that the unequal strengths of transposase binding and the positions of the DRs within the inverted repeats are required for ordered assembly of transposase-DNA complexes at the ends of the transposon that has a fundamental effect on the outcome of the transposition reaction. The IRs are not identical either; the left IR contains a sequence motif called the HDR, which resembles the 3'-half of the transposase binding sites (Izsvák et al. 2002). A construct containing two left IRs transposes better than the wild-type transposon, but another version that has two right IRs has very poor mobility, indicating that the left and right IRs are functionally distinct (Izsvák et al. 2002).

1.5.1.1.2 The transposase

As discussed above, transposons are very diverse genetic entities, however their enzymes carry out similar chemical reactions e.g. hydrolysis for strand cleavage and transesterification for strand transfer. The similar activities of TEs are manifested in the remarkable overall structural similarity of the transposition proteins.

Both the transposases of ISs and transposons and integrases of retroelements show structural similarities for their functional organization. Most of them can be divided into topological distinct functional domains. Partial proteolysis experiments revealed that the transposon specific DNA binding domains are generally localized in the N-terminal part, whereas the catalytic domain responsible for the strand cleavage and transfer is located in the C-terminal of the transposase protein (Machida and Machida 1989; Lavoie and Chaconas 1996) (Fig. 5). One possible explanation for this characteristic arrangement in prokaryotic elements is that during translation the N-terminal part of the premature transposase protein can fold independently of the C-terminal catalytic domain, and interact with its specific transposon binding sites close to the point of synthesis. This hypothesis is reinforced by the observation that the presence of the C-terminal part of some bacterial transposases decrease the affinity of the IR binding (Weinreich et al. 1994). This arrangement can facilitate that the transposase is going to act on the transposon that produced it (a phenomenon called *cis*-preference) (Jain and Kleckner 1993).

1.5.1.1.2.1 DNA recognition

It is a key feature of all transposases that they recognize their specific transposon ends. TEs that move by transposon-specific transposases possess recognition sequences in their IRs. The majority of ISs has simple 10-40 bp long IRs while others, exhibit long and complex IRs. Most transposon ends are composed of two functional parts. The 2-3 terminal base pairs of the ends are the recombinationally active sequences involved in the cleavage and the strand

transfer reactions. The other functional part is situated within the IRs and it ensures the sequence-specific positioning of the transposase on the transposon ends (Derbyshire et al. 1987; Ichikawa et al. 1990). ISs have single transposase binding sites whereas for example Mu and *Tn7* have complex, asymmetric recognition sites (Hauer and Shapiro 1984; Craigie 1996). The bi-functionality of the transposon ends is reflected in the arrangement of the transposase on its cognate transposon. Due to the flexibility of the transposase, the N-terminal region of the enzyme attaches to the inner segment of IRs while the C-terminal contacts the external ends. The sequence-specific DNA-binding of both eukaryotic and prokaryotic transposases is often carried out by a helix-turn-helix (HTH) motif. This domain can be simple as it is the case of IS transposases (Mahillon and Chandler 1998), or can be complex and bipartite as found in *Ac*, Mu or in Tc transposases (Plasterk 1996; Becker and Kunze 1997). The catalytic C-terminal domains of transposases are also involved in DANN-binding, however, this activity is not sequence specific and contributes to the correct positioning of the transposon end into the catalytic pocket (Haren et al. 1999).

The overall domain structure of the transposase is conserved in the entire Tc1/*mariner* superfamily (Plasterk et al. 1999). Specific substrate recognition is mediated by an N-terminal, bipartite DNA-binding domain of the transposase (Fig. 5) (Vos and Plasterk 1994; Pietrokovski and Henikoff 1997; Izsvák et al. 2002). This DNA-binding domain has been proposed to consist of two HTH motifs, similar to the paired domain of some transcription factors in both amino acid sequence and structure (Franz et al. 1994; Vos and Plasterk 1994; Ivics et al. 1996). The modular paired domain has evolved versatility in binding to a range of different DNA sequences through various combinations of its subdomains (PAI+RED) (Czerny et al. 1993). The nucleotide sequences recognized by the composite paired domain are degenerate, the DNA-binding specificity is relaxed (Pellizzari et al. 1999). The origin of the paired domain is not clear, but phylogenetic analyses indicate that it might have been derived from an ancestral transposase (Breitling and Gerber 2000). Partially overlapping with

the RED subdomain in the transposase is a nuclear localization signal (NLS in Fig. 5), flanked by phosphorylation target sites of casein kinase II (Ivics et al. 1996). Phosphorylation of these sites is a potential checkpoint in the regulation of transposition. The NLS indicates that these transposons, unlike murine retroviruses, can take advantage of the receptor-mediated transport machinery of host cells for nuclear uptake of their transposases. A characteristic GRPR-like motif (GRRR) between the two HTH motifs is similar to an AT-hook (Izsvák et al. 2002), responsible for minor groove interactions in the Hin invertase of *Salmonella* (Feng et al. 1994) and in the RAG1 recombinase of V(D)J recombination (Spanopoulou et al. 1996).

1.5.1.1.2.2 The catalytic domain

The majority of known transposases and INs possess a well-conserved triad of amino acids, known as the aspartat-aspartat-glutamat, in short the DDE motif (actually, more of a signature than a “motif” in a usual sense) in their C terminal catalytic domain (Kulkosky et al. 1992) (Fig. 5). The DDE motif is found in a large group of recombinases, including retrotransposon and retrovirus integrases, bacterial IS element transposases (Doak et al. 1994) and RAG1 (Doak et al. 1994; Kim et al. 1999; Landree et al. 1999). However, transposases of P element, *Ac* and *En/Spm* transposons exhibit different structures. Structural analyses of the HIV INs and mutational studies revealed that the DDE triad lies in the heart of the catalytic domain of transposases and INs (Dyda et al. 1994; Vos and Plasterk 1994). These amino acids play essential role in catalysis by coordinating, in general, two divalent cations necessary for activity. Retroviral INs were shown to be able to coordinate Ca^{++} , Zn^{++} and Mn^{++} ions, but the biologically relevant cation is thought to be Mg^{++} (Bujacz et al. 1997; Goldgur et al. 1998). One metal ion acts as a Lewis acid, and stabilizes the transition state of the penta-coordinated phosphate, the other one acts as a general base and deprotonizes the incoming nucleophil during transesterification and strand transfer (Haren et al. 1999).

Within the catalytic domains of Tc1-like transposases, a conserved glycine-rich subdomain can be found (Ivics et al. 1997). The function of this subdomain is unknown. In addition to the DDE-containing transposases and integrases (Dyda et al. 1994; Davies et al. 2000), crystallographic analyses of the catalytic domains of proteins whose functions are not obviously related to transposition, such as RNAase H (Katayanagi et al. 1990) or RuvC (Ariyoshi et al. 1994) have revealed a remarkably similar overall fold. The existence of a common structural motif that catalyses polynucleotidyl transfer reactions in diverse biological contexts suggests that the different specificities in binding to DNA might have evolved by the apparent acquisition of different DNA-binding domains in the evolution of DDE recombinases (Capy et al. 1996).

1.5.1.2 The molecular mechanism of *Sleeping Beauty* transposition

The transposase protein and the inverted repeats together engage in a series of molecular events that lead to the excision of the element from its DNA context and reintegration into a different locus, a process termed cut-and-paste transposition. 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

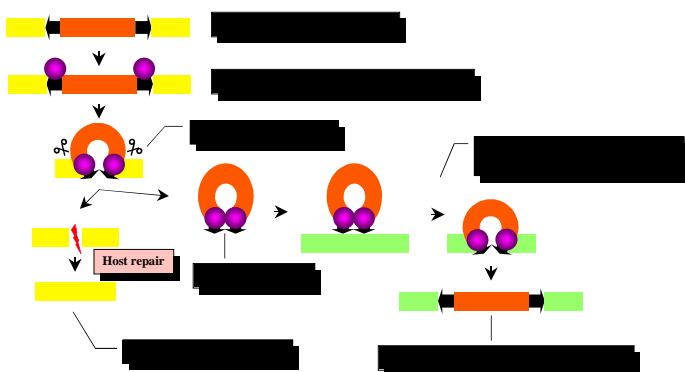


Figure 6. Mechanism of *Sleeping Beauty* transposition. The transposase (purple spheres) binds to its sites within the transposon inverted repeats (black arrows). Excision takes place in a synaptic complex. Excision separates the transposon from the donor (yellow) DNA, and the double-strand DNA breaks that are generated during this process are repaired by host factors. The excised element integrates into a TA site in the target (green) DNA that will be duplicated and will be flanking the newly integrated transposon.

elements are paired and held together by transposase subunits; 3)

excision from the donor site; 4) reintegration at a target site (Fig. 6).

On the molecular level, mobility of DNA-based transposable elements

can be regulated by imposing constraints on transposition. One

important form of transpositional

control is represented by regulatory “checkpoints”, at which certain molecular requirements have to be fulfilled for the transpositional reaction to proceed. These requirements can operate at any of the four different stages of transposition listed above, and can be brought about by both element-encoded and host-encoded factors.

1.5.1.2.1 Specific DNA-binding by the *Sleeping Beauty* transposase

Similar to the DNA-binding domain of the transposase, the binding sites also have a bipartite structure in which the 3'-part of the binding site is recognized by the PAI subdomain, whereas the 5'-sequences interact with the RED subdomain of the transposase (Izsvák et al. 2002). Specificity of DNA-binding is predominantly determined by base-specific interactions mediated by the PAI subdomain (Izsvák et al. 2002). The PAI subdomain also binds to the HDR motif within the left inverted repeat of SB, and mediates protein-protein interactions with other transposase subunits. Thus, the PAI subdomain is proposed to have at least three distinct functions: interaction with both the DRs and the HDR motif, and transposase oligomerization. In cooperation with the main DNA-binding domain, the GRRR motif was shown to function as an AT-hook, contributing to specific substrate recognition (Izsvák et al. 2002). Although part of the NLS is included in the RED subdomain, it does not appear to contribute to DNA recognition. Domain swapping experiments have shown that primary DNA-binding is not sufficient to determine specificity of the transposition reaction. Zebrafish Tdr1 elements are closely related to SB, but are not mobilized by SB transposase. Comparison of the transposase binding site sequences of SB and Tdr1 elements revealed main differences in the 5'-half of the DRs. This sequence is contacted by the RED subdomain, indicating that the function of the RED is to enforce specificity at a later step in transposition. Substrate recognition of SB transposase is therefore sufficiently specific to prevent activation of transposons of closely related subfamilies.

The spacing between the DRs is conserved in the IR/DR group, and decreasing the distance between the DRs has a negative effect on transposition (Izsvák et al. 2000). The transposase does not bind the DRs with equal affinity, it preferentially binds the internal recognition sequences (Cui et al. 2002; Zayed et al. 2003). Perhaps due to the two-base-pair difference in length, the helical phasing of the outer binding sites make transposase binding unfavored at these sites. The significance of this unequal affinity in binding is discussed in the next section.

1.5.1.2.2 Synaptic complex assembly and the role of multiple binding sites for the transposase

A uniform requirement among transposition reactions is the formation of a nucleoprotein complex, before the catalytic steps can take place. This very early step, synaptic complex assembly, is the process by which the two ends of the elements are paired and held together by transposase subunits (Fig. 6). *Sleeping Beauty* transposition is controlled at the level of complex assembly (Izsvák et al. 2002). The paired-like DNA-binding domain forms tetramers in complex with transposase binding sites (Izsvák et al. 2002). The necessary factors that are required for synaptic complex assembly of SB include the complete inverted repeats with four transposase binding sites, the HDR motif and tetramerization-competent transposase. These tetrameric complexes form only if all the four binding sites are present and they are in the in proper context. The HDR motif is important but not essential in transposition, and therefore can be viewed as a transpositional enhancer that, together with the PAI subdomain of the transposase, stabilizes complexes formed by a transposase tetramer bound at the IR/DR. In contrast to Mu transposase, where the two specificities of binding to the enhancer and to the recombination sites are encoded in two distinct domains (Leung et al. 1989), the paired-like region of SB transposase combines these two functions in a single protein domain.

1.5.1.2.2.1 The role of HMGB1 in *Sleeping Beauty* transposition: Ordered assembly of synaptic complexes

Differential interactions between the transposon and host-encoded factors may result in limitation of host range. The high mobility group protein HMGB1 is required for efficient *Sleeping Beauty* transposition in mammalian cells (Zayed et al. 2003). HMGB1 is an abundant, non-histone, nuclear protein associated with eukaryotic chromatin, and has the ability to bend DNA (Bustin 1999). SB transposition was significantly reduced in HMGB1-deficient mouse cells. This effect was complemented by expressing HMGB1 and HMGB2, but not with the more distantly related HMGA1 protein. Overexpression of HMGB1 in wild-

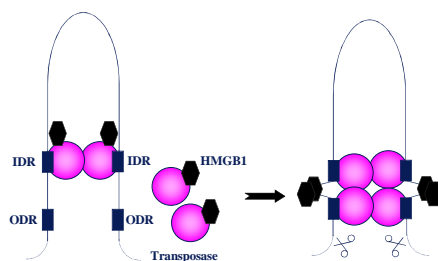


Figure 7. A proposed model for the role of HMGB1 in *Sleeping Beauty* synaptic complex formation. Sleeping Beauty transposase (pink spheres) recruits HMGB1 (dotted hexagons) to the transposon inverted repeats. First, HMGB1 stimulates specific binding of the transposase to the inner binding sites (IDRs). Once in contact with DNA, HMGB1 bends the spacer regions between the DRs, thereby assuring correct positioning of the outer sites (ODRs) for binding by the transposase. Cleavage (scissors) proceeds only if complex formation is complete. The complex includes the four binding sites (black boxes) and a tetramer of the transposase.

type cells enhanced transposition, indicating that HMGB1 is a limiting factor of transposition. HMGs have low affinity to standard, B-form DNA, and interactor proteins need to guide them to certain sites (Bustin 1999). SB transposase was found to interact with HMGB1 *in vivo*, and to form a ternary complex with the transposase and transposon DNA, suggesting that the transposase may actively recruit HMGB1 to transposon DNA via protein-protein interactions.

Considering the significant drop of transposition activity in HMGB1-deficient cells, the role of HMGB1 in transposition is a critical one. HMGB1 was proposed to promote communication between DNA motifs within the transposon that are otherwise distant to each other, including the DRs, the transpositional enhancer and the two IRs (Fig. 7). However, as mentioned above, physical proximity of the DRs is not sufficient for SB transposition; a highly specific configuration of functional DNA elements within the inverted repeats has a critical importance. As mentioned earlier, SB transposase preferentially binds the inner DRs within the transposon inverted repeats. It was also found that HMGB1 enhances transposase

binding to both DRs, but its effect is significantly more pronounced at the inner sites. It appears, therefore, that the order of events that take place during the very early steps of transposition is binding of transposase molecules first to the inner sites, and then to the outer sites. The pronounced effect of HMGB1 on binding of the transposase to the inner sites suggests that HMGB1 enforces ordered assembly of a catalytically active synaptic complex (Fig. 7). Indeed, interference with this sequence of events by replacing the outer transposase binding sites with the inner sites abolishes SB transposition (Cui et al. 2002). This ordered assembly process probably controls that cleavage at the outer sites occurs only if all the previous requirements had been fulfilled. An assembly pathway similar to the one proposed for SB has been described for bacteriophage λ (Richet et al. 1986).

The IR/DR-type organization of inverted repeats introduces a higher level regulation into the transposition process. The repeated transposase binding sites, their dissimilar affinity for the transposase, and the effect of HMGB1 to differentially enhance transposase binding to the inner sites are all important for a geometrically and timely orchestrated formation of synaptic complexes, which is a strict requirement for the subsequent catalytic steps of transposition.

1.5.1.2.3 The biochemistry of cut-and-paste transposition

Central to all transposition reactions are the excision and integration of a polynucleotide, therefore transposons execute polynucleotide transfer reactions. The transposition reactions can be divided into three stages: liberation of the element from its donor DNA by single-, or double-strand DNA cleavage, transfer of the transposon to its target site and processing of the transposition product by host-encoded enzymes (Haren et al. 1999). All transposition reactions involve DNA breakage and joining; the nature of the emerging transposition products depends on which strand of the DNA is cleaved and joined.

1.5.1.2.3.1 Transposon excision

The key process of all transposon excision is the exposure of the 3'-OH groups of the transposon ends, which will later be used at the strand transfer reaction for integration (Mizuuchi 1992) (Fig. 8). In the case of phage Mu and retroviral transposition the DNA cleavage involves only a single strand cut at each transposon ends. The vast majority of transposases, however, cleave both DNA strands of the corresponding transposon. During the excision of bacterial cut-and-paste elements, it is the first nick that generates the 3'-OH groups at the transposon ends. On the contrary, transposases of eukaryotic cut-and-paste

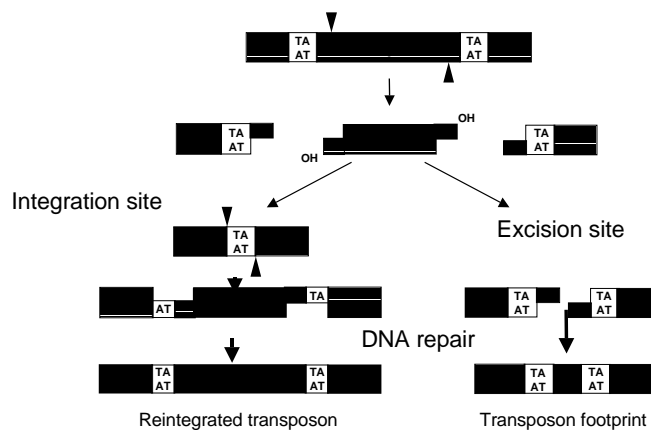


Figure 8. Cut-and-paste transposition of Tc1/mariner transposons. The element (black box) is removed from its original site with staggered cuts, which leaves some transposon nucleotides at the site of excision. The excised element reintegrates elsewhere in the genome at a TA target dinucleotide. Repair of the single stranded gaps of the integration site results in the duplication of the target TA. The excision site is predominantly repaired by non-homologous end joining, which leaves behind a transposon footprint.

transposons first generate a 5' P on the transposon ends and the 3'-OH groups are exposed only as a result of the second strand cut (Curcio and Derbyshire 2003). In case of retroviruses, this process operates on the double stranded cDNA of the element, and results in the cleavage of only two bases from the 3'-end of the cDNA (Mizuuchi 1997).

Every DNA strand cleavage in all transposition reactions is a transposase- or integrase-catalyzed, Mg^{++} - dependent hydrolysis of the phosphodiester bonds of the DNA backbone, executed by a nucleophilic molecule. All the DDE recombinases catalyze similar chemical reactions (Craig 1995), which begin with a single-strand nick that generates a free 3'-OH group. In the case of the first strand cleavage the nucleophilic molecule is H_2O (Mizuuchi 1992). During cut-and-paste transposition, nicking of the element is followed by the cleavage of the complementary DNA strand too. To catalyze second strand cleavage, DDE recombinases developed versatile strategies (Turlan and Chandler 2000). This cleavage can

occur at different positions relative to the transposon ends. The position of 5'-cleavage of the second strand required for the liberation of the element occurs directly opposite to the 3'-cleavage site in V(D)J recombination (Gellert 2002; van Gent et al. 1996) and for the bacterial *Tn10* element (Kennedy et al. 1998) (thereby generating blunt ended products). For *Tn7* the cleavage occurs three nucleotides toward the 5'-end of the element (Haren et al. 1999). In case of the *Tc1/mariner* elements the non-transferred strand is cleaved a few nucleotides within the transposon (Fig. 8) [two nucleotides for the *Tc1* and *Tc3* elements (van Luenen et al. 1994), and three nucleotides inwards the element in case of *mariner* (Lampe et al. 1996) and SB (Luo et al. 1998)]. Thus, transposon excision leaves behind three-nucleotide-long 3'-overhangs in SB transposition. The double strand DNA breaks generated by transposon excision are repaired either by the non-homologous end joining pathway (NHEJ), or by template-dependent gap repair (Engels et al. 1990; Plasterk 1991; Lohe et al. 2000; Izsvák et al. 2004). NHEJ generates transposon "footprints" (Fig. 8) that are therefore identical to the first or last three nucleotides of the transposon in SB transposition (Luo et al. 1998; Izsvák et al. 2004). In V(D)J recombination, the single-strand nick is converted into a double-strand break by a transesterification reaction in which the free 3'-OH attacks the opposite strand, thereby creating a hairpin intermediate (van Gent et al. 1996; Gellert 2002). *Tn5* and *Tn10* transposons also transpose *via* a hairpin intermediate, with the difference that the hairpin is on the transposon and not on flanking DNA (Kennedy et al. 1998; Bhasin et al. 1999). In contrast to V(D)J recombination, the excision sites do not have a hairpin structure in SB transposition (Izsvák et al. 2004). Whether second-strand cleavage occurs by transesterification or by hydrolysis in SB transposition needs to be investigated.

1.5.1.2.3.2 Transposon integration

The second step of the transposition reaction is the transfer of the exposed 3'-OH transposon tip to the target DNA molecule by transesterification (Figs. 6 and 8). Similarly to the initial

DNA cut, the strand transfer is done by a nucleophilic attack. In this case, the 3'-OH groups of the already liberated transposon ends serve as a nucleophile that couples the element to the target, without previous target DNA cleavage. As a result, the transposon ends are covalently attached to staggered positions: one of the transposon ends joining to one of the target strand, the other end joining to a displaced position of the target strand. Similarly to the initial strand cleavage, the strand transfer reaction does not need an external energy source, which suggest that it is the energy of the target phosphodiester bond that is used for the new transposon-target joint (Mizuuchi 1992). Although the initial excision and the strand transfer reactions are isoenergetic, many transposons such as *Tn7* and the P element, need molecules with high-energy bonds (ATP and GTP, respectively) for transposition *in vitro*. However, these molecules do not serve as an energy source, rather they only play regulatory roles (Bainton et al. 1991; Kaufman and Rio 1992).

The final steps of transposition reaction are performed by host proteins. Due to the staggered way of insertion during the strand transfer step, there are short, single stranded gaps flanking the new integrant (Fig. 8). Host DNA repair factors then repair these gaps generating characteristic short direct repeats, the hallmarks of transposition. Similar to most other transposable elements, SB does not integrate randomly into target DNA, and displays a certain degree of specificity in target site utilization. Namely, SB exclusively integrates into TA dinucleotides that are duplicated upon transposition, and flank the integrated element (referred to as target site duplication, Fig. 8) (Ivics et al. 1997).

1.6 Regulation of transposition

Transposition is under strict control, and the accumulation of hundreds or thousands of transposon copies can take millions of years. This is because transposition can potentially endanger the survival of the host organism and, consequently, that of the transposable element. Therefore, transposons and their hosts have coevolved, and developed strategies that

reduce the negative effects on the host but ensure proliferation of the element. The mechanisms of transpositional regulation are diverse (Fedoroff 2002). The expression of transposon-encoded factors can be kept in check by transcriptional silencing, through DNA methylation. Post-transcriptional silencing can also contribute to downregulation of factors that are required for transposition. For example, RNA interference (RNAi) suppresses gene expression in a wide variety of organisms, and has been shown to constitute an antiviral defence mechanism in plants. RNAi has been proposed to be a major mechanism for transposon silencing in nematodes (Sijen and Plasterk 2003; Vastenhouw and Plasterk 2004), and perhaps contributes to transposon regulation in other animals as well. Another form of regulation is site-selective insertion of transposons into “safe” places in the genome. For example, the Ty LTR-retrotransposable elements in yeast show considerable site-specificity of insertion, thereby reducing the negative impact of transposition. Transposition of the bacterial transposon *Tn5* and that of P elements and *mariner* elements in *Drosophila* can be regulated by repressor proteins, which are truncated or point mutant versions of the transposase polypeptide (Hartl et al. 1997). For example, defective transposases can compete with wild-type transposase for binding sites located in the transposon ends. Furthermore, P element transposition is restricted to the fly germline, by alternative splicing of the transposase mRNA (Laski et al. 1986). A particularly interesting feature of the bacterial *Tn7* element is that it does not insert into DNA that already contains a copy of *Tn7*, a phenomenon called target immunity. Target immunity helps to avoid multiple copies of the element in the same DNA molecule, which might result in deleterious recombination between the two elements (Peters and Craig 2001). Taken together, there is a great variety of mechanisms which put a limit on transpositional activity. The outcome of this regulation is that transposable elements move at very low frequencies in natural populations.

1.7 Transposons as genetic tools

Genome sequences of many model organisms of developmental or agricultural importance are becoming available. The tremendous amount of sequence data is fuelling the next phases of challenging research: annotating all genes with functional information, and devising new ways for the experimental manipulation of vertebrate genomes. Transposable elements are known to be efficient carriers of foreign DNA into cells. Importantly, the transposase gene

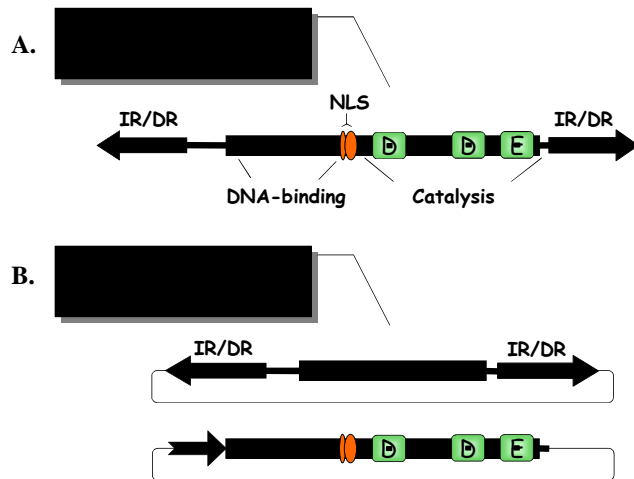


Figure 9. The *Sleeping Beauty* transposon system. (A) Structure of the *Sleeping Beauty* transposon. The central transposase gene (purple box) is flanked by terminal inverted repeats (IR, black arrows) that contain binding sites for the transposase (white arrows). The transposase consists of an N-terminal DNA-binding domain, a nuclear localization signal (NLS) and a catalytic domain characterized by the DDE signature. **(B) Gene transfer vector system based on *Sleeping Beauty*.** The transposase coding region can be replaced by a gene of interest (yellow box) within the transposable element. This transposon can be mobilized if a transposase source is provided in cells; for example, the transposase can be expressed from a separate plasmid vector containing a suitable promoter (black arrow).

can be physically separated from the IRs, and replaced by other DNA sequences (Fig. 9). These transposase-deficient elements can be mobilized if the transposase is provided *in trans*; thus, it is possible to stably integrate a desired DNA molecule into the genome using transposable elements as transgene vectors in a controlled manner (Ivics and Izsvák 2004). This represents the basis of utilizing

transposable elements as transgene vectors; essentially any DNA of interest can be cloned between the IRs, and mobilized by supplying the transposase function in cells (Fig. 10).

P element and Tc1 transposon-based vectors have been extremely valuable in exploring gene function in the invertebrate model organisms *Drosophila melanogaster* and *Caenorhabditis elegans*, respectively (Cooley et al. 1988; Zwaal et al. 1993). However, efficiently manipulating vertebrate genomes with TEs was until recently not feasible. This is because, unfortunately, vertebrate model organisms seem to lack active, endogenous DNA transposons like P and Tc1; the only exception so far is the *Tol2* element in the medaka fish (*Oryzias latipes*) (Koga et al. 1996). To address this problem, a variety of invertebrate TEs,

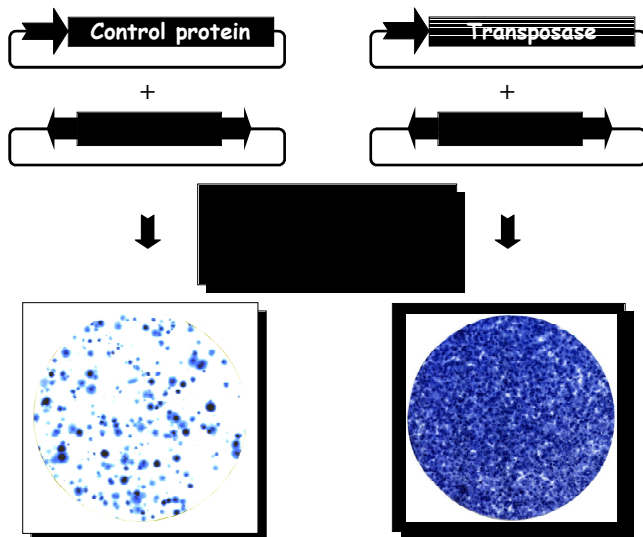


Figure 10. Transposition in tissue culture. The transposon containing a selectable antibiotic resistance gene (neo) is transfected either with or without a transposase-expressing helper plasmid. Transfected cells are placed under antibiotic selection. The dramatic increase in the number of resistant cell colonies in the presence of transposase is the result of transposition of the element from the plasmid vector into chromosomes.

including Tc1/*mariners*, were adopted for gene transfer in vertebrates. However, invertebrate transposons tend to have moderate activity in vertebrates (Fischer et al. 2001), most likely due to restricting activities, or to the lack of specific cofactors [e.g. (Rio et al. 1988)]. Molecular reconstruction of *Sleeping Beauty* represents a milestone in transposon-based technologies that

expanded our abilities in genome manipulations, including insertional mutagenesis, transgenesis and gene therapy, in vertebrate organisms.

1.7.1 Insertional mutagenesis

Alongside with computational approaches and gene expression studies, mutational analysis is the most straightforward way of identifying gene function. One approach of creating mutants is to target and disrupt a gene of interest by homologous recombination; also referred to as reverse genetics. However, in spite of our growing acquaintance with protein domains, protein-protein interactions and molecular structures, our knowledge is yet inadequate to reliably predict the biological process that will be affected by knocking out a particular gene.

Another approach of obtaining mutant phenotypes is to introduce loss-of-function mutations into genomes of model organisms in a random and genome-wide fashion, termed forward genetics. Mutagenesis efforts have been carried out mainly based on X-ray irradiation and chemicals. However, it turned out that X-ray irradiation can cause a variety of chromosomal rearrangements affecting several genes simultaneously, which makes the identification of functions of individual genes difficult. Ethylnitrosourea (ENU) is a potent

chemical mutagen that primarily introduces point mutations into DNA (Russell et al. 1979). Two large-scale mutagenesis screens have been performed in zebrafish (*Danio rerio*) using ENU (Driever et al. 1996; Haffter et al. 1996), and it is routinely used in functional genetic analyses of the mouse genome (Kile et al. 2003). The major advantages of ENU are easy use and highly efficient mutagenic rates in high-throughput screens. Nonetheless, a common disadvantage of these mutagenesis approaches is the time consuming and labor intensive molecular identification of the affected genes by positional cloning. While in some cases mutant phenotypes implicate certain signal transductional or developmental processes or genes, such a candidate gene approach can only be used in a fraction of the mutants. There are >20,000 genes in mammals (Lander et al. 2001), which necessitates the development of methods for rapid identification and functional annotation of genes.

An alternative approach of introducing mutations into the genome is insertional mutagenesis. Discrete pieces of foreign DNA can be harnessed to disrupt host gene function by creating random insertions in the genome. As opposed to chemical mutagenesis, inserting DNA fragments into genes simultaneously provide a molecular tag, which can be used to rapidly identify the mutated allele. Viral and non-viral technologies have been devised to facilitate the penetration of transgenes through biological membranes. Non-viral methods, including naked DNA injection, electroporation, liposomes, “gene-guns” can be useful to introduce DNA into the cells, but chromosomal integration of the introduced DNA is still very inefficient. Moreover, a common drawback of the integration created by these techniques is the concatamerization of the foreign DNA at the insertion locus. Such events can facilitate chromosomal rearrangements (Babinet et al. 1989), aberrant splicing, heterochromatin formation, gene silencing (Garrick et al. 1998), and can interfere with cloning. The above problems can be circumvented by using retroviruses. The overt advantage of using viruses as vehicles for delivering DNA into cells is their capability to penetrate membranes and to catalyze the integration of single copies of the proviral DNA into chromosomes. However,

retroviruses have pronounced preferences for their sites of integration (Bushman 2003), thereby limiting the spectrum of mutations. Moreover, retroviral vectors have limited packaging size and, due to their long terminal repeats, they can induce gene silencing (Garrick et al. 1998) and ectopic reporter gene expression. Additionally, the observations coming from mutagenesis screens in zebrafish suggest that virus-based techniques are labor-intensive, and achieving high-throughput requires a large facility for screening (Amsterdam et al. 1999). Therefore, as an alternative approach to viruses, techniques of transposon-based whole-genome manipulation launched a new wave of research in functional genomics.

1.7.1.1 Cut-and-paste elements in functional genomics

Cut-and-paste DNA elements have been routinely used for studying bacterial, fungal and plant genes in forward genetic screens. Similarly to retrovirus-based methods transposons can be utilized for insertional mutagenesis, followed by the easy identification of the mutant gene. However, DNA transposons have several advantages compared to the above approaches. For example, unlike proviral insertions, transposons can be remobilized *in trans*. Thus, instead of performing time-consuming microinjections, it is possible to generate *de novo* transposon insertions by simply crossing stocks transgenic for the two component of the transposon system (transposon and transposase). This scenario is especially useful when transposition events are directed to the germline of the experimental animal in order to mutagenise germ cells. Also, transposase expression can be directed to particular tissues or developmental stages by a variety of specific promoters. Furthermore, remobilization of a mutagenic transposon out of its insertion site can be used to isolate revertants and, if transposon excision is associated with a deletion of flanking DNA, it can be used to generate deletion mutants. Since transposon are composed of DNA and can be maintained in plasmids, they are much safer and easier to work with than highly infectious retroviruses. Furthermore, timing of

transposase activity is feasible by supplying the transposase in the form of DNA, mRNA or protein in the desired experimental phase.

When transposons are used in insertional mutagenesis screens, transposon vectors often comprise three major classes of constructs to identify the mutated genes rapidly (Fig. 11). These contain a reporter gene, which should be expressed depending on the genetic context of the integration. These vectors are only expressed if they land in-frame in an exon or close downstream to a promoter of an expressed gene. In *polyA traps*, the marker gene lacks a polyA signal, but contains a splice donor (SD) site. Thus, when integrating into an intron, a fusion transcript can be synthesized comprising the marker and the downstream exons of the trapped gene. *Gene traps* (or exon traps) lack promoters, but are equipped with a splice acceptor (SA) preceding the marker gene. Reporter activation occurs if the vector is integrated into an expressed gene, and splicing between the reporter and an upstream exon takes place. The gene trap and polyA trap cassettes can be combined. In that case, the marker of the polyA trap part is amended with a promoter so that the vector can also trap downstream

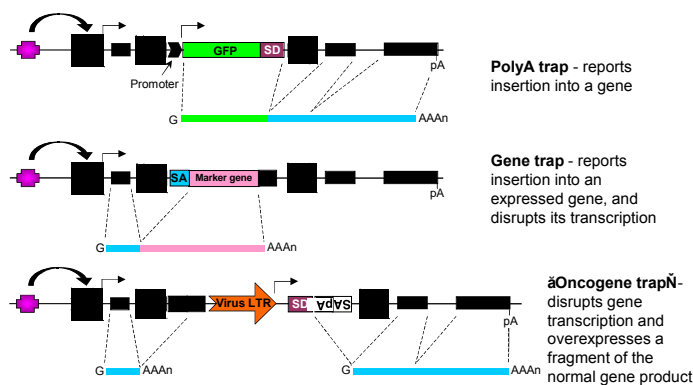


Figure 11. Transposon-based gene trapping vectors. On top, a hypothetical transcription unit is depicted with an upstream regulatory element (purple box), a promoter (red arrow), three exons (blue boxes) and a polyadenylation signal (pA). Major classes of transposon-based trapping constructs and spliced transcripts are shown below. Transposon inverted repeats are indicated by black arrows, different promoters are depicted as green arrows, SD and SA represent splice donor and splice acceptor sites, respectively.

exons, and both upstream and downstream fusion transcripts of the trapped gene can be obtained (Zambrowicz and Friedrich 1998).

The above constructs also offer the possibility to visualize spatial and temporal expression patterns of the mutated genes by using *LacZ* or fluorescent proteins as markers.

The *Minos* transposase has also been shown to mobilize nonautonomous *Minos* elements in mice by transposase expression in the oocytes using ZP3 (Drabek et al. 2003) and in the lymphocytes using CD2 promoters (Zagoraiou et al. 2001). *PiggyBac* has also been

used in coinjection experiments in mice (Ding et al. 2005). The activity of *Tol2* element has already been demonstrated in mouse embryonic stem (ES) cells (Kawakami and Noda 2004) and *in vivo* in the mouse liver (Balciunas et al. 2006).

1.7.1.1.1 Insertional mutagenesis with *Sleeping Beauty*

SB transposition is efficient in cells of different vertebrate classes in tissue culture (Izsvák et al. 2000; Huang et al. 2006) and in somatic as well as germline tissues of fish (Davidson et al. 2003; Grabher et al. 2003; Balciunas et al. 2004), frogs (Sinzelle et al. 2006; Yergeau and Mead 2007), mice (Yant et al. 2000; Dupuy et al. 2001; Fischer et al. 2001; Carlson et al. 2003; Horie et al. 2003; Geurts et al. 2006) and rats (Kitada et al. 2007; Lu et al. 2007) *in vivo*. Therefore, SB is a valuable tool for functional genomics in several model organisms (Miskey et al. 2005; Mates et al. 2007).

In zebrafish, SB and Tol2 have been shown to be useful for insertional mutagenesis in coinjection experiments (Davidson et al. 2003; Balciunas et al. 2004; Kawakami et al. 2004; Parinov et al. 2004; Sivasubbu et al. 2006). SB has also been successfully used for forward genetics approaches in the mouse. Double transgenic mouse lines were generated bearing

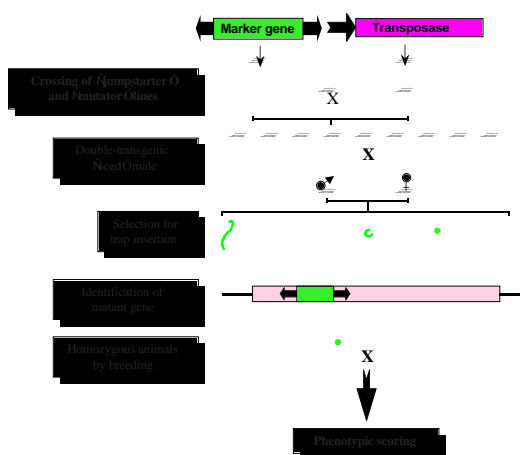


Figure 12. *In vivo* germline mutagenesis of the mouse with transposable elements. Breeding of “jumpstarter” and “mutator” stocks induces transposition in the germline of double-transgenic “seed” males. The transposition events that take place in germ cells are segregated in the offspring. Animals with transposition events need to be bred to homozygosity in order to visualize the phenotypic effects of recessive mutations. Mutant genes can easily be cloned by different PCR methods making use of the inserted transposon as a unique sequence tag.

chromosomally present transposons and an either ubiquitously (Dupuy et al. 2001; Horie et al. 2001; Carlson et al. 2003; Horie et al. 2003) or male germline-specifically (Fischer et al. 2001) expressed transposase gene (Fig. 12). Segregating the transposition events by mating the founder males to wild-type females (Fig. 12) revealed that up to 80% of the progeny can carry transposon insertions (Horie et al. 2001), and a single sperm of a founder can contain, on average, two

insertion events (Dupuy et al. 2001). Additionally, subsequent studies elegantly showed that the germline of such a founder can harbor approximately 10,000 different mutations (Horie et al. 2003).

All the vectors used in vertebrate insertional mutagenesis to date are versions of gene trapping insertional mutagenic constructs (Fig. 11), equipped with elevated mutagenicity and other useful properties. The mutagenicity of gene trap vectors is higher than that of simple insertional vectors, and they enable easy identification of the mutagenized gene by RT-PCR of composite transcripts made up by sequences of the insertional vector and the endogenous gene. Indeed, transposition of gene trap transposons identified mouse genes with ubiquitous and tissue-specific expression patterns, and mutant/lethal phenotypes were easily obtained by generating homozygous animals (Carlson et al. 2003; Horie et al. 2003). Similarly to the GAL4/UAS system in *Drosophila*, a conditional, tetracycline-regulated system has been shown to be applicable to TE-mediated insertional mutagenesis in mice (Geurts et al. 2006).

As an alternative to the loss-of-function approaches, targeted over- and/or misexpression has been shown to be efficient in somatic tissues of mice using SB. Viral enhancer-promoter elements incorporated into SB vectors (Fig. 11) were shown to be useful to induce cancer in experimental animals (Collier et al. 2005; Dupuy et al. 2005). These screens can also capitalize on TEs with an intronic preference of insertion, such as members of the Tc1 family. In order to devise customized screens for cancer development, a current approach is pointing towards establishing mouse lines conditionally expressing the transposase (Dupuy et al. 2006). One approach is to express the transposase from tissue-specific promoters. The second is to generate a Cre recombinase-inducible transposase allele, and take advantage of the many existing Cre strains to induce mutagenesis in specific tissues in mice (Dupuy et al. 2006).

1.7.1.1.1 Local hopping

The studies described above established very efficient transposition of SB in the mouse germline, showed no integration preference with respect to gene structure (Carlson et al. 2003), but revealed that SB tends to reintegrate to sites that are relatively close to the donor locus, a phenomenon called local hopping. Local hopping of SB was first described in mouse ES cells (Luo et al. 1998), and then observed in the mouse germline, indicated by cosegregation of new transposon insertions with their donor sites (Dupuy et al. 2001; Fischer et al. 2001). Additional data showed that most of the reintegration events occur within 3 Mb (Horie et al. 2003), and that the total transposition interval of local hopping is between 5-15 Mb, which is significantly broader than the 100 kb local hopping interval of the P element (Tower et al. 1993). Given that *Minos* has also been found to exhibit preference for local transposition in mice (Drabek et al. 2003), the phenomenon seems to be a general property of the Tc1/*mariner* family.

Local hopping offers the possibility to direct extensive insertional mutagenesis to gene clusters and particular chromosomal regions (Keng et al. 2005). The feasibility of such application has recently been demonstrated by generating four mutant mice having different transposon insertions in a single gene. Does local hopping interfere with the intention to perform genome-wide transposon mutagenesis from a limited number of donor sites? The high number of transposition sites in the germ cells of founder mice and the fact that approximately every fourth excised transposon can be randomly reintegrated into chromosomes other than the donor chromosome suggest that whole-genome mutagenesis is feasible (Horie et al. 2003). Alternatively, the problem of local hopping can be circumvented by injecting SB transposons and transposase mRNA into one-cell mouse embryos. In this case, integration into any chromosome has equal likelihood (Dupuy et al. 2002). These results may also indicate that local hopping is not an intrinsic feature of the transposition machinery, but is due to unequal availability of the different chromosomes as a transposition target.

1.7.2 Transgenesis

The other major field of applications of transposon-based technologies is somatic and germline transgenesis. Transposon-based technologies can be exploited for gene transfer in cultured cells (Fig. 10). Once integrated, transposase-deficient nonautonomous transposons are stable in the absence of the transposase. Transposons can be harnessed to integrate plasmid-based siRNA expression cassettes into chromosomes to obtain stable knockdown cell lines by RNA interference (Heggestad et al. 2004; Kaufman et al. 2005). Also, TEs hold potentials for generating transgenic model organisms, or animals of agricultural and biotechnological importance.

Classical methods to express foreign genes in vertebrates rely on microinjection of nucleic acids into oocytes or fertilized eggs. Two main drawbacks of these approaches are the low rates of genomic integration, and that the injected DNA generally integrates as a concatemer. Both drawbacks can be circumvented utilizing transposition-mediated gene delivery, as it can increase the efficiency of chromosomal integration and facilitates single-copy insertion events. Single units of expression cassettes are presumably less prone to transgene silencing than the concatemeric insertions created by classical methods. Retroviral vectors are also useful tools for the same purpose, but their integration pattern is potentially more mutagenic, due to their preference for the 5'-end of transcription units [reviewed in (Bushman et al. 2005)]. In case of transgenesis, a single-copy insertion away from endogenous genes is clearly desired. The insertional spectrum of *Tc1/mariner* elements satisfies this need the best, as these elements integrate randomly at the genome level, and do not show pronounced bias for integration into genes. Another particular problem concerning transgenesis is that founders that develop from the injected oocytes or eggs are predominantly mosaic for the transgene, because integration generally occurs relatively late during embryonic development. Therefore, in order to potentiate successful transmission of the transgene through the germline to the next generation, it is necessary to shift the window of

integration events as early as possible. This can be facilitated by co-injection of engineered transposons with transposase mRNA. This method has been employed to generate transgenic zebrafish with Tc3 (Raz et al. 1998), *Mos1* (Fadool et al. 1998), *Tol2* (Kawakami et al. 2000) and SB (Nasevicius and Ekker 2000) transgenic *Xenopus* with SB (Sinzelle et al. 2006) and *Tol2* (Hamlet et al. 2006) and transgenic mice with SB (Dupuy et al. 2002; Carlson et al. 2005; Wilber et al. 2006). The far end on the scale of transposition-based somatic gene transfer is human gene therapy. Indeed, a large body of work has already been done in mice investigating possibilities of transposon-based human gene therapy.

1.7.3 *Sleeping Beauty*: an integrating, nonviral gene delivery vector for gene therapy

Considerable effort has been devoted to the development of gene delivery strategies for the treatment of inherited and acquired disorders in humans. For effective gene therapy it is necessary to: 1) achieve delivery of therapeutic genes at high efficiency specifically to relevant cells, 2) express the gene for a prolonged period of time, 3) ensure that the introduction of the therapeutic gene is not deleterious. There are several methods and vectors in use for gene delivery for the purpose of human gene therapy (Verma and Somia 1997). These methods can be broadly classified as viral and non-viral technologies, and all have advantages and limitations, none of them providing a perfect solution.

Adapting viruses for gene transfer is a popular approach, but genetic design of the vector is restricted due to the constraints of the virus in terms of size, structure and regulation of expression. In addition, safety, immunogenicity and production issues hamper clinical progress (Dobbelstein 2003; Thomas et al. 2003). For example, onco-retroviral and lentiviral vectors are efficient at integrating foreign DNA into the chromosomes of transduced cells, and have enormous potential for life-long gene expression (Sinn et al. 2005). However, there are several other considerations including safety (VandenDriessche et al. 2003); preferential integration of retroviral and lentiviral vectors into expressed genes (Scherdin et al. 1990)

poses the risk of inadvertent oncogene activation and congruent development of cancer. In addition, the requirement of cell replication for integration limits the use of retroviral vectors to dividing cell types. Adenovirus vectors have been shown to be capable of *in vivo* gene delivery of transgenes to a wide variety of both dividing and non-dividing cells, as well as mediating high level transgene expression. However, adenoviruses lack the ability to integrate the transferred gene into chromosomal DNA, and their presence in dividing cells is short-lived. Whereas early generation adenoviral vectors still contained residual viral backbone genes that contributed to inflammatory immune responses, toxicity and short-term expression, the latest generation adenoviral vectors (so-called gutless or helper-dependent adenoviral vectors) do not contain any residual viral genes and hence have a significantly improved safety and expression profile compared to early generation adenoviral vectors (Schiedner et al. 1998; Ehrhardt and Kay 2002). Nevertheless, even these latest generation adenoviral vectors still activate the innate immune system, particularly in larger animals and in patients (Thorrez et al. 2004) by virtue of their interaction with antigen-presenting cells (Chuah et al. 2003). Although long-term transgene expression has been achieved in mouse models using gutless adenoviral vectors, expression is typically transient in larger animal models. Hence, repeated vector administration would be required to boost expression levels, but the induction of a humoral (and possibly also cellular) (Kafri et al. 1998) immune response against the capsid proteins precludes vector readministration. Adeno Associated Virus (AAV) vectors have several potential advantages to be explored, including the ability to transduce both dividing and non-dividing cells and the potential for stable transgene expression, even in large preclinical animal models, including non-human primates. Limitations of AAV include low maximal insert size, preferential integration into genes, and the induction of chromosomal rearrangements at the site of insertion (Miller et al. 2002). Moreover, AAV administration in patients has been associated with the induction of a possible cellular immune response directed against the processed AAV capsid antigens (Zaiss and Muruve 2005), leading to

transient and acute hepatotoxicity and precluding long-term transgene expression (Chuah et al. 2004; Manno et al. 2006).

Problems associated with virus vectors have led to an emphasis on development of non-viral methods (Abdallah et al. 1995; Li and Ma 2001; Niidome and Huang 2002; Glover et al. 2005). DNA condensing agents, liposomes, microinjection, electroporation and “gene guns” might be easier and safer to use than viruses. Advantages of non-viral systems include their reduced immunogenicity, no strict limitation of the size of therapeutic expression cassette and improved safety/toxicity profiles. In addition, non-viral vectors are easier and less expensive to manufacture; for example, plasmid-based vectors can be produced in bacteria such as *E. coli*. However, non-viral approaches have been suffering from inefficient delivery, lack of chromosomal integration and resulting transient transgene expression. Recent advances indicate that efficient, long-term gene expression can be achieved by non-viral vectors based on transposable elements.

Transposable elements represent nonviral vector systems that possess the capacity to stably integrate into the genome, and thus provide long-lasting expression of transgene constructs in cells. SB is the most thoroughly studied vertebrate transposon to date, and it has been shown to provide long-term transgene expression in preclinical animal models [see (Ivics and Izsvak 2006)] for a recent review). Since, unlike viruses, transposons are not infectious, they have to be actively delivered into the cell. Various methods for non-viral DNA delivery including hydrodynamic injection, electroporation, microinjection and complexing of the transposon components with PEI, have been tested in conjunction with transposable element vectors [reviewed in (Ivics and Izsvak 2006)]. Alternatively, transposon vectors can be delivered into cells by coupling the integration machinery of the transposable element to the cell infection machinery of a virus. Transposon-virus hybrid vectors delivering the components of the SB transposon system into cells by infection of adenovirus (Yant et al. 2002) or herpes simplex virus (Bowers et al. 2006) have been developed.

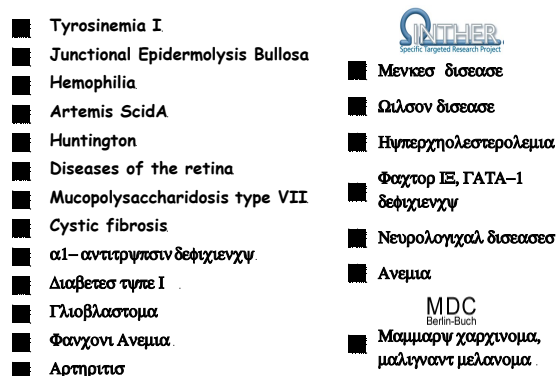


Figure 13. Current preclinical gene therapy experiments using *Sleeping Beauty*.

The past couple of years have seen a steady growth in interest in applying the SB system for the treatment of a number of conditions including haemophilia A and B (Yant et al. 2000; Ohlfest et al. 2005; Liu et al. 2006), junctional epidermolysis bullosa (Ortiz-Urda et al. 2002), tyrosinemia I

(Montini et al. 2002), glioblastoma (Ohlfest et al. 2005), Huntington disease (Chen et al. 2005) and type 1 diabetes (He et al. 2004) (Fig. 13). In addition, important steps have been made towards SB-mediated gene transfer in the lung for potential therapy of α -1-antitrypsin deficiency, cystic fibrosis and a variety of cardiovascular diseases (Belur et al. 2003; Liu et al. 2004) (Fig. 13). Thus, the establishment of non-viral, integrating vectors generated considerable interest in developing efficient and safe vectors for human gene therapy (Izsvák and Ivics 2004; Essner et al. 2005; Hackett et al. 2005).

1.7.4 Target site selection of integrating gene transfer vector systems

About 23 % of gene therapy clinical trials have used retroviral and lentiviral vectors based on the murine leukemia virus (MLV), the avian sarcoma-leukosis virus (ASLV) or the human immunodeficiency virus (HIV) (<http://www.wiley.co.uk/genmed/clinical/>). However, with any vector that integrates into chromosomes in a nearly random manner comes the potential

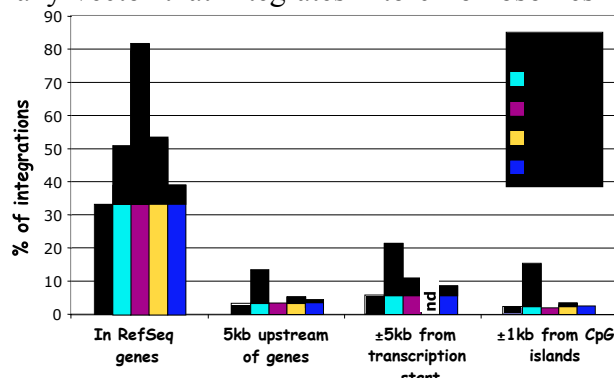


Figure 14. Genomic insertion preferences of integrating vector systems.

risk of insertional activation or inactivation of cellular genes (Baum et al. 2004). MLV has been shown to have a strong tendency to insert into transcription start sites of genes (Wu et al. 2003), whereas HIV exhibits a bias

towards insertions into transcription units but without bias to transcription start sites (Schroder et al. 2002) (Fig. 14). ASLV shows the weakest preference for insertion into active genes in this group, but still at a frequency higher than that of random integration (Mitchell et al. 2004) (Fig. 14). Integration of the vector into a gene or its regulatory elements can knock out the gene, alter its spatio/temporal expression pattern or lead to truncation of the gene product (Fig. 15). Such genotoxic effects can have devastating consequences for the cell and the whole organism, including the development of cancer (Baum et al. 2004). Such unfortunate events were observed in clinical trials using an MLV-based vector for gene therapy of X-linked severe combined immunodeficiency (SCID-X1). 9 out of 11 patients could be cured upon *ex vivo* transfer of a gene construct encoding the γ chain of the common cytokine receptor (γ_c) into autologous CD34+ bone marrow cells (Hacein-Bey-Abina et al. 2002). However, several years after the gene therapy treatment, two patients developed T-cell leukaemia. In both patients, development of the leukaemia was due to insertion of the transgene close to the promoter region of the LIM domain only 2 (*LMO2*) gene (Hacein-Bey-Abina et al. 2003), and deregulated cell proliferation driven by retrovirus enhancer activity on the *LMO2* promoter. Since then, the number of severe adverse events in this particular clinical trial has grown to four (Baum 2007), and yet a new case has been reported in a separate

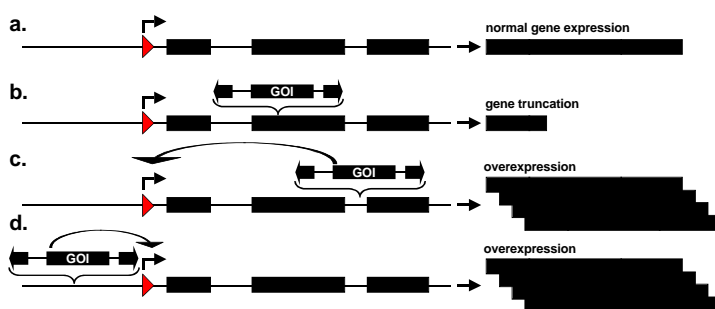


Figure 15. Possible mutagenic consequences of transgene integration in or close to a transcription unit. (a) The figure depicts a hypothetical transcription unit with a promoter (red arrow) and three exons. Normal gene expression results in physiological levels of the correctly spliced protein. (b) A gene of interest (GOI) carried by an integrating vector inserts into an exon, thereby resulting in a truncated gene product. The black arrows flanking the GOI represent retroviral long terminal repeats or transposable element terminal inverted repeats. (c) Transgene insertion occurs in an intron. An enhancer linked to the GOI upregulates transcription of the endogenous gene, resulting in overexpression and/or ectopic expression. (d) Transgene insertion occurs upstream of the targeted gene. An enhancer linked to the GOI upregulates transcription of the endogenous gene, resulting in overexpression and/or ectopic expression.

SCID-X1 trial (Thrasher and Gaspar 2007). These incidents very drastically underscored the peril of insertional mutagenesis upon transgene integration.

Taken together, potential genotoxic effects elicited by integrating viral vector systems give rise to serious risk for

patients undergoing gene therapy. Targeted integration of the therapeutic gene to a “safe” site in the human genome would prevent possible hazards to the host cell and organism due to the problems mentioned above.

1.7.4.1 Naturally occurring specificity in target site selection of integrating genetic elements

1.7.4.1.1 Site-selectivity in viral integration

As discussed above, most viral vectors show an integration bias towards transcriptionally active regions in the genome. Because no sequence-specific integration preference of the retroviral/lentiviral integrase (IN) protein itself has been observed, biased genomic integration can be due to interaction of the viral components with certain host proteins or recognition of different chromatin states of the chromosomes during integration (Mitchell et al. 2004). For example, in contrast to MLV, the integration pattern of HIV does not correspond to the genomic distribution of DNaseI hypersensitivity sites that are associated with open chromatin found in regions upstream of genes and in active transcription units (Lewinski et al. 2006). Instead, the bias of HIV towards integration into active cellular transcription units was proposed to be due to tethering interactions with cellular proteins rather than to chromatin accessibility. In particular, the cellular lens-epithelium-derived-growth-factor (LEDGF)/p75 was shown to influence HIV target site selection (Ciuffi et al. 2005). LEDGF/p75 acts as a transcriptional co-activator, and interacts with components of the basal transcription machinery (Ge et al. 1998). LEDGF/p75 binds tightly to HIV IN, and drives IN into the nucleus when both proteins are produced at high levels (Llano et al. 2004). LEDGF/p75 is conserved among vertebrate species, indicating that insertion site selection of HIV-derived lentiviral vectors could be maintained among vertebrates (Barr et al. 2005). Cells in which LEDGF/p75 expression is knocked down to <10 % by RNAi are still capable of production of infectious HIV, indicating that LEDGF/p75 is dispensable for virus replication (Llano et al.

2004; Ciuffi et al. 2005), but showed reduced integration into transcribed units as compared to normal control cells.

Adeno-associated virus (AAV) is a single-stranded DNA virus that depends on the protein machinery of a helper virus such as adenovirus or Herpes Simplex Virus (HSV) to enter its lytic cycle (McCarty et al. 2004). In the absence of helper virus, the rep proteins encoded by AAV catalyze chromosomal integration and formation of a provirus. AAV shows a strict sequence-specificity for integration (Table 1). In the absence of helper virus, two of the four rep proteins termed rep78 and rep68 encoded by AAV catalyze integration at a single locus named *AAVSI* on human chromosome 19. The exact mechanism of site-specific integration of AAV is still unknown. The viral components involved in targeted DNA integration include the inverted terminal repeats (ITRs) and either the rep68 or the rep78 protein. The ITR spans the terminal 145 nt of the AAV genome and contains a rep-binding element (RBE) and a terminal resolution site (trs). An RBE and a trs-like site can also be found in the *AAVSI* locus in the human genome, and this region is required for site-specific integration of AAV into the human genome (Linden et al. 1996). By binding to both the genomic as well as viral DNA, rep68/rep78 bring the viral genome to close proximity to the *AAVSI* locus (Weitzman et al. 1994). Rep68/rep78 bound to the RBE at *AAVSI* introduce a nick at the trs, and initiate unidirectional DNA synthesis (Urcelay et al. 1995). Rep68 bound to the RBE in the AAV genome also introduces a nick at the viral trs, and viral DNA is integrated into the *AAVSI* locus by template strand switches during unidirectional DNA synthesis (Linden et al. 1996).

1.7.4.1.2 Site-specific recombinases

Sequence-specific DNA integration is also mediated by some recombinases (Table 1). Two groups of recombinases can be distinguished: the serine and tyrosine recombinases that differ in the mechanisms by which they catalyze recombination.

The structural domains of serine recombinases are often spatially separated as opposed to tyrosine recombinases whose domains are interwoven. Cre is a type I topoisomerase from bacteriophage P1 that mediates recombination of DNA between *loxP* sites. Cre has been shown to be active in eukaryotic, including human, cells and is widely used for genome engineering in mice (Yu and Bradley 2001). DNA flanked by *loxP* sites in a direct orientation will be excised and integrated into a *loxP* site previously placed into the human genome (Sauer and Henderson 1990). Recombination at pseudo *loxP* sites (endogenous human DNA

sequences that show similarity to *loxP*) in the human genome occurs with a 4-fold lower efficiency than for wild-type *loxP* sites (Thyagarajan et al. 2000). A directed evolution approach was employed to create a new site-specific Cre recombinase. The newly created recombinase, termed Tre, recombines sequences in the LTRs of integrated HIV proviruses, resulting in excision of the HIV provirus from genomic DNA (Sarkar et al. 2007).

Recombinase	Integration site	Efficiency of targeting	Natural host/origin	Cofactors
Cre recombinase	<i>loxP</i> ; pseudo <i>loxP</i> sites exist in mammalian genomes	0.12% for human <i>lox h7q21</i> (site specific/total integrations) (~100% excision for wt <i>loxP</i> in <i>E. coli</i>)	<i>Escherichia coli</i> (Bacteriophage P1)	No
Flp recombinase	FRT, pseudo sites in mammalian cells unknown	unknown	<i>Saccharomyces cerevisiae</i>	No
ΦC31 integrase	<i>attP</i> ; pseudo <i>attP</i> sites; at least 11 in humans	5% for human <i>psA</i> site; 15% at inserted <i>attP</i> site	<i>Streptomyces lividans</i>	No
Tn7 transposase	Original target site: <i>attTn7</i> in the <i>E. coli glmS</i> gene; <i>gfpt-1</i> and <i>gfpt-2</i> are human orthologs of <i>glmS</i>	Frequency of targeting <i>attTn7</i> in <i>E. coli</i> ~100%; Frequency of targeting <i>gfpt-1</i> (31%) is comparable to targeting <i>glmS</i> (32%) <i>in vitro</i> ; for <i>gfpt-2</i> (23%)	Bacterial	TnsD transposase subunit binding to <i>attTn7</i> in <i>glmS</i>
Ty1	within 750 bp of tRNA or other RNA pol III transcribed genes	90% of insertions	<i>Saccharomyces cerevisiae</i>	Interaction of integrase with TFIII components
Ty3	within 750 bp of tRNA or other RNA pol III transcribed genes	95% of insertions	<i>Saccharomyces cerevisiae</i>	Interaction of integrase with TFIIIB and C
Ty5	silent heterochromatin	21% of insertions adjacent or within transcriptional silencers flanking HML and HMR or subtelomeric repeat in chromosome III	<i>Saccharomyces cerevisiae</i>	Interaction of integrase with Sir4
TRE3	100-150 bp downstream of tRNA genes	100%	<i>Dictyostelium discoideum</i>	Not known
TRE5	50bp upstream of tRNA genes	100%	<i>Dictyostelium discoideum</i>	Interaction of ORF1 with TFIIIB
DIRS	cluster formation on extremities of chromosomes	87% into transposons	<i>Dictyostelium discoideum</i>	Not known
AAV Rep	<i>AAVS1</i>	70-90%	<i>Homo sapiens</i>	Not known

Table 1. Integrating genetic elements showing targeted insertion in their natural hosts.

FLP, a recombinase from *Sacharomyces cerevisiae*, recombines DNA between its recognition sites called *FRT*. Though wild-type FLP shows lower affinity to its target than Cre, mutants created by directed evolution displayed improved performance in human 293

and mouse embryonic stem cells (Buchholz et al. 1998). Both Cre and FLP are bidirectional recombinases that catalyze DNA excision and integration, but favoring the excision reaction. This feature leads to inefficient integration and expression of transgene constructs. Furthermore, genotoxic effects including chromosomal rearrangements and growth inhibition observed for Cre recombinase when expressed persistently at high levels make it a possible hazard to genome integrity (Loonstra et al. 2001). The same holds true for a site-specific IN from the *Streptomyces* phage ϕ C31 (Thorpe and Smith 1998) that catalyzes recombination between so-called attachment (*att*) sites (Table 1). The *attP* site is found in the ϕ C31 genome, whereas *attB* is located in the host *Streptomyces* genome. ϕ C31-mediated integration in human as well as mouse cells frequently occurs into pseudo *att* sites such as *psA* in human or *mpsA* in the mouse genome (Thyagarajan et al. 2001; Chalberg et al. 2006). *PsA* shares 44% identity with *attP* (Ginsburg and Calos 2005). In human 293 cells harbouring an inserted *attP* site, 15% of the integrations were detected at *attP*, 5% of the rest of the integration events occurred at *psA*, 5-10% were random, whereas the rest of integrations was believed to be distributed over the other ~100 pseudo sites in the human genome (Thyagarajan et al. 2001). In several studies, reasonably efficient delivery and stable expression of genes relevant in human genetic diseases (Glover et al. 2005) was achieved in mouse or human cells using ϕ C31 recombinase. However, ϕ C31 is mutagenic, because it can cause chromosomal aberrations due to recombination between pseudo sites or imperfect recombination reactions (Ehrhardt et al. 2006; Liu et al. 2006). It remains to be tested if insertions of transgenes at pseudo sites in the human genome can cause alterations of host gene expression patterns leading to abnormal cell behavior.

1.7.4.1.3 Site-specific transposable elements

Unlike viruses, transposons do not possess envelope genes, and hence lack an extracellular phase in their life-cycle. This makes their fate closely linked to the fate of the host cell, and

may result in integration patterns less mutagenic to the cell. The higher the gene density of a genome, the higher the chance for transposable elements to insert into coding sequences, resulting in potentially fatal consequences to the cell. Significant fractions of genomes with a small proportion of coding regions and extensive intergenic regions can be composed of transposon-derived sequences (e.g., 45% of the human genome), in contrast to organisms having a small genome with high gene density, such as yeast. Ty LTR-retrotransposons in *Saccharomyces cerevisiae* are structurally and functionally related to retroviruses. Integration of Ty1, Ty3 and Ty5 retrotransposons is tethered to certain sites in the genome by host proteins (Table 1). The Ty1 element shows a strong insertion preference for genes transcribed by RNA polymerase III (Pol III). 90% of Ty1 insertions can be found about 1 kb upstream of tRNA genes (Kim et al. 1998). A second preferred integration area of Ty1 is found upstream of the 5S RNA genes that are also transcribed by Pol III (Bryk et al. 1997). Targeting of this site by Ty1 elements may thus depend on the same factors as targeting of the tRNA genes. Indeed, components of the Pol III transcription machinery were found to be required for targeting of Ty1 (Bachman et al. 2005); however, other factors such as chromatin components, physical properties of DNA or subnuclear localization of the target may as well specify integration sites.

Ty3 integrates one or two base pairs upstream of Pol III transcription start sites. TFIIB and TFIIC are important factors for assembly of Pol III complexes at transcription start sites of Pol III-transcribed genes, and are also involved in the recruitment of Ty3 (Kirchner et al. 1995). Though TFIIB is sufficient to target Ty3, TFIIC orientates binding of TFIIB to the TATA box (Yieh et al. 2002), and weakly interacts with Ty3 IN (Aye et al. 2001). The Ty5 element interacts with the host protein Sir4p (Xie et al. 2001), which targets insertions to heterochromatic regions of the genome such as telomers and silent mating locus (Zou et al. 1996). Interaction of Ty5 IN with Sir4p is mediated by its targeting domain, a 6-amino-acid motif at the C-terminus of Ty5 IN. Mutations within this domain abolish

interaction between IN and Sir4p, and result in random integration of Ty5 retrotransposons. Concordantly, random integration of Ty5 is observed in cells deficient in Sir4p (Xie et al. 2001).

Targeting of a specific genomic site may be specified by primary DNA sequence recognized by specific DNA-binding domains (DBDs). In addition, physical properties of the DNA such as kinks due to protein binding, triplex DNA or altered/abnormal DNA structures due to base composition may cause preferential binding of proteins or protein complexes at certain sites. For the bacterial transposon *Tn7*, both sequence- and structure-specific binding apply. The *Tn7* transposon encodes five different proteins: TnsABCD and E. Depending on proteins involved in the transposition process, either a particular DNA structure found during conjugation or a specific site in the bacterial genome is targeted (Peters and Craig 2001). During bacterial conjugation, TnsE seems to recognize DNA structures with recessed 3'-ends during lagging strand DNA synthesis, and directs integration of the transposon to this site. TnsD binds to a specific DNA sequence called *attTn7* in the 3'-end of the bacterial glutamine synthetase (*glmS*) gene in the bacterial genome, followed by insertion of the transposon several base pairs downstream of *glmS* (Table 1). Binding of TnsD creates DNA distortion probably responsible for recruitment of TnsC, which in turn interacts with TnsAB promoting insertion of *Tn7* at *attTn7*. Importantly, *Tn7* inserts into the human homologue of *glmS* in *Escherichia coli* and test tube reactions (Kuduvalli et al. 2005), but *Tn7* transpositional activity in human cells has not been reported.

The eukaryotic microorganism *Dictyostelium discoideum* has a highly compact genome of 34 Mb with 76% coding regions and a surprisingly high transposon load of 10%. Transposons in *D. discoideum* have developed two strategies to avoid genotoxic insertion into coding sequences (Table 1). One of these strategies is nested integrations of transposons forming clusters. For example, the DIRS LTR-retrotransposon family shows no initial target site selectivity, but can be found in few clusters, made up of several copies of themselves

(Loomis et al. 1995), located in centromeric and telomeric regions of chromosomes. The other strategy is targeted integration into “safe” regions of the genome free from protein-coding sequences. This strategy is primarily used by non-LTR retrotransposons that insert up- and downstream of tRNA genes (Winckler et al. 2002). The non-LTR retrotransposons collectively called TRE (tRNA gene-targeting retrotransposable elements) can be divided into two groups: TRE5 elements preferentially integrate about 50 bp upstream of tRNA genes, whereas TRE3 elements favour integration 100-150 bp downstream to tRNA genes. An *in vivo* assay using a reporter gene tagged with a tRNA coding region showed targeted integration of TRE5 in the same manner as in a genomic context, indicating that targeted insertion of TRE5 is dependent on interactions with Pol III transcription factors (Winckler et al. 2005). Indeed, the ORF1 protein encoded by the TRE5 element was recently shown to interact with TFIIB, suggesting a role of this interaction in targeting integration into tRNA genes (Chung et al. 2007). Altogether, these observations suggest a general model wherein interactions between transposase/IN and DNA-bound proteins mediate insertional target choice. In sum, the existence of transposable elements with natural targeting abilities raises promise that recombinase/transposase/IN proteins with target-selective insertion properties can be engineered.

1.7.5 Artificial (imposed) targeting of DNA integration into preselected sequences

None of the vector systems currently used either in preclinical experiments or in clinical trials described above displays DNA sequence preferences specific enough for targeted insertion into a defined location in the human genome. Integration into selected sites in the genome would simultaneously ensure appropriate expression of the transgene (lack of position effects), and prevent hazardous effects to the organism due to insertional mutagenesis of cellular genes (lack of genotoxicity). Targeted gene delivery can rely on distinct molecular strategies. One possibility implies fusion of the recombinase/transposase/IN to a DNA

binding domain. Upon binding of the engineered recombinase to a specific target site, integration of the DNA component of the vector system may occur in adjacent regions. A more indirect approach uses DNA-binding specificity of interacting proteins. Interaction of proteins bound to specific target sequences can tether either the DNA or the protein component of the vector system to this region of DNA, resulting in integration into nearby regions.

1.7.5.1 Targeting through fusion to DNA binding domains

Altering sequence-specificity of most recombinases may prove difficult, since they do not have spatially separated catalytic and target DBDs that could be modularly replaced irrespectively of each other. Target-specificity can potentially be altered by directed evolution (random mutagenesis techniques followed by activity screening under selective conditions) or by substitution of key amino acids implicated in target recognition. Both approaches yielded mutants of proteins showing more relaxed target-site specificity or even a complete shift in target site preference [reviewed in (Collins et al. 2003)]. Engineering of proteins that specifically bind to desired DNA sequences is expected to pose a major challenge, and may not only lead to altered site-specificity but also to impaired or modified catalytic activity. Fusions of proteins to a specific DBD appear to be a much easier and more direct approach (Table 2). However, some proteins display sensitivity to fusions with foreign peptides, domains or proteins, possibly due to altered folding of the resulting chimeric protein. Thus, fusions may result in abolished or limited enzymatic activity. Another factor to consider is that the native DNA-binding capacity of the protein can compete with the foreign DBD of the fusion partner. Requirements for integration of a vector system, such as a TA dinucleotide within an appropriate structural context for the SB transposon, should also be taken into account when selecting a site to be targeted in the genome. Keeping this in mind, fusions

between a DBD and a recombinase protein may overall be a promising approach to targeted gene insertion.

In vitro targeting studies of the IN of avian sarcoma virus (ASV) fused to the DNA binding domain of the *Escherichia coli* LexA protein showed altered insertion patterns and an insertion hot spot near a tandem LexA operator as compared to unfused IN (Katz et al. 1996).

Protein	Efficiency of targeting	Activity of chimeric enzyme	Experimental system
ASV IN/LexA	Significantly enriched integration adjacent to binding site	Full processing activity; similar joining activity	<i>In vitro</i>
HIV IN/ λ -repressor	Enriched integration near binding sites	Fusion retained known activities of HIV IN	<i>In vitro</i>
LexA	Enriched integration near target site	No appreciable change in activities compared to wt	<i>In vitro</i>
Zif268	Different integration patterns when compared to wt with integration hot spots near targeted sites	Infectivity abolished, but restored when mixing in wt IN (up to 93 % of wt only when mixed 1:1)	<i>In vitro</i>
E2C (ZF)	>60 % of insertions within 10 bp of binding site (5 % for wt) for GC-rich strand (for G-rich strand 14 % bzw 32 %; wt 5 %) ~10-fold higher preference for integration near the E2C site as compared to wt IN	no appreciable change in processing and joining activity up to ~20 % activity of viruses compared to wt	<i>In vitro</i> cell culture (HeLa, chromosomal target)
IS30/ cI repressor	More than 10-fold	Similar activity to wt (2.5×10^{-2} vs 2×10^{-2})	<i>E. coli</i> (plasmid target)
Gli1	6 insertions near binding site (5 illegitimate, one legitimate); none found on target lacking the binding site	Low excision and integration efficiency	Zebrafish (plasmid target)
Tn3 resolvase/ Zif268	100 % of plasmid molecules recovered were resolved (indicating efficient binding to binding site and recombinational activity)	44-98 % (of resolved substrate molecules in recovered plasmid DNA)	<i>E. coli</i>
PB/Gal4	67 % of insertions into location ~1 kb upstream of binding site	>10-fold increased transpositional activity (compared to transposition into plasmid lacking target site)	mosquito embryos (plasmid target)
Mos1/Gal4	96 % of insertions into location ~1 kb upstream of binding site	>10-fold increased transpositional activity (compared to transposition into plasmid lacking target site)	mosquito embryos (plasmid target)

Table 2. Targeting of gene delivery systems by direct fusion to DNA-binding domains.

HIV IN fusions to the DBD of phage λ repressor protein (Bushman 1994) or to the DBD of the LexA repressor protein (Goulaouic and Chow 1996) were also capable of targeting integrations near their specific binding sites *in vitro*. These experiments demonstrated the feasibility of using fusions between DBDs and INs to target viral insertions to a certain extent to specific sites.

Transcription factors (TFs) recognize and bind specific DNA sequences followed by recruitment of proteins affecting the transcriptional status of the associated gene. These processes are usually mediated by distinct domains, making it possible to separate these functions. Consequently, the DBD of a TF by itself would preserve its unrestrained DNA binding capacity (specificity and affinity), serving as a potent source as a fusion partner. TFs are typically classified according to the structure of their DBDs, such as zinc finger (ZF), leucine zipper, HTH, helix-loop-helix and high mobility group boxes. One naturally occurring

ZF is the DBD of transcription factor Gli1 present in vertebrates that recognizes and binds a 9-bp DNA sequence. The transposase of the bacterial insertion sequence element *IS30* was fused to either the cI repressor of phage λ or the Gli1DBD, and the resulting fusion proteins showed targeted integration into plasmid targets in *E. coli* and zebrafish (Szabo et al. 2003) (Table 2). This study was the first demonstration that targeted transposition by an engineered transposase could work *in vivo*.

The DBD of the yeast Gal4 TF contains a ZF domain of the Zn₂Cys₆ type. It recognizes a specific, 17-bp DNA sequence called upstream activating sequence (UAS). Fusions of the Gal4 DBD to the *Mos1* (a Tc1/*mariner* transposon from *Drosophila mauritiana*) and *piggyBac* (PB) transposases were tested for their transpositional activities and targeting potentials by applying plasmid-based transposition assays in mosquito embryos (Maragathavally et al. 2006) (Table 2). Transposition mediated by the chimeric *Mos1* transposase into the UAS-containing target plasmid occurred at a 96 % frequency at the same TA located 954 bp away from the targeted UAS sequence. Transposition by the Gal4-PB fusion protein into a plasmid containing the UAS target sequence occurred at a 67 % frequency into a TTAA site located 1103 bp upstream of the UAS. These results present quite efficient targeting by *Mos1*- and PB-Gal4 fusions. Binding of the Gal4 DBD to its recognition site presumably brings the fused transposase to close proximity, thereby enhancing the chance of transposon insertions nearby. Chimeric transposases may structurally be limited after UAS binding, allowing transgene integration into only few sites.

Naturally occurring ZFs also include the three-finger transcription factor Zif286 originally identified in the mouse. A chimeric recombinase composed of the DBD of Zif268 and the catalytic domain of the bacterial *Tn3* resolvase was successfully assayed for targeting of two inverted Zif268 recognition sites flanking a *Tn3 res* site in *E. coli* (Akopian et al. 2003) (Table 2). *Tn3* belongs to the serine recombinases that have spatially separated catalytic and DNA-binding domains. Functionality of the chimeric protein proves that exchange of the

physiological DBD of *Tn3* resolvase with a foreign DBD yields a recombinationally competent enzyme. It remains to be investigated whether such a fusion construct is also functional in eukaryotic cells. Zif268 fusions with the HIV IN were also shown to have biased insertion patterns near specific binding sites *in vitro* (Bushman and Miller 1997).

Naturally occurring DBDs have some limitations for use as gene targeting agents. First, some of the DBDs discussed above are derived from proteins that do not have physiological targets in the human genome; thus, specific target sites would need to be introduced into the genome prior to delivery of a transgene. Second, those DBDs that do have physiological binding sites in the human genome recognize short DNA sequences present in multiple copies throughout the human genome, making targeted insertion with these DBDs impractical (for example, a 9-bp recognition sequence of a ZF would be expected to occur >10,000 times in the human genome). Recognition sites of 18 bp would be expected to be unique in the human genome. Artificial ZFs, especially the C₂H₂ type, offer a potential solution. Their modular character in structure and function is the key advantage in engineering of proteins that are able to recognize theoretically any sequence in the human genome (Mandell and Barbas 2006). Each individual zinc finger binds 3-4 bp DNA, thus a set of 64 domains would cover recognition of any desired DNA sequence. ZF nucleases (ZFNs) consisting of the *FokI* cleavage domain fused to a ZF represent an attractive technology for targeted gene repair by homologous recombination (Lombardo et al. 2007).

Fusions of engineered ZFs to recombinase proteins could enable selective insertion of a transgene into a desired region of the genome. The synthetic E2C ZF protein is a six-finger ZF recognizing an 18-bp target site in the 5'-untranslated region of the human *erbB-2* gene. E2C fusions to transcriptional activator and repressor domains have been used to regulate endogenous *erbB-2* gene expression (Beerli et al. 1998). Fusions of E2C to HIV IN were shown to target retroviral integration near the 18-bp E2C binding site in cell-free reactions (Tan et al. 2004) (Table 2). The E2C/IN fusion protein was then tested for targeting of the

E2C locus in cultured human cells using a quantitative real-time PCR assay showing a ~10-fold increase of insertions near the E2C binding site in the genome as compared to unfused IN. However, virions containing the fusion proteins exhibited poor infectivity ranging from 1 to 24 % compared to viruses containing wild-type IN (Tan et al. 2006).

Taken together, direct fusions of DBDs to integrase/transposase proteins appear to interfere with the production of genetically stable virions (in case of viral vectors) and with the biochemical activities of transposase proteins. Nevertheless, engineered recombinases do show biased insertion patterns near targeted DNA sites *in vitro*, as well as in cultured cells using plasmids as targets. Site-selected transgene insertion by engineered IN and transposase proteins at the genome level remains a challenge.

1.7.5.2 Targeting through interaction with DNA-binding proteins

An alternative approach to target DNA integration is based on employing DNA-binding proteins that interact with either the transposon DNA and/or with the transposase protein. Either naturally occurring or engineered transposon/transposase interactors may tether the transpositional machinery to specific DNA sites, potentially leading to integration into nearby regions (Table 3). As outlined above, there are examples for the existence of such targeting mechanisms in nature. For example, based upon observations for a role of LEDGF/p75 in directing HIV integration into expressed transcription units, *in vitro* studies have shown

Protein	Interaction partner	Efficiency of targeting	Efficiency of transposition	Experimental system
HIV IN	LEDGF/p75 fused to λ repressor	Significantly enriched integration near the λ operator	Wild-type activity	<i>In vitro</i>
Ty 5 IN TD replaced with a Rad9p motif TD replaced with a 12 aa NpwBP motif)	Sir4p fused to LexA	11 %	100 %	Yeast (plasmid target)
	LexA-FHA1	11 %	100 %	Yeast (plasmid target)
	LexA-Npw38	14 %	100 %	Yeast (plasmid target)

Table 3. Targeting of gene delivery systems by fusing a DNA-binding domain to a protein domain that interacts with the recombinase.

increased integration near λ repressor binding sites by fusing either the full-length LEDGF/p75 or the LEDGF/p75 IN-binding domain to the DBD of phage λ repressor protein (Ciuffi et al. 2006) (Table 3). In an analogous fashion, Sir4p (which, as

described above, mediates targeted insertion of the yeast retrotransposon into heterochromatin in yeast) fused to the *E. coli* LexA DBD was shown to result in integration hot spots for Ty5 near LexA operators (Zhu et al. 2003) (Table 3). Domain swaps in recombinase proteins by changing protein-protein interaction domains could also lead to modified integration patterns. Indeed, replacing the targeting domain of Ty5 IN, which interacts with Sir4p, with a heterologous domain interacting with a protein fused to LexA also leads to insertions near the LexA operators (Zhu et al. 2003) (Table 3).

2 MATERIALS and METHODS

Plasmid constructs. The pT/neo transposon donor plasmid and the pCMV-SB transposase-expressing helper plasmid have been described (Ivics et al. 1997). To construct the pT/zeo vector, the *neo* gene in pT/neo was replaced by a zeocin gene and a plasmid origin of replication. In this plasmid, the *zeo* gene is driven by the SV40 enhancer/promoter.

The LexA gene was PCR-amplified from plasmid pEG202 using primers 5' gct gac cgc gga tca tga aag cgt taa cgg cca ggc and 5' agg tgc tcg agc cag tcg ccg ttg cg, digested with *Sac*II and *Xho*I and inserted into the respective sites in the pFV4a expression vector containing the carp β -actin promoter (Liu et al. 1990). A double-stranded oligo encoding the SV40 T antigen nuclear localization signal was inserted downstream of LexA, followed by an in-frame insertion of 44 amino acids (excluding the initiator methionine) of the SAF-box, PCR-amplified using primers 5' gct ata ctc gag agt tcc tcg cct gtt aat gta aaa and 5' gct ata ctc gag cta ctc gtc gtc cag cgc agc c and digested with *Xho*I. The sequence containing the LexA operator site (Fig. 3a) was inserted into pT/zeo322 (Izsvák et al. 2000) using PCR mutagenesis. The tetracycline repressor-LexA fusion was constructed by PCR amplification of LexA using primers 5' gtt cag cta gcg aaa gcg tta acg gcc agg caa c and 5' gtt cag gat cct tac agc cag tcg ccg ttg cg and insertion into the CMV promoter-driven TetR expression vector pUHD141-1 with *Nhe*I and *Bam*HI. The TetR/NLS/N-57 construct was made by PCR amplification of 56 amino acids (excluding the initiator methionine) of the N-terminal HTH domain of the SB transposase using primers 5' gtt cag cta gca ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt gga aaa tca aaa gaa atc and 5' gtt cag gat cct agc ggt atg acg gct gcg tgg, and insertion into pUHD141-1 with *Nhe*I and *Bam*HI. This construct was designed to contain a glycine-bridge between the fusion partners to form a flexible linker between the two functional folding units. The TetR/SB fusion was constructed by replacing a *Sac*II-*Nco*I fragment of FV4a-SB with a corresponding restriction fragment of TetR/NLS/N-57. In both of the previous fusions, the TetR domain and the transposase domain are connected by an

NLS/glycine-bridge linker with the following amino acid composition: PKKKRKLAGGGGGGGGGG.

The SB/E2C fusion was made by PCR amplification of the SB transposase gene with primers 5' gct ata *ccg cgg* atc atg gga aaa tca aaa gaa atc agc and 5' gct ata *ctc gag* acc tcc gcc acc acc tcc gcc acc tcc tcc gta ttt ggt agc att gcc ttt aaa ttg, cutting with *SacII* and *XhoI* and cloning together with an *XhoI/ApaI*-cut E2C gene fragment amplified with 5' gct ata *ctc gag* gcc cag gcg gcc ctg gag ccc ggg gag aag ccc and 5' gct ata *ggg ccc* tca gcc ggc ctg gcc act agt ttt ttt acc ggt g into FV4a. The E2C/SB fusion was constructed by inserting E2C amplified with 5' gct ata *ccg cgg* act atg gcc cag gcg gcc ctc gag ccc and 5' gct ata *gct agc* ccg gcc tgg cca cta gtt ttt tta ccg gtg and cut with *SacII/NheI* into the corresponding sites of FV4a-TetR/SB. The Jazz/SB fusion was done in a similar way, by inserting PCR-amplified, HA-tagged Jazz gene with 5' gct ata *ccg cgg* act atg tat cca tat gat gtt cca gat tat gca agc gat g and 5' gct ata *gct agc* tcg aga tca ttt tgc ctc aaa tg. In these three constructs, the zinc finger domains and the transposase domain are connected by a linker consisting of ten glycine residues.

Cell culture and transfections. Human HeLa cells were cultured in DMEM supplemented with 10% fetal bovine serum. Transposition assays in HeLa cells were done as described (Ivics et al. 1997). Typically, 10⁵ cells were transfected with 90 ng of each the transposon donor plasmid pT/neo or pT/zeo322 or pTzeo-322/LexOP, the transposase-expressing vector pCMV-SB10 (Ivics et al. 1997), and a plasmid expressing a targeting fusion protein, using Eugene6 transfection reagent (Roche). Selection with doxycycline was done at 1 µg/ml as recommended by the manufacturer (Clontech). The transgenic HeLa derived cell line was generated by cotransfection of pTRE-d2EGFP (Clontech) and an SV40-Hygro selectable marker, and selection with hygromycin. One resulting clone that responded well to transactivation by the pTet-Off plasmid (Clontech) was kept for all further experiments.

Frequency of targeted transposition events was assessed by transfecting TRE-transgenic HeLa cells in 6-well plates using 90 ng pT/neo, 450 ng pCMV-TetR/NLS/N-57 and 0.9 ng pCMV-SB10, and plating the cells two days post-transfection directly into 96-well tissue culture plates under conditions that all wells contained at least one transgenic cell. This way, the chance for subclonal propagation of transgenic clones within a transfected cell population is kept at a minimum.

Isolation of integrated transposons from chromosomal DNA by plasmid rescue. For the recovery and analyses of integration target sites, pT/zeo was used as a transposon donor plasmid. Genomic DNA was isolated from zeocin-resistant cell clones using Qiagen's DNeasy tissue kit. Approximately 5 µg DNA was digested with restriction enzymes *NheI*, *SpeI* and *XbaI*, which generate compatible ends and do not cleave within the transposon. Digested DNA was ligated in 500 µl volume with 1200 units of T4 ligase (NEB) at 16° C overnight, and chloroform-extracted following addition of 50 µl 2.5 M potassium acetate (pH 8.0). The ligated DNA was precipitated and electroporated into DH10B *E. coli* cells (Gibco BRL).

Mapping of transposon insertion sites. Cloning of transposon insertion sites from human genomic DNA was done using splinkerette PCR or transposon rescue as described (Ivics et al. 1997). MAR-Wiz (<http://www.futuresoft.org/MAR-Wiz/>) predicts MARs based on the co-occurrence of DNA sequence patterns and structural features, such as AT-richness, bendability, and presence of topoisomerase II recognition sites, that have been shown to occur in the neighborhood of MARs. The default parameters of MAR search were used. Statistical analysis was done using student's t-test.

Structural analyses. B-DNA twist, A-philicity, DNA bendability and protein-induced deformability were measured as previously described (Brukner et al. 1995; Gorin et al. 1995; Ivanov and Minchenkova 1995; Olson et al. 1998). All calculations (except percentage of GC) were done using a three base pair sliding window to incorporate effects of adjacent bases on the various biophysical properties. For B-DNA twist, A-philicity and protein-induced deformability only dinucleotide values were available for determining the sequences' structures, whereas for bendability, trinucleotide values were used. Profiles were averaged over all sequences. DNA sequence at the site of insertion was compared to control DNA sequences, randomly selected from sequences of the relevant host organisms or the plasmid targets. MANOVA (multivariate analysis of variance) was used to measure if differences between the controls and the insertion sites were significant. A commercial software developed by the SAS Institute Inc. was used for statistical analyses, with 90% and 95% confidence levels, to determine significant differences between sample and control data.

Electrophoretic mobility shift assay. The probes used were an *Eco*RI fragment of pTzeo-322/LexOP containing the LexA operator site within the left IR of SB and an *Xho*I/*Sac*II fragment of pTRE-d2EGFP containing the TRE region, in assay conditions described in (Ivics et al. 1997). Reactions contained labelled probe, 1 µg poly[dI][dC], 100 pg labeled fragment and 1 µl nuclear protein extract.

DNase I digestion analysis. The following four, 32-nucleotide oligos and their complementary strands were synthesized, such that there is a three-base, 5'-overhang of the bottom strand. SB-S 5'-gagtaaacttgATATATATgaagtttta-3', SB-AS 5'-ctttaaacttcATATATATccaagtttactc-3'; Tc1-S 5'-gagtaaacttgCACATATGTGgaagtttta-3', Tc1-AS 5'-ctttaaacttcCACATATGTGcaagtttactc-3'; Bad Bender-S (BB-S) 5'-gagtaaacttgAAAAAAAAgaagtttta-3', Bad Bender-AS (BB-AS) 5'-

ctttaaacttcTTTTTTTccaagtttactc-3'; BB2-S 5'-gagtaaacttggAAATAAAgaagtttta-3', BB2-AS 5'-ctttaaacttcTTTATTTTccaagtttactc-3'. The lower case letters represent the identical flanking sequence while the upper case letters represent the test sequence. Each oligo and its complementary strand were mixed in equimolar concentrations and annealed by briefly boiling, then slow cooling. The double-stranded oligos were labeled with α -³²P-dATP and Klenow. The labeled, double-stranded oligos were digested with DNase I as described (Brukner et al. 1995), under conditions that allow less than one cut per molecule. The digests were incubated for 90 seconds at 37°C, stopped with addition of 1 volume loading buffer (80% formamide, 0.1% bromophenol blue, 10 mM EDTA), and separated on an 8% polyacrylamide denaturing gel containing 8 M urea. Sizes were determined by use of an end-labeled Oligo Size Marker (Amersham). The dried gel was exposed to a phosphorimager screen and quantitated with a Molecular Dynamics PhosphorImager System.

Pull-down SAF-binding assay. A DNA-binding assay was done essentially as described (Kipp et al. 2000). The ZZ-45 protein is a recombinant SAF-A peptide expressed as a C-terminal fusion to an artificial double-Z domain tag (tandem of the IgG-binding domain of Protein A). Radiolabelled DNA fragments to be tested and the control DNAs (pMII human DNA as positive control and bacterial pUC19 DNA as negative control) were incubated together with ZZ-45 immobilized on IgG resin in binding buffer. Sheared *E. coli* DNA was added to the binding reactions as an unlabelled unspecific competitor in 1000-fold excess. Unbound DNA was removed by washing the complex extensively with the binding buffer. DNA binding was quantified by scintillation counting.

3 RESULTS

3.1 Target site selection of *Sleeping Beauty*

3.1.1 *Sleeping Beauty* shows a random pattern of integration in the human genome

In order to analyze SB's insertion profile on the genomic level, transposon insertions were generated in human HeLa cells using an *in vivo* transposition assay (Ivics et al. 1997), which is based on mobilization of a zeocin resistance gene (*zeo*)-marked SB element from extrachromosomal plasmids into chromosomes. The only level of selection in recovering transposition events in this assay is that the *zeo* gene within the integrated transposon has to be expressed. 138 insertion sites were identified and mapped on human chromosomes by computer analysis, using NCBI's human genome BLAST service. As shown in Fig. 16, although some chromosomes were hit more frequently than others, no clear preference is apparent for any chromosome, or for certain subchromosomal regions. The Y chromosome is not present in human HeLa cells, thus no hits were recorded. This observation indicates that most (if not all) chromosomes can serve as good targets for transposition. One insertion was found in the 3'-UTR region of a gene, 46 were mapped to intron sequences and one

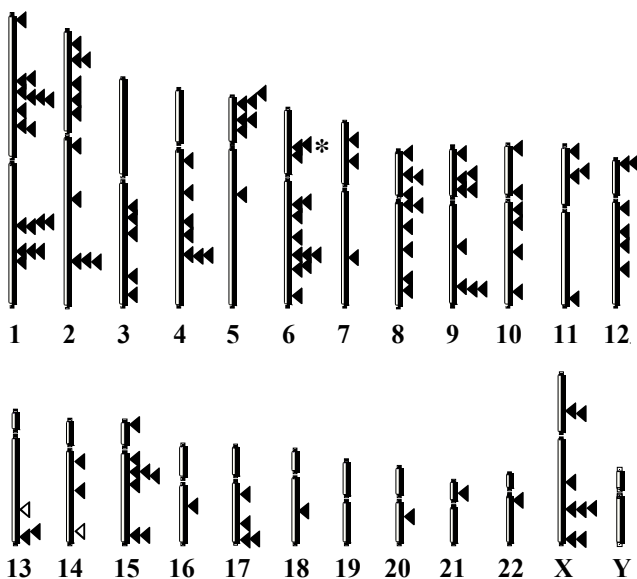


Figure 16. Mapping of *Sleeping Beauty* insertion sites on human chromosomes. Schematic representation of human chromosomes with 138 unique SB insertions. Insertion sites are marked with triangles, whereas filled triangles represent insertions in genes. Asterisk marks the single transposition event that occurred in an exon of a gene.

transposon landed in an exon. Thus, 48 out of the 138 integrations (35%) occurred in transcribed regions. Because about one third of the human genome is estimated to be transcribed (Lander et al. 2001), this frequency suggests no preference for or against insertion into genes. Such a subgenomic distribution of SB is unlike that of P elements in *Drosophila*, which have the preference to insert into 5'-UTRs of genes, close to

the transcriptional start site (Spradling et al. 1995). The predominant targeting of introns suggests that these sequences are hit more frequently either because their base composition makes them more attractive targets for the transposon, or because they tend to be significantly longer than exons or promoters (Lander et al. 2001), therefore representing a larger target into which a transposon can integrate. Eight insertions were found in repetitive sequences: five in other transposable elements such as Alu, L1 and MER1, and three in centromeric repeats. Three elements landed closer than 1 kb to a 5' region of a gene. Taken together, these results indicate a fairly random pattern of integration of SB elements in human chromosomes.

3.1.2 *Sleeping Beauty* prefers a palindromic AT-repeat for insertion

Sleeping Beauty, like all other Tc1/*mariner* elements, integrates at TA dinucleotides, which occur approximately once every 20 basepairs, on average, in vertebrate genomes. We next investigated whether all TAs are equally good targets, or if there are other sequence determinants influencing SB's target site selection. Integrated transposons were recovered from cells, and 71 chromosomal sequences flanking the integrated transposons were used to determine the DNA sequence of a consensus target site. All sequences were aligned at the

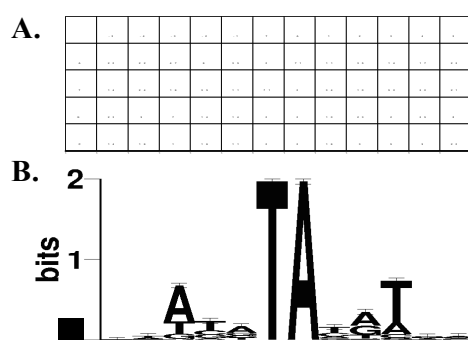


Figure 17. Consensus sequence of *Sleeping Beauty* insertion sites. Five base pairs upstream and downstream of the TA target site were analyzed for 71 insertions. (a) Base distribution of target sites; (b) Seqlogo analysis. The y-axis represents the strength of the information, with 2 bits being the maximum for a DNA sequence.

canonical TA insertion site in the same orientation, relative to the transposon. We found six bases directly surrounding the insertion site forming a short, palindromic AT-repeat: ATATTATAT, in which the central underlined TA is the insertion site (Fig. 17). Particularly conserved are the 5'- and 3'-most bases in the consensus, represented by an A and a T, respectively, in 66% and 70% of the target sites.

3.1.3 Structural properties of DNA at *Sleeping Beauty* insertion sites

Having found a particular sequence into which SB preferentially integrates, we next asked whether integration sites have anything in common on the structural level. The structural properties of DNA examined in this study included GC content, B-DNA twist, A-philicity, DNA bending and protein-induced deformability. B-DNA twist affects the tightness of the DNA coil and the ability of molecules to interact within the grooves of the DNA. These interactions allow DNA to serve as areas of binding for proteins (Gorin et al. 1995). A-philicity represents the propensity of DNA to form an A-DNA like double helix (Ivanov and Minchenkova 1995). A-DNA has a wide and shallow minor groove that is believed to provide proteins easier access to form hydrogen bonds with bases within the DNA helix. Along with A-philicity, DNA bending can lead to changes in the width and depth of the major and minor grooves, affecting a protein's access to bases of the DNA (Bruckner et al. 1995). Protein-induced deformability is the ability of DNA to change shape when in contact with a protein,

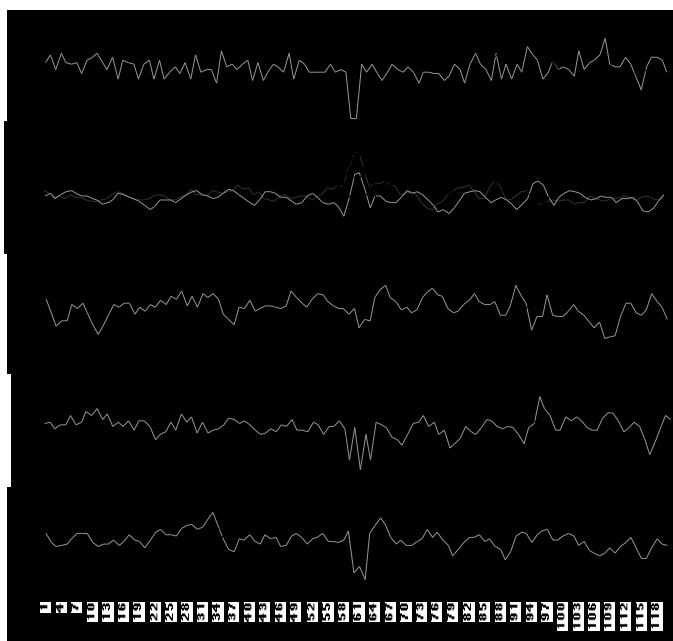


Figure 18. Physical properties of *Sleeping Beauty* insertion sites. *Sleeping Beauty* insertion sites and random human sequences were compared for five different physical properties. The random sequences from human DNA were taken from chromosome 21 and aligned at a TA. In the GC profile the base composition of the insertion and random sequences is given. Lower values in the A-philicity chart mean that the sequence is more likely to form A-DNA. Black lines represent averaged sample data, gray lines represent averaged control data. Asterisks mark base pairs which were found significantly different at a confidence level higher than 95%. Base position 60 corresponds to the beginning of the TA insertion site.

which in turn affects the binding of other proteins or the action of the protein already bound (Olson et al. 1998).

Physical properties of a data set containing 58 sequences, each with 61 bases flanking the TA insertion site on each side, were analyzed to determine significant features of SB target sites other than the actual DNA sequence.

Random DNA sequences from human chromosome 21 were analyzed for comparison with SB insertion sites. All random human sequences were also

aligned at a TA dinucleotide, allowing us to evaluate physical properties of the DNA that were not due to the canonical TA target dinucleotide. At the 90% confidence level, all five physical properties deviated from the control data around the area of SB insertion (data not shown). As shown in Fig. 18, bendability deviated from the control data at the 95% confidence level, as evidenced by the number of positions that are significantly different from the control dataset. These significant positions appeared clustered in the immediate vicinity of the insertion sites (Fig. 18). A strong signal clustered around the transposon insertion sites was also observed for protein-induced deformability (Fig. 18). Areas of significance outside of the insertion site formed no discernible pattern. These data suggest that *Sleeping Beauty* insertion sites have unique physical properties.

3.1.4 Insertion sites of Tc1/*mariner* transposons have similar structural properties

We wondered if any of the physical properties of DNA identified at SB insertion sites is shared by other Tc1/*mariner* elements. Insertion sites of the Tc1, Tc3 and *Himar1* transposons were analyzed with respect to the DNA structural features described above. Like SB, these three transposons insert into TA target sites. The first dataset was generated by taking 61 base pairs surrounding 19 Tc1 element insertions in the *C. elegans* genome. The 19 Tc1 elements were chosen by comparing a prototype Tc1 element (Genbank accession number X01005.1) to the *C. elegans* genome and taking those elements that were the most similar. For comparison, 19 random DNA sequences from chromosome I of *C. elegans* were taken. 40 unique *Himar1* insertions generated by *in vitro* transposition into plasmid DNA (Lampe et al. 1998) were also analyzed, and 40 random DNA sequences from the target plasmid were used as a control. In addition, 24 unique Tc1 and 23 unique Tc3 insertions into the *C. elegans gpa-2* gene (van Luenen and Plasterk 1994) were included in the analyses. As a control, we chose random sequence from the *gpa-2* gene. The control sequences for all of these datasets were aligned at a TA dinucleotide.

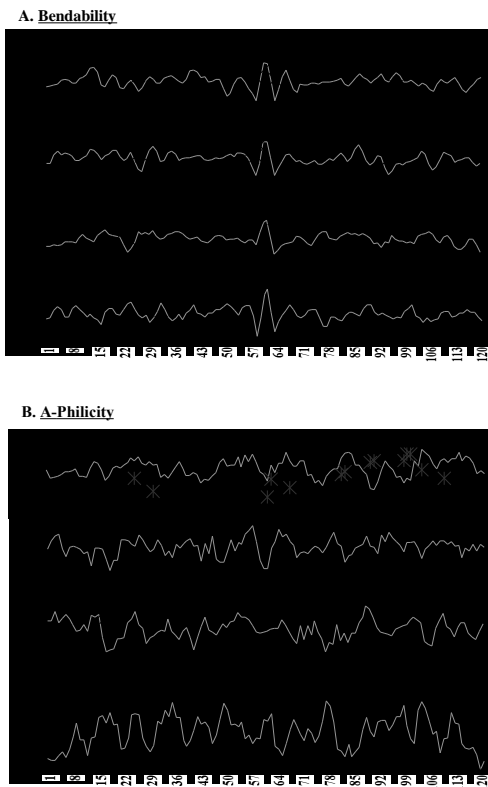


Figure 19. Physical properties of insertion sites of other Tc1/mariner elements. Shown are profiles generated for bendability (a) and A-philicity (b). Elements closely related to SB were compared to random sequences for four different physical properties. For Tc1 genomic insertions, random sequence is taken from chromosome I of *C. elegans*. For the Tc1 and Tc3 insertions in the *gpa-2* gene, random sequences were taken from the *gpa-2* gene. *Himar1* insertions are in the pBS plasmid, and random sequence is taken from the pBS plasmid. All random sequences were aligned at a TA. Charts are organized as in Fig. 18.

A uniform pattern was seen in the physical properties of both bendability and A-philicity among all elements at the 90% confidence level (data not shown). At 95% confidence, bendability appears to be a significant and conserved feature of all transposon insertion sites (Fig. 19a). For the Tc1 *gpa-2*, Tc3 *gpa-2* and Tc1 genomic datasets, clustering of significant positions around the transposon insertion sites was observed, whereas for the *Himar1* insertions, groups of significant positions were detected about 40 bps upstream and downstream of the sites of insertion (Fig. 19a).

Although the profiles for A-philicity were not consistent among the cases tested, clustering of some of the significant positions around the insertion sites was noted for the Tc1 genomic

dataset (Fig. 19b). Taken together, these data suggest that insertion sites of Tc1/mariner transposons have a property beyond just the TA sequence that gives them a more bendable structure and a looser helical conformation.

To directly test the prediction of increased bendability of the insertion target sequences, a DNase I digestion assay (Brukner et al. 1995) was performed on the consensus integration sequences of SB and Tc1. The eight-base SB (ATATATAT) and ten-base Tc1 (CACATATGTG) (van Luenen and Plasterk 1994) consensus sequences were compared to two control sequences predicted to have low bendability (AAAAAAAA and AAATAAAA) (Brukner et al. 1995). The second “bad bender” sequence contains a central TA dinucleotide to reflect that of the SB and Tc1 consensus sequences. These four sequences were flanked by

identical sequences. The digestion parameters were such that, on average, DNase I cleaved

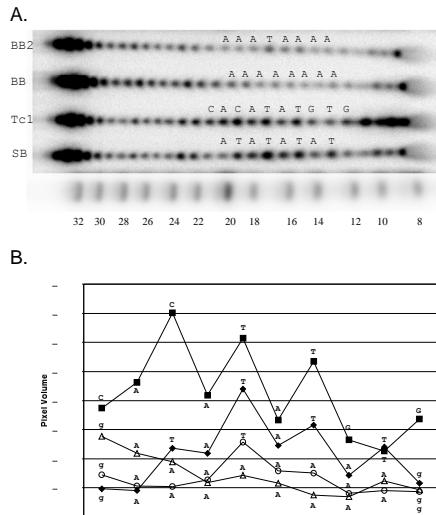


Figure 20. DNase I digestion assay of bendability. (a) Electrophoretic patterns of DNase I digestion of 32-bp oligos containing either the 8-bp SB or 10-bp Tc1 consensus insertion sequences, or two different 8-bp sequences with low predicted bendability. The bands corresponding to the consensus sequence are labeled with the respective nucleotide, and the sizes of the Oligo Size Marker ladder bands are marked. (b) The bands were quantitated and graphed relative to each other. The uppercase letters indicate the core target sequence, while the lowercase letters indicate the identical flanking sequence. (◆) SB consensus, (■) Tc1 consensus, (△) Bad Bender, (○) Bad Bender2.

each DNA molecule less than once, and therefore cleavage occurred at the most favorable position, one that is the most bendable (Bruckner et al. 1995). Thus, the more bendable a particular sequence is, the more often DNase I will digest there, and the more intense the resulting radioactive band will be. Quantitation of the digestion patterns showed that both the SB and Tc1 oligos were digested more often within their consensus target sequences than were the control oligos (Figs. 20a and b). These data confirm the computer predictions of increased bendability of DNA sequences at transposon insertion sites.

3.1.5 Patterns of hydrogen bonding at target sites

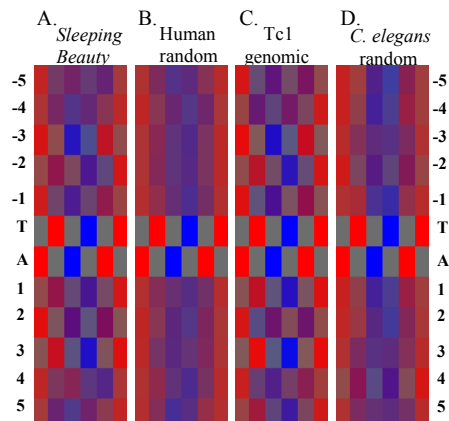


Figure 21. Hydrogen-bond analysis of insertion sites. Columns indicate six potential sites with which proteins can form hydrogen bonds. The rows indicate base pairs 3' (negative) and 5' (positive) of the TA insertion site. The colors of a given cell denote the type of hydrogen bond that can be formed: donor is red, acceptor is blue and sites that cannot hydrogen bond are gray. The analysis was performed with multiple insertions sites, and so the final color is determined by the percentages of hydrogen-bond donors, acceptors and non-hydrogen-bonding sites at a given position. (a) SB insertions in human DNA; (b) random human DNA; (c) Tc1 insertions in *C. elegans* DNA; (d) random *C. elegans* DNA.

Transposase, like other DNA-binding proteins, likely forms hydrogen bonds with its DNA substrate. Previously, analysis of a large number of P element insertion sites in *Drosophila* identified a 14-bp palindromic pattern of hydrogen bonding sites using a computer program called HbondView (Liao et al. 2000). This graphical tool identifies patterns of bond donors or acceptors in the major groove of DNA sequences by converting a set of aligned DNA sequences into a display of potential hydrogen-bonding positions. We compared insertion sites of

Tc1/*mariner* elements and random DNA for their respective propensities to form hydrogen bonds in their major grooves. Although there was no discernible pattern of hydrogen-bonding found for the Tc1 and Tc3 insertions in the *gpa-2* gene and for the *Himar1* insertions in pBlueScript (data not shown), both *Sleeping Beauty* and Tc1 genomic insertions showed a symmetrical pattern (Figs. 21a and c). The transposon insertion sites have a 10-bp palindromic pattern, including the TA target plus four base pairs on each side. Because the non-insertion control data sets for both human and *C. elegans* genomic DNA lack such a pattern (Figs. 21b and d), these results indicate that in addition to structural features, a specific pattern of hydrogen-bonding sites at the target DNA contributes to target site selection of transposons.

3.2 Targeted *Sleeping Beauty* transposition in human cells

3.2.1 Transpositional activities of transposase fusions

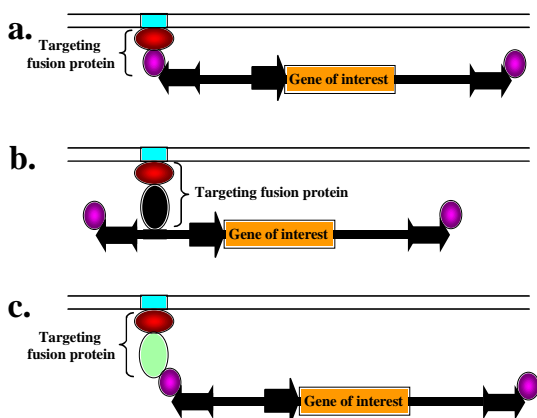


Figure 22. Experimental strategies for targeting *Sleeping Beauty* transposition. The common components of the targeting systems include a transposable element that contains the IRs (arrowheads) and a gene of interest equipped with a suitable promoter. The transposase (purple circle) binds to the IRs and catalyzes transposition. A DNA-binding protein domain (red oval) recognizes a specific sequence (turquoise box) in the target DNA (parallel lines). **(a) Targeting with transposase fusion proteins.** Targeting is achieved by fusing a specific DNA-binding protein domain to the transposase. **(b) Targeting with fusion proteins that bind the transposon DNA.** Targeting is achieved by fusing a specific DNA-binding protein domain to another protein (white oval) that binds to a specific DNA sequence within the transposable element (yellow box). In this strategy, the transposase is not modified. **(c) Targeting with fusion proteins that interact with the transposase.** Targeting is achieved by fusing a specific DNA-binding protein domain to another protein (light green oval) that interacts with the transposase. In this strategy, neither the transposase nor the transposon is modified.

Because the transposon system consists of two main functional components: the transposon DNA and the transposase protein, tethering the transpositional complex to a given site in the genome can be brought about by interactions with either of these two components. Therefore, to achieve targeted transposition of SB, we considered experimental strategies employing a DNA-binding protein domain responsible for binding to a chromosomal target and fused to either 1) the transposase (Fig. 22a); 2) or a protein domain that binds to a site within the

transposable element (Fig. 22b); 3) or a protein domain that makes contacts with the transposase through protein-protein interactions (Fig. 22c).

Fusion proteins containing the SB transposase and either the bacterial tetracycline repressor (TetR) that specifically binds the tetracycline operator sequence, or the Jazz and

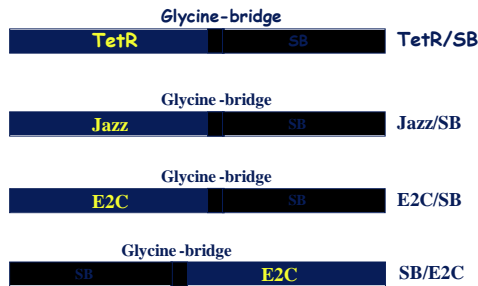


Figure 23. Design and transpositional activities of transposase fusions. (a) Schematic representation of the fusion proteins that consist of the SB transposase fused to the tetracycline repressor (TetR), the Jazz or the E2C ZF proteins. The TetR/SB fusion contains the SV40 nuclear localization signal (NLS) in addition to the NLS naturally present in SB. All fusions contain a glycine-bridge consisting of ten consecutive glycine residues to provide a flexible linker between the fusion partners.

E2C ZF peptides were engineered (Fig. 23).

Transposon excision activity of the fusion proteins was tested using a PCR-based excision assay (Izsvák et al. 2004). Out of four constructs tested, only the

Jazz/SB fusion showed detectable activity in human HeLa cells, although at a clearly reduced efficiency compared to unfused transposase (Fig. 24, compare lanes 2 and 4). In line with the excision data, the Jazz/

SB fusion was found to retain transpositional activity at about 10-15% of the wild-type level. However, a PCR survey on genomic DNA isolated from transformant cells generated using Jazz/SB as transposase source revealed no indication of targeted transposition into the

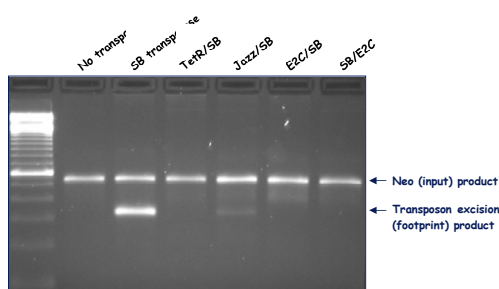


Figure 24. A fusion protein consisting of the SB transposase and the Jazz zinc finger protein retains transposon excision activity. HeLa cells were cotransfected with a *neo*-marked transposon plasmid and vectors expressing the proteins indicated. Transposon excision is assayed with PCR that amplifies a footprint product. PCR-amplification of the *neo* marker inside the transfected transposon donor serves as a loading control.

utrophin locus, and no occurrence of the 9-bp binding site of Jazz within a 1-kb window around the transposon insertion sites (data not shown).

Taken together, the results establish that most direct fusions to the SB transposase have negative effects on transpositional activity, and suggest that ZFs with higher specificity in terms of DNA binding will be required for targeted transposition.

3.2.2. Targeting using fusion proteins that interact with the transposon DNA

3.2.2.1 Targeted *Sleeping Beauty* transposition into chromosomal matrix attachment regions in cultured human cells

Based on the negative effects of direct transposase modifications on transposition activity, we sought to test a targeting strategy that employs modification of the transposon DNA (Fig. 22b). We began our experiments with targeting into chromosomal regions that are represented many times in a complex genome, whose base composition is AT-rich so that SB can

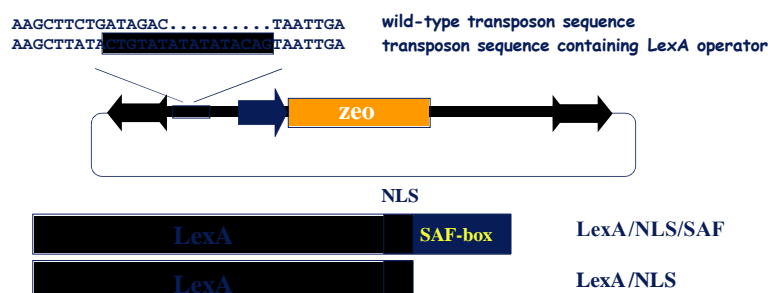


Figure 25. Components of a targeting system for transposon insertion into scaffold attachment regions (SARs). The targeting fusion protein consists of the bacterial LexA DNA-binding protein, a nuclear localization signal (NLS) and the SAF-box. A control protein lacks the SAF-box. The transposable element contains a 16-bp binding site for the LexA protein between the left inverted repeat of the transposon and the SV40 promoter that drives expression of the zeocin resistance gene.

efficiently integrate into them,
and for which interacting
protein determinants are known
and well characterized.

Scaffold attachment regions (SARs), also called matrix attachment regions or MARs

that partition the genome into distinct, independent loops by binding to the nuclear matrix, satisfy the above criteria. Scaffold attachment factor A (SAF-A) is known to bind MAR DNA, and contains an evolutionarily conserved protein domain, termed the SAF-box, which is necessary and sufficient to mediate MAR binding (Kipp et al. 2000). A fusion protein in which the SAF-box is fused to the C-terminus of the *E. coli* LexA protein was engineered

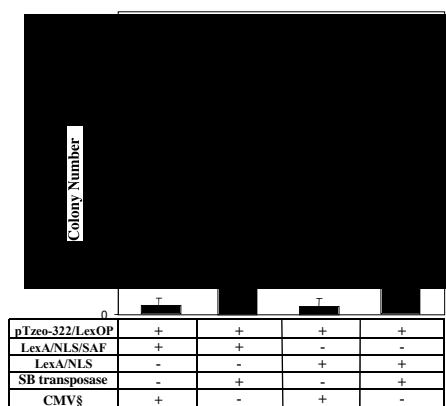


Figure 26.SB transposition in human HeLa cells, in the presence or absence of the targeting fusion protein.

(Fig. 25). The function of the LexA protein in this experimental approach is to specify binding of the targeting fusion protein to a LexA operator site engineered into a SB transposon vector (pTzeo-322/LexOP in Fig. 25). First, the effect of expression of the targeting protein on SB transposition was examined by comparing transpositional efficiencies in human

HeLa cells transfected with a plasmid expressing the SB transposase (pCMV/SB), pTzeo-322/LexOP and a plasmid either expressing the LexA/NLS/SAF fusion or LexA/NLS only. No difference in transpositional activity was observed (Fig. 26), indicating that binding of the fusion protein to the transposon does not interfere with transposition.

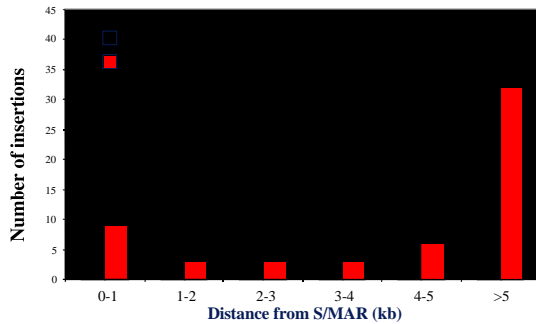


Figure 27. Preferential insertion into S/MARs by transposon targeting. Transposition events were recovered from transformed cells, and human chromosomal DNA flanking the insertion sites was analyzed with respect to proximity to chromosomal S/MARs. The MAR-Wiz program was used to predict the presence of an S/MAR in the vicinity of a transposon insertion. Distances were categorized, and the numbers of insertions obtained in the presence and in the absence of the targeting fusion protein in each category are shown.

We hypothesized that targeted transposition could occur by simultaneous interaction of LexA/NLS/SAF with both the transposon vector and chromosomal MAR regions, thereby forcing the transposition complex to integrate into nearby sites.

In the absence of suitable methods for directly selecting transposition events in MARs, transfected HeLa cells were pooled, and SB

integration sites were cloned out from genomic DNA and analyzed using a software called MAR-Wiz that detects the presence of MARs within DNA sequences. We have analyzed 56 unique SB insertions recovered from cells expressing LexA/NLS/SAF and 57 from cells expressing LexA/NLS only. The proximity of each insertion site to a predicted MAR was categorized into six groups. A statistically significant ($p=0.024$) difference between the two data sets was found, with the most notable difference being in the group of insertions that are the closest to a MAR. Specifically, nine insertions in the targeting group occurred within 1-kb

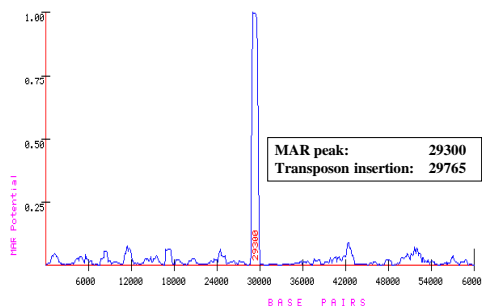


Figure 28. MAR-Wiz output showing the position of a predicted S/MAR. A transposon insertion close to this S/MAR is shown.

distance from a predicted MAR versus two events in the control group (Fig. 27). An output of the MAR-Wiz program showing the positions of the MAR peak and that of a transposon insertion is shown in Fig. 28. Some of the cloned chromosomal DNA fragments flanking the

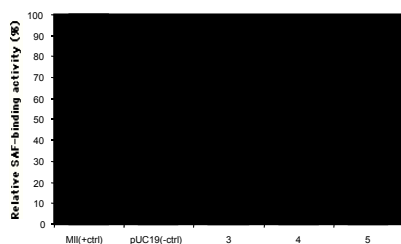


Figure 29. *In vitro* binding of SB target sites by the SAF-box. Resin-bound SAF-box peptide was incubated with radiolabeled MII DNA (positive control); pUC19 plasmid DNA (negative control), and 3-5) transposon target sites, and radioactivity retained on the resin after extensive washing was measured. Efficiency of binding of the SAF-box to MII DNA is arbitrarily set to 100%.

transposon insertions and predicted to contain MAR sequences were tested in an *in vitro* assay for binding by the SAF-A protein. As shown in Fig. 29, two out of three SB target sites tested showed considerable binding to the SAF-box peptide, suggesting that SB insertions in these regions were indeed targeted by the DNA-binding activity of the SAF box. Taken together, our results indicate a shift

in the insertional spectrum of SB in the presence of the targeting fusion protein, and show a bias for integration in the vicinity of genomic MAR sequences.

3.2.2.2 Targeted transposition into a specific locus in human cells

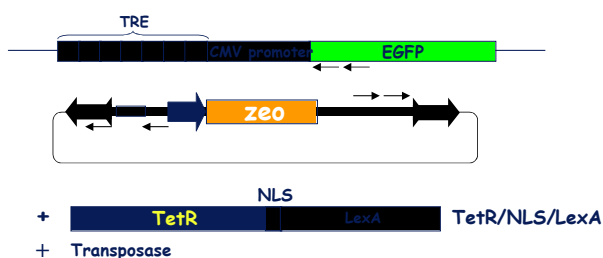


Figure 30. Components of a targeting system for transposon insertion into the tetracycline response element (TRE). The target DNA is a chromosomally integrated tetracycline response element (TRE) upstream of the CMV minimal promoter and the EGFP gene, in transgenic human HeLa cells. Arrows indicate the approximate positions of nested PCR primers that were used to identify targeted transposition events. The targeting fusion protein consists of the tetracycline repressor (TetR) that binds to the TRE, a nuclear localization signal (NLS), and the LexA DNA-binding protein that binds to its 16-bp binding site (yellow box) engineered into the transposon vector.

Encouraged by the above results, we next sought evidence for targeted transposition into a unique site in the human genome. For this purpose, we concentrated our efforts to the tetracycline repressor (TetR)-operator system (Gossen et al. 1995). Unlike binding of the SAF-box to scaffold attachment regions, interaction of TetR with its operator site is highly sequence-specific. We have generated a transgenic, HeLa-derived cell line containing a tetracycline response element (TRE, encompassing seven units of the tetracycline operator)-driven EGFP gene, which served as a chromosomal target for transposon integration (Fig. 30). The targeting fusion protein consisted of TetR, an NLS and the LexA protein (TetR/NLS/LexA in Fig. 30). The ability of TetR/NLS/LexA to bind to the TRE and the LexA operator-containing transposon DNA was tested in an electrophoretic mobility shift

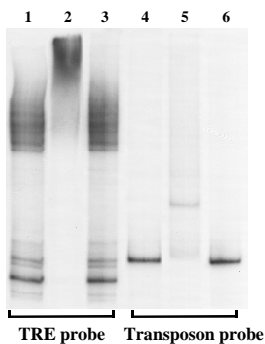


Figure 31. Mobility shift experiment showing the ability of the targeting fusion protein to specifically bind to radiolabeled TRE and transposon probes. Lanes 1 and 4: no protein; lanes 2 and 5: HeLa extract containing TetR/NLS/LexA; lanes 3 and 6: untransfected HeLa extract.

TRE and transposon probes (Fig. 31, lanes 2 and 5), but not with a prepared from untransfected cells (Fig. 31, lanes 3 and 6).

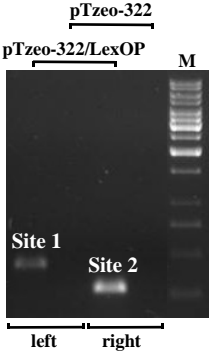


Figure 32. Site-specific transposon insertions in human cells. The agarose gel shows PCR products obtained from cells transfected with pT/zeo-322/LexOP or with pT/zeo-322 with primers amplifying the left or the right IR of the transposon. M: size marker.

cells transfected with radiactively labeled probes (Fig. 31). transposon probes TetR/NLS/LexA 5), but not with a prepared from (Fig. 31, lanes 3 and

The TRE-EGFP transgenic cell line was cotransfected with a transposase expression plasmid, pTzeo-322/LexOP and

TetR/NLS/LexA. After selection, approximately 400 cell colonies were pooled, genomic DNA prepared and subjected to PCR analysis with primers designed to amplify transposition events upstream of the EGFP gene (Fig. 30). PCR products consistent with transposon integration in both orientations with respect to the EGFP gene were obtained from

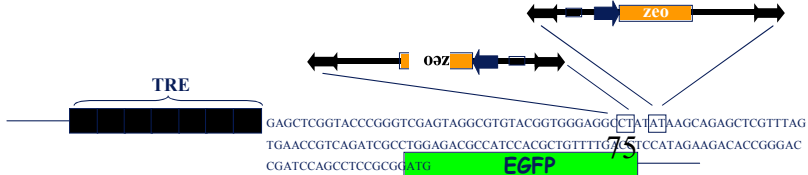


Figure 33. Targeted transposition close to the TRE. PCR products in Fig. 32 represent two transposon insertions in close proximity of the targeted TRE region, in the two possible orientations, within two TA dinucleotides of the CMV promoter TATA-box (Sites 1 and 2).

transfections with pTzeo-

322/LexOP but not with pTzeo-322, a transposon vector without a LexA binding site (Fig. 32), indicating that a functional interaction between the LexA protein and its binding site within the transposon DNA is required for targeted insertion into the TRE locus. Sequencing of the PCR products revealed transposition of SB in the two different possible orientations into two TA sites in the promoter region of the EGFP gene, 44 and 48 bps downstream of the TRE region (Sites 1 and 2 in Fig. 33). In conclusion, our strategy based on targeting proteins that can bind both transposon and target DNA was successful in directing SB element transposition into the vicinity of a specific DNA sequence in the human genome.

3.2.3 Targeting using fusion proteins that interact with the transposase

Many naturally occurring transposable elements utilize protein-protein interactions for targeted insertion into defined sites. A requirement for adapting such a strategy to SB is to identify proteins that interact with the transposase. We previously identified an N-terminal

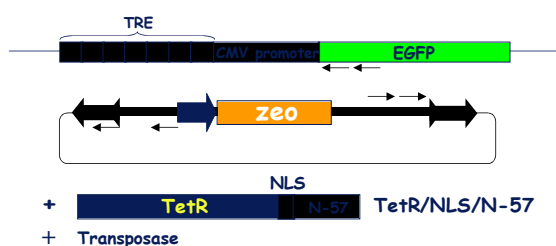


Figure 34. Transposon targeting using a strategy based on protein-protein interactions between a targeting fusion protein and the SB transposase. The targeting fusion protein consists of the tetracycline repressor (TetR) that binds to the TRE, a nuclear localization signal (NLS), a glycine-bridge and the N-terminal protein interaction domain of the SB transposase (N-57). The targeted chromosomal locus as described in Fig. 30, the transposon is an unmodified, antibiotic-marked SB element.

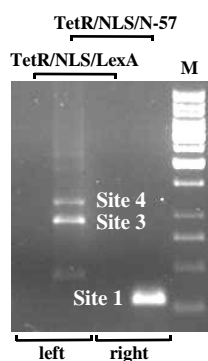


Figure 35. Locus-specific SB transposition in human cells. The agarose gel shows PCR products obtained from cells transfected with a vector expressing TetR/NLS/N-57 or with TetR/NLS/LexA with primers amplifying the left or the right IR of the transposon. M: size marker.

fragment of the SB transposase to encode a transposase interaction function (Izsvák et al. 2002). The region required for transposase interaction has been mapped to the N-terminal 57 amino acids of the transposase containing the PAI subdomain of the paired-like (PAI + RED) DNA-binding domain of the SB

transposase (Izsvák et al. 2002). We have shown before that coexpression of a peptide covering the PAI subdomain (hereafter called N-57) with the full-length transposase does not impair transposition (Izsvák et al. 2002). Therefore, we chose to pursue a strategy to generate fusion proteins of DNA-binding proteins with N-57, in the hope that these fusions will retain their abilities to bind to

desired target DNA sequences as well as to interact with the SB transposase without compromising its function. An experimental strategy similar to that described above was

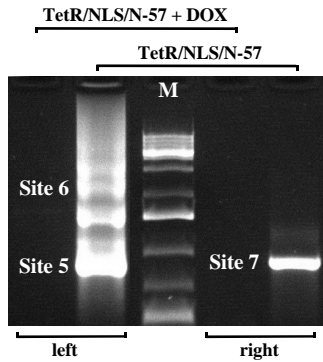


Figure 36. Targeted transposition events require a functional interaction between TetR and TRE. The agarose gel shows PCR products obtained from cells transfected with a vector expressing TetR/NLS/N-57 and selected in the absence and presence of doxycycline (DOX), with primers amplifying the left or the right IR of the transposon. M: size marker.

employed, except that the targeting fusion protein consisted of TetR and N-57 (Fig. 34). The TRE-EGFP transgenic cell line was cotransfected with a transposase expression plasmid, pTzeo-322 and the targeting fusion protein. Approximately 400 antibiotic-resistant cell colonies were pooled, and targeted transposition events assessed by PCR on genomic DNA, as above.

PCR products consistent with transposon insertions in both orientations with respect to the EGFP gene near the targeted

TRE region were recovered from transfections with TetR/NLS/N-57 but not with TetR/NLS/LexA (Fig. 35). Thus, N-57 is required for transposon insertion into the targeted TRE locus. Two of the insertions correspond to transposition events into Site 1 in different orientations, whereas two other insertions correspond to transposition events into TA target dinucleotides

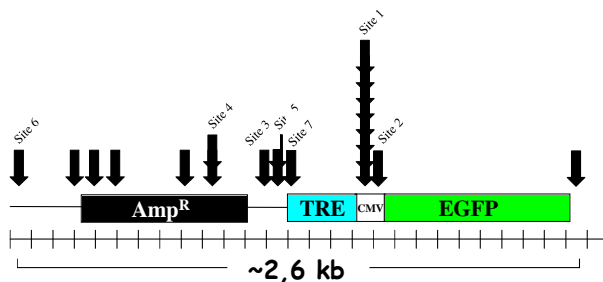


Figure 37. Mapping of targeted SB insertions. Mapping with respect to the TRE-EGFP target isolated from six independent experiments is shown. Multiple arrows represent independent insertions into the same site.

170 bp (Site 3) and 423 bp (Site 4) upstream of the TRE region (Figs. 36 and 37). As a further control, cells transfected with TetR/NLS/N-57 were selected in the absence and presence of doxycycline (DOX), an antibiotic that disrupts interaction of TetR with TRE. As shown

in Fig. 36, transposon insertions into the TRE region can only be recovered in the absence of DOX. Collectively, the results demonstrate that the TRE target itself does not serve as a transpositional hotspot for SB, because both N-57 and a functional interaction of TetR with TRE (i. e., components of an active targeting mechanism) are required for transposition into

this region. Altogether 12 different insertion sites within the TRE-EGFP target were recovered from six independent transfections, from which three, Site 1, Site 4 and Site 5, were hit multiple times independently (Fig. 37).

The efficiency at which transgenic cells harboring targeted insertion events can be generated was assessed by doing the PCR test on DNA isolated from individual transgenic cell clones. Five out of 24 clones contained targeted transposition events into Site 1 (Fig. 38, upper gel). In an independent experiment, four out of 48 individual transgenic cell clones were found to have targeted transposition events (Fig. 38, lower gel). Thus, on average, the frequency at which transgenic cells with targeted transposon insertions can be generated is >10%. We conclude that transposon targeting based on protein-protein interactions between

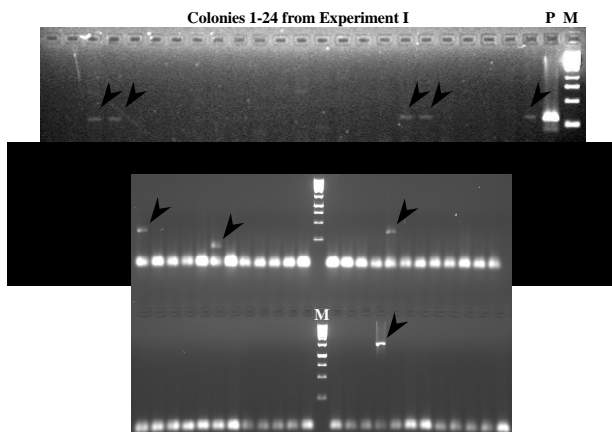


Figure 38. Frequency of targeted transposition. The agarose gels show PCR products obtained from individual, transgenic cell clones picked from two independent cell transfections. A PCR product recovered from pooled (P) DNA samples serves as a reference. M: size marker.

the SB transposase and a targeting fusion containing the PAI subdomain is a successful strategy to direct SB integrations into a given locus in the human genome. We also conclude that SB insertions can occur into multiple sites within a targeted chromosomal region, and that particular sites are favored targets (hotspots).

3.3 New scientific results

- We established that the distribution of experimentally induced *Sleeping Beauty* insertions in the human genome can be considered fairly random, because most chromosomes can serve as a target; no obvious hotspots with multiple insertions were found, and no preference for coding versus non-coding DNA was observed.
- We showed that the SB element displays a certain degree of specificity in target site utilization at the DNA sequence and structural level. A palindromic AT-repeat consensus sequence with bendability and a symmetrical pattern of hydrogen bonding sites in the major groove of the target DNA define preferred sites for integration.
- We demonstrated targeted chromosomal insertion of the SB transposon in human cells. This provides proof-of-principle for directing chromosomal insertion of an otherwise randomly integrating genetic element into preselected sites.
- We established a successful strategy based on targeting proteins that can bind both transposon and target DNA to direct SB element transposition into the vicinity of a specific DNA sequence in the human genome.
- Transposon targeting based on protein-protein interactions between the SB transposase and a targeting fusion containing the N-terminal protein interaction domain of SB is a successful strategy to direct SB integrations into a given locus in the human genome. This approach was found to enable a $\sim 10^7$ -fold enrichment of transgene insertion at a desired locus.

4 DISCUSSION

4.1 Common physical properties of DNA affecting target site selection of *Sleeping Beauty* and other Tc1/*mariner* transposable elements

We have analyzed a collection of *Sleeping Beauty* integration sites both on the genomic level, by mapping the insertions on chromosomes, and by inspection of sequence and structural properties of DNA locally, at the sites where transposon integrations occurred. The distribution of experimentally induced SB insertions in the human genome can be considered fairly random (Fig. 16), because most chromosomes can serve as a target; no obvious hotspots with multiple insertions were found, and no preference for coding versus non-coding DNA was observed. In another study, analysis of a significant dataset of >1300 SB insertions recovered from primary and cultured mammalian cells (mouse and human) established that SB transposition shows 1) a general preference for AT-rich DNA, 2) a strong bias toward microsatellite repeats, and 3) a small but significant bias toward genes and their upstream regulatory sequences (Yant et al. 2005). However, compared to virus-based integrating vector systems, including retrovirus-, HIV- or AAV-based vectors, that were found to have a propensity for integrating into genes versus non-genic regions (Scherdin et al. 1990; Schroder et al. 2002; Nakai et al. 2003; Thomas et al. 2003; Wu et al. 2003), the regional preferences associated with SB-mediated integration were much less pronounced (35% of SB insertions in RefSeq genes, versus 53% for ASV, 51% for MLV, 83% for HIV-1 (Yant et al. 2005) and 72% for AAV (Nakai et al. 2003). Importantly, in contrast to most integrating virus-based, microarray analyses revealed no correlation between the integration profile of SB and transcriptional status of targeted genes (Yant et al. 2005), suggesting that SB might be a safer vector for therapeutic gene delivery than most viruses that are currently used. Indeed, it is important to note that no dominant adverse effects associated with SB vector integration have been so far found in experimental animals (Essner et al. 2005). The safety profile of SB transposon-based vectors is further improved by recent findings that they are fairly inert in

their transcriptional activities, and that insulator elements can successfully be incorporated in the next generation of transposon vectors (Walisko et al. 2008). Nevertheless, the genotoxic potential of SB-based vectors will have to be systematically assessed in the future, probably by applying high throughput, cell-based assays.

Two observations might suggest some selectivity in chromosomal integration of the SB transposon on the genome level. First, although about 45% of the human genome is derived from transposable elements (Lander et al. 2001), only six percent of the SB insertions occurred into repetitive DNA. Second, no insertion was recovered from chromosome 19, or the Y chromosome, which is lacking in human female-derived HeLa cells. Intriguingly, chromosome 19 has the lowest A/T content among all the chromosomes, therefore our finding might indicate a general affinity of SB for A/T-rich DNA. There seems to be a significant difference in the distribution of SB jumps depending on whether transposition is initiated from an extrachromosomal plasmid or from a chromosome. SB displays local hopping when transposition is initiated from a chromosome; about fifty percent of the reintegration events occur on the same chromosome (Luo et al. 1998; Fischer et al. 2001; Horie et al. 2001). Because we found that almost all chromosomes can serve as a target, the phenomenon of local hopping cannot be due to an effect specific for certain chromosomes; rather, it likely is an attribute of the transposition reaction itself. Compared to P elements (Spradling et al. 1995) and retroviruses (Scherdin et al. 1990), which preferentially insert into the 5'-regions of genes, the genomic distribution of *Sleeping Beauty* insertions can be advantageous for insertional mutagenesis, because relatively random integration is expected to generate a wider mutational spectrum.

We found three properties of DNA that together define preferred sites for integration of Tc1/*mariner* transposons (Figs. 18-21). These are bendability, A-philicity and a symmetrical pattern of hydrogen bonding sites in the major groove of the target DNA. Bendability of the target site seems to be an important factor, because phosphodiester bonds

need to be accessed by the incoming transposase/transposon complex. A bendable structure may allow transposase and/or auxiliary host factors, to deform the bound DNA into a spatial optimum for strand transfer. The consensus target site of SB is made up of AT dinucleotides (Fig. 17), which have been shown to be associated with bendable DNA structures (Brukner et al. 1995; Olson et al. 1998). Previous studies found DNA bending important for retroelement integration (Muller and Varmus 1994; Jurka et al. 1998); thus, inserting into bent DNA appears to be a shared mechanism between the two groups. A statistically significant difference between transposon insertion sites and random DNA in terms of A-philicity was only observed at the 90% confidence level in our analyses. Nevertheless, we believe that a role of A-philicity in insertion site preference is likely because of a direct interaction of the transposase with DNA. Because A-DNA has a wide and shallow minor groove, it is believed to provide proteins easier access to the DNA helix. Indeed, A-philicity of DNA was found to be an important factor in determining target site preference of the Mu bacteriophage (Haapa-Paananen et al. 2002).

We have detected a 10-bp palindromic pattern of hydrogen bonding for both SB and Tc1 genomic insertions (Fig. 21). Such palindromic pattern and the symmetry of the consensus target site sequence (Fig. 17) together indicate that the target DNA is recognized by a dimeric or multimeric form of the transposase. Indeed, we have shown that SB transposase forms tetramers in solution, suggesting the involvement of a transposase tetramer in SB transposition (Izsvák et al. 2002). A possible reason for the lack of a hydrogen bonding pattern for the other data sets might be that those studies were not done on a genomic scale. Structural features of the target DNA may be more significant for recognition by the transposition complex when DNA is bound in chromatin. In the case of SB and Tc1 genomic insertions, the hydrogen bonding patterns indicate an interaction between the transposase and the major groove of the target DNA. There are remarkable parallels between the structural properties of transposon insertion sites identified in this study and the manner in which the

EcoRV restriction endonuclease binds its recognition sequence GATATC (Winkler et al. 1993). Upon binding, *EcoRV* induces a major conformational change of the DNA, which deviates considerably from canonical B-form DNA. The DNA is bent by 50° at the central TA step, which leads to a widening of the minor groove. Base-specific hydrogen bonds between the enzyme and the recognition base pairs occur exclusively in the major groove (Winkler et al. 1993). These similarities underscore the evolutionary conservation of some of the fundamental mechanisms in DNA-protein interactions.

We have shown here that target site selection of transposable elements is considerably more specific than it was assumed before, and that it is primarily determined on the DNA structural rather than on the sequence level. Our results indicate that a combination of particular physical properties (Figs. 18-20) generate a spatial optimum of the DNA for transposase interaction. Such a spatial optimum, together with a specific hydrogen-bonding capacity (Fig. 21) recruits the transposase with a substantial degree of specificity. The significance of our findings is supported by the observation that this pattern of structural preference is conserved in the Tc1/*mariner* family and in other, relatively randomly integrating transposons in the DDE recombinase family such as the bacterial elements *Tn5* (data not shown), *Tn7* (Kuduvalli et al. 2001), *Tn10* (Pribil and Haniford 2000), Mu bacteriophage (Haapa-Paananen et al. 2002), *IS231* (Hallet et al. 1994) and retroviral integrases (Muller and Varmus 1994; Pruss et al. 1994). Significantly, transposition by the RAG V(D)J recombinase, yet another DDE transposase (Jones and Gellert 2004), is preferentially targeted to distorted DNA structures (Lee et al. 2002).

However, these factors cannot be the only determinants of target site selection, because the Tc1 and Tc3 elements have different insertion profiles in *C. elegans* (van Luenen and Plasterk 1994). Therefore, it appears that there exist at least two levels of selection that together determine how favorable a particular DNA sequence is for transposon insertion. Physical properties of the DNA primarily specify a set of sequences in a genome that are in a

spatial optimum to receive a transposon insertion, whereas the ability of the transposase polypeptide to efficiently interact with such sequences specify a subset within these sites where insertions occur.

4.2 Steps towards target-selected *Sleeping Beauty* insertion

We demonstrated targeted chromosomal insertion of the *Sleeping Beauty* transposable element in human cells. The very fact that targeted transposition can be achieved is remarkable, because the SB element integrates in a fairly random manner into human chromosomal DNA.

We have evaluated three distinct molecular strategies for targeted SB transposition, making use of heterologous DNA-binding domains that are either fused to: 1) the SB transposase itself, 2) a protein domain that binds the transposon DNA, or 3) a protein domain that interacts with the SB transposase (Fig. 22). Fusions of the bacterial *IS30* transposase with the λ repressor and with the DNA-binding domain of the transcription factor Gli1 showed altered insertions profiles in *E. coli* and in zebrafish embryos, respectively, using plasmid targets (Szabo et al. 2003). Furthermore, direct fusions of the *Mos1* and *piggyBac* eukaryotic transposases with the GAL4 DNA-binding domain were recently shown to retain transpositional activity, and to result in site-selective transposon insertion in a plasmid-to-plasmid experimental setup in mosquito embryos (Maragathavally et al. 2006). We have assessed the feasibility of fusing DNA-binding proteins directly to the SB transposase polypeptide (Fig. 23). Only one out of four fusions, the artificial 3-finger protein Jazz added to the N-terminus of SB, showed detectable transposition activity (Fig. 24). We obtained no evidence for targeted transposition by Jazz/SB either into the utrophin locus or any other region containing the 9-bp binding site of Jazz. Conceivably, specificity for a 9-bp DNA sequence of Jazz is far too low for targeted transposition. An independent study examined the transpositional activities of three different transposase proteins after fusion to Gal4 in cultured

human cells (Wu et al. 2006). Fusions completely abolished transpositional activity of *Tol2* and SB11 (an early-generation hyperactive mutant of SB), whereas only a slight decrease in activity was observed for Gal4-PB when compared to unfused PB transposase. Targeted transposition by the fusion transposases was not investigated in this study. Yet another group reported that only N-terminal fusions to the SB transposase retained transpositional activity, and that fusion of the Gal4 DBD to HSB5 (a third-generation improved SB transposase) resulted in a drop in transposition efficiency to ~26% of unfused HSB5 (Yant et al. 2007). This fusion transposase showed targeted transposon integration in a plasmid-based assay in cultured human cells. Targeted transposition events were enriched about 11-fold in a 443-bp window around a 5-mer UAS site in the target plasmid, as compared with integration patterns mediated by unfused transposase. A recent publication reported the generation of fusion proteins of E2C and the HSB5 hyperactive SB transposase (Yant et al. 2007). As seen before (Wilson et al. 2005), fusion proteins showed reduced transpositional activity as compared to unfused transposase, but about 20% transposition activity could be rescued by applying a glycine/serine linker between the ZF and transposase domains and by using a human codon-optimized E2C gene. This optimized fusion protein showed targeted transposon integration in a plasmid-based assay in cultured human cells. Targeted transposition events were enriched about 8-fold in a 443-bp window around a 5-mer repeat of the E2C binding site in the target plasmid, as compared with integration patterns mediated by unfused HSB5. However, cell-based assays failed to detect targeting of the E2C binding site in a genomic context. One possibility to explain failure of targeting in a genomic context could be physical constraints on the transposase upon site-specific binding in that the transposase is unable to interact with a TA dinucleotide to integrate the transposon. This may especially hold true for GC-rich DNA sequences at the *erbB-2* promoter region. Future efforts to improve this technology will thus have to focus on the identification of highly specific DNA-binding domains, and a systematic

evaluation of protein spacer sequences linking the two fusion partners in order to allow rational design of direct transposase fusions.

The second, more successful strategy was based on a fusion protein with dual DNA-binding activity that has the capacity to bind to two DNA molecules that contain binding sites of the respective fusion partners, thereby bringing them into close proximity (Fig. 22b). A similar mechanism of bridging of DNA molecules by proteins might act in targeting some P element transposon vectors in *Drosophila*. P element insertion is essentially random at the genome scale. However, P elements containing regulatory sequences from the *engrailed* gene show some insertional specificity by frequently inserting near the endogenous, parental gene (Hama et al. 1990; Kassis et al. 1992). This phenomenon, called transposon "homing", tends to be region-specific (Kassis et al. 1992) with transposon integrations distributed over several kilobase pair regions near the targeted loci.

Potential SB targeting by such mechanism was assessed by engineering a LexA operator into a benign site within an SB transposon vector (Fig. 25), and by engineering a fusion protein consisting of LexA and the SAF-box. The SAF-box is a protein domain first identified in the human scaffold attachment factor (SAF-A) that specifically binds to scaffold/matrix attachment regions (S/MARs) (Kipp et al. 2000). S/MAR elements are bound to the nuclear matrix, thereby structuring chromosomal DNA by forming chromatin loops. Transgenes flanked by S/MARs have shown expression independent from their site of integration. Therefore, a possible way to minimize silencing effects on transgene expression could be the insertion of a transgene into S/MARs. For targeted transposition into S/MARs to occur, the LexA-SAF-box fusion protein was expected to bind the LexA operator-containing transposon. This protein-DNA complex would then be tethered to S/MAR regions of chromosomes through SAF-box binding, whereas transposition into linked sites would occur upon recruitment of SB transposase. Indeed, an increase in transposon insertions within a 1-kb range of genomic S/MAR sequences was observed as compared to controls with fusion

proteins lacking the SAF-box (Figs. 27 and 28). In this study, targeting by a protein with highly specific DNA-binding properties, the tetracycline repressor (TetR), was also sought. A transgenic HeLa cell line incorporating a tetracycline response element (TRE)-driven EGFP gene as a targeted locus was created. In this experiment, a targeting fusion protein consisting of TetR and LexA was applied (Fig. 30). Integrations upstream of the EGFP gene were determined, yielding insertions into two TA sites within the EGFP promoter region 44 and 48 bp downstream of the TRE region (Fig. 33). No insertions into this region were detected with transposons lacking the LexA operator sequence, suggesting that interaction between the targeting protein and the transposon DNA is indeed required for targeted transposition events (Fig. 32). Thus, the targeted transposition events identified in these experiments were likely mediated by simultaneous binding of the targeting fusion protein to both transposon and target DNA. In sum, this strategy shows promise, because it does not measurably interfere with the transposition process, but its success may be limited by the ability of the targeting fusion protein to interact with the modified transposon in the cell.

A third strategy for targeted SB transposition, based on protein-protein interactions between a targeting protein and the SB transposase (Fig. 22c), was also evaluated. As shown for HIV IN and LEDGF/p75, protein-protein interactions can tether integration complexes to certain regions of the genome, suggesting that such a mechanism can be adapted for targeted transposon insertion as well (Table 3). We have successfully adapted such a strategy for targeted SB transposition by coexpressing the SB transposase with a targeting fusion protein consisting of a specific DNA-binding domain and a subdomain of the SB transposase that mediates protein-protein interactions between transposase subunits (Fig. 34). Importantly, coexpression of N-57 together with full-length transposase had no dominant negative effect on transposition (Izsvák et al. 2002). A significant advantage of this technology as compared to direct transposase fusions is that the transposase polypeptide does not have to be modified; thus, potential negative effects on transposase activity are eliminated. Analysis of the

insertion sites obtained in the presence of the TetR/NLS/N-57 fusion protein allows us to draw some general conclusions concerning the mechanism of targeted SB transposition. First, although a preferred integration hotspot 44 bps downstream of TRE was identified, targeted insertion events occurred within a 2.6-kb window around the targeted TRE (Fig. 37). This is in contrast to targeted retroviral and retrotransposon insertions that have been found to occur within a narrow (<150 bp) integration window around the targeted DNA sites (Bushman 1994; Goulaouic and Chow 1996; Katz et al. 1996). A possible explanation for this difference is that these retroelements are fairly promiscuous in terms of the DNA sequences into which they can integrate; thus, tethering the integration complex can result in integration into nearby sites. In contrast, SB requires TA dinucleotides for integration that have to be accessible for the transpositional complex. Our observations are compatible with a model of targeted transposition wherein the SB integration complex is drawn to a chromosomal region by protein-protein interactions, but the sites where integration can take place will be limited by the biochemical and biophysical constraints affecting SB transposition. In this respect, it is interesting to note that the TRE itself was never targeted (Fig. 37), even though the tetracycline operator sequence contains two TA sites. It is likely, that these sites are not available for integration because they are occupied by TetR/NLS/N-57. Finally, it is evident from the present study that, in addition to the targeted insertion events, transposition nevertheless also occurs into numerous non-targeted sites. Since the efficacy of targeted insertion of naturally occurring transposons can approach 100%, the question arises as to how to improve the frequency of SB targeting. An important consideration is that the SB transposase used in our experiments was fully functional in target DNA binding/capture. Thus, once in the nucleus, the SB transposase is probably confronted by vast numbers of potential TA target sites in the chromosomal DNA. Nonspecific binding of the transposase to human chromosomal DNA likely competes with specific binding to a desired targeted sequence, thereby limiting the chance and frequency at which a targeted transposition event

can occur. Such limitation could, in principle, be circumvented by interfering with the target DNA binding function of the transposase. It is not immediately evident whether SB transposase mutants deficient in target DNA binding but proficient in catalysis can be made.

We demonstrated efficient transposon targeting in the human genome; >10% of cells receiving transposon insertions contained at least one transposition event within the targeted chromosomal region (Fig. 38). Because the estimated theoretical frequency of hitting any TA in the human genome by random transposition is about 1 in a total of 10^8 transposition events, this technology represents on the order of 10^7 -fold enrichment of transgene insertion at a desired locus. It should be noted that the targeted site in our experiments was a 7-mer repeat of the tetracycline operator, and it is possible that the rate of targeted integration would decrease if the binding site were monomeric. Indeed, the efficiency of experimentally retargeted Ty5 retrotransposon integration was found to correlate with the number of target sites, suggesting that targeting efficiency was determined by the amount of targeting protein tethered to the target DNA (Zhu et al. 2003). Technologies for site-directed transgene integration could bear practical relevance in at least three potential areas of molecular genetics and therapy. First, targeted transposition could provide means for target-selected chromosomal engineering in experimental model systems where methods based on homologous recombination do not exist. Second, the ability to target transgene integration into loci associated with open chromatin and the potential to reliably and persistently express a transgene could minimize position effects and silencing of transgene expression. Finally, designed integration into safe regions in the human genome would reduce the potential genotoxic effects of transposon insertion, thereby contributing to an overall improvement of the safety profile of transposon-based gene vectors for human applications. Future work will have to focus on the identification of applicable, endogenous chromosomal target sites and the selection of DNA-binding proteins that can be exploited for efficiently targeting transposition into those sites *in vivo*.

5 CONCLUSIONS

There are several factors affecting site-selectivity of integrating vector systems. These include accessibility of specific chromosomal sites by chromatin components, primary sequence and physical structure of the DNA at the targeted region, endogenous expression of proteins that may compete for binding, and the specificity as well as capacity of chimeric proteins in DNA-binding as well as in catalytic functions. Both naturally targeted recombinase systems (such as ϕ C31) as well as targeting systems engineered from promiscuously integrating vectors (such as *Sleeping Beauty*) show off-target effects in the context of the human genome. For the former, the capacity of the recombinase to act at endogenous pseudo sites can lead to genomic rearrangements. For the latter, despite the fact that targeted integrations can be generated, non-targeted insertions can still occur at high frequencies, because the natural DNA-binding capacities of the transposase competes with that of the foreign DBD used for targeting. Keeping such off-target effects at a minimum remains a major challenge. Although several hurdles are yet to be overcome before technologies of targeted gene insertion can be considered for applications, recent evidence suggests that target-selected transgene insertion into desired regions in the human genome is a realistic goal.

6 SUMMARY

Transposons are discrete segments of DNA that have the distinctive ability to move and replicate within genomes. Transposons were discovered in the 1940's by Barbara McClintock (who later was awarded with the Nobel Prize for this discovery) in the maize genome, and have since been found ubiquitous in essentially all living organisms. The process of element movement is generally called transposition, and can contribute to insertional mutagenesis, altered gene expression and recombination. Transposons make up significant fractions of genomes; for example, about 45% of the human genome is composed of sequences of a variety of different elements. Transposons are best viewed as molecular parasites that

propagate themselves using resources of the host cell. Despite their parasitic nature, there is increasing evidence that transposable elements are a powerful force in gene evolution. Indeed, about 50 human genes are derived from transposable elements, among them genes that are responsible for immunoglobulin gene recombination in all vertebrates.

Transposons are natural gene delivery vehicles, and have been revolutionizing genomic manipulations in diverse model systems. Molecular reconstruction of the *Sleeping Beauty* (SB) transposon represents a milestone in applying transposon-based technologies for vertebrate genomics, including applications for functional genomics, transgenesis and gene therapy. SB shows efficient transposition and long-term transgene expression in cells of vertebrates, including humans. A variety of integrating vectors for gene delivery exist. Some of them exhibit random genomic integration, whereas others have integration preferences based on attributes of the targeted site, such as primary DNA sequence and physical structure of the DNA, or through tethering to certain DNA sequences by host-encoded cellular factors. Uncontrolled genomic insertion bears the risk of the transgene being silenced due to chromosomal position effects, and can lead to genotoxic effects due to mutagenesis of cellular genes. I review here the advantages and disadvantages of both viral and non-viral gene delivery technologies, discuss mechanisms of target site selection of integrating genetic elements (viruses and transposons), and suggest distinct molecular strategies for targeted gene delivery.

I provide experimental data describing attributes of target site selection of SB. Inspection of the DNA flanking the sites of element integration revealed significant differences from random DNA in both primary sequence and physical properties. The consensus sequence of SB target sites was found to be a palindromic AT-repeat, ATATATAT, in which the central TA is the canonical target site. We found however, that target site selection is primarily determined on the level of DNA structure, and not by specific base pair interactions. Computational analyses revealed that insertion sites tend to have a

bendable structure and a palindromic pattern of potential hydrogen-bonding sites in the major groove of the DNA. These features appear conserved in the Tc1/*mariner* family of transposons and in other, distantly related elements that share a common catalytic domain of the transposase, and integrate fairly randomly.

Random chromosomal transposition is clearly undesired for human gene therapeutic applications due to potential genotoxic effects associated with transposon integration. I set out to manipulate SB's target site selection for targeted transposition into predetermined chromosomal regions. I evaluated experimental strategies based on engineered proteins composed of DNA-binding domains, responsible for binding to chromosomal target DNA, fused to either the transposase or to another protein that interacts with the transposase or with the transposon DNA. I demonstrate targeted transposition into endogenous matrix attachment regions, and a chromosomally integrated tetracycline response element in cultured human cells, using targeting proteins that bind to the transposon DNA. An approach based on interactions between the transposase and a targeting protein containing the N-terminal protein interaction domain of SB was found to enable a $\sim 10^7$ -fold enrichment of transgene insertion at a desired locus. My experiments provide proof-of-principle for targeted chromosomal transposition of an otherwise randomly integrating transposon. Targeted transposition could be a powerful technology for safe transgene integration in human applications.

7 ACKNOWLEDGEMENTS

I am grateful to all members of the "Transposition" and "Mobile DNA" groups at the MDC for their dedicated work. Work in the author's laboratory is supported by EU FP6 grant INTHER LSHB-CT-2005018961, and a grant from the Deutsche Forschungsgemeinschaft SPP1230 "Mechanisms of gene vector entry and persistence".

8 REFERENCES

- Abdallah B, Sachs L, Demeneix BA (1995) Non-viral gene transfer: applications in developmental biology and gene therapy. *Biol. Cell* 85:1-7
- Akopian A, He J, Boocock MR, Stark WM (2003) Chimeric recombinases with designed DNA sequence recognition. *Proc Natl Acad Sci U S A* 100:8688-91
- Amsterdam A, Burgess S, Golling G, Chen W, Sun Z, Townsend K, Farrington S, et al. (1999) A large-scale insertional mutagenesis screen in zebrafish. *Genes Dev.* 13:2713-2724
- Ariyoshi M, Vassilyev DG, Iwasaki H, Nakamura H, Shinagawa H, Morikawa K (1994) Atomic structure of the RuvC resolvase: a Holliday junction-specific endonuclease from *E. coli*. *Cell* 78:1063-72
- Aye M, Dildine SL, Claypool JA, Jourdain S, Sandmeyer SB (2001) A truncation mutant of the 95-kilodalton subunit of transcription factor III λ reveals asymmetry in Ty3 integration. *Mol Cell Biol* 21:7839-51
- Babinet C, Morello D, Renard JP (1989) Transgenic mice. *Genome* 31:938-49
- Bachman N, Gelbart ME, Tsukiyama T, Boeke JD (2005) TFIIB subunit Bdp1p is required for periodic integration of the Ty1 retrotransposon and targeting of Isw2p to *S. cerevisiae* tDNAs. *Genes Dev* 19:955-64
- 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
- Balciunas D, Davidson AE, Sivasubbu S, Hermanson SB, Welle Z, Ekker SC (2004) Enhancer trapping in zebrafish using the Sleeping Beauty transposon. *BMC Genomics* 5:62
- Balciunas D, Wangenstein KJ, Wilber A, Bell J, Geurts A, Sivasubbu S, Wang X, et al. (2006) Harnessing a high cargo-capacity transposon for genetic applications in vertebrates. *PLoS Genet* 2:e169
- Barr SD, Leipzig J, Shinn P, Ecker JR, Bushman FD (2005) Integration targeting by avian sarcoma-leukosis virus and human immunodeficiency virus in the chicken genome. *J Virol* 79:12035-44
- Barry EG, Witherspoon DJ, Lampe DJ (2004) A bacterial genetic screen identifies functional coding sequences of the insect mariner transposable element Famar1 amplified from the genome of the earwig, *Forficula auricularia*. *Genetics* 166:823-33
- Baum C (2007) What are the consequences of the fourth case? *Mol Ther* 15:1401-2

- Baum C, von Kalle C, Staal FJ, Li Z, Fehse B, Schmidt M, Weerkamp F, et al. (2004) Chance or necessity? Insertional mutagenesis in gene therapy and its consequences. *Mol Ther* 9:5-13
- 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
- Beerli RR, Segal DJ, Dreier B, Barbas CF, 3rd (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc Natl Acad Sci U S A* 95:14628-33
- Belur LR, Frandsen JL, Dupuy AJ, Ingbar DH, Largaespada DA, Hackett PB, Scott McIvor R (2003) Gene insertion and long-term expression in lung mediated by the Sleeping Beauty transposon system. *Mol Ther* 8:501-7
- Bhasin A, Goryshin IY, Reznikoff WS (1999) Hairpin formation in Tn5 transposition. *J. Biol. Chem.* 274:37021-9
- Bowers WJ, Mastrangelo MA, Howard DF, Southerland HA, Maguire-Zeiss KA, Federoff HJ (2006) Neuronal precursor-restricted transduction via in utero CNS gene delivery of a novel bipartite HSV amplicon/transposase hybrid vector. *Mol Ther* 13:580-8
- Bradley D, Carpenter R, Sommer H, Hartley N, Coen E (1993) Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the *plena* locus of *Antirrhinum*. *Cell* 72:85-95
- Breitling R, Gerber JK (2000) Origin of the paired domain. *Dev Genes Evol* 210:644-50
- Brosius J (1991) Retroposons - seeds of evolution. *Science* 251:753
- Brukner I, Sanchez R, Suck D, Pongor S (1995) Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides. *EMBO J.* 14:1812-8
- Bryk M, Banerjee M, Murphy M, Knudsen KE, Garfinkel DJ, Curcio MJ (1997) Transcriptional silencing of Ty1 elements in the RDN1 locus of yeast. *Genes Dev* 11:255-69
- Buchholz F, Angrand PO, Stewart AF (1998) Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat Biotechnol* 16:657-62
- Bujacz G, Alexandratos J, Wlodawer A, Merkel G, Andrade M, Katz RA, Skalka AM (1997) Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity. *J Biol Chem* 272:18161-8
- Bushman F, Lewinski M, Ciuffi A, Barr S, Leipzig J, Hannenhalli S, Hoffmann C (2005) Genome-wide analysis of retroviral DNA integration. *Nat Rev Microbiol* 3:848-58

- Bushman FD (1994) Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. *Proc Natl Acad Sci U S A* 91:9233-7
- Bushman FD (2003) Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. *Cell* 115:135-8.
- Bushman FD, Miller MD (1997) Tethering human immunodeficiency virus type 1 preintegration complexes to target DNA promotes integration at nearby sites. *J Virol* 71:458-64
- Bustin M (1999) Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. *Mol. Cell. Biol.* 19:5237-5246
- Capy P, Vitalis R, Langin T, Higuete D, Bazin C (1996) Relationships between transposable elements based upon the integrase-transposase domains: is there a common ancestor? *J. Mol. Evol.* 42:359-68
- Carlson CM, Dupuy AJ, Fritz S, Roberg-Perez KJ, Fletcher CF, Largaespada DA (2003) Transposon mutagenesis of the mouse germline. *Genetics* 165:243-56.
- Carlson CM, Frandsen JL, Kirchhof N, McIvor RS, Largaespada DA (2005) Somatic integration of an oncogene-harboring Sleeping Beauty transposon models liver tumor development in the mouse. *Proc Natl Acad Sci U S A* 102:17059-64
- Casa AM, Brouwer C, Nagel A, Wang L, Zhang Q, Kresovich S, Wessler SR (2000) Inaugural article: the MITE family heartbreaker (Hbr): molecular markers in maize. *Proc Natl Acad Sci U S A* 97:10083-9
- Casola C, Hucks D, Feschotte C (2008) Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals. *Mol Biol Evol* 25:29-41
- Chalberg TW, Portlock JL, Olivares EC, Thyagarajan B, Kirby PJ, Hillman RT, Hoelters J, et al. (2006) Integration specificity of phage phiC31 integrase in the human genome. *J Mol Biol* 357:28-48
- Chen ZJ, Kren BT, Wong PY, Low WC, Steer CJ (2005) Sleeping Beauty-mediated down-regulation of huntingtin expression by RNA interference. *Biochem Biophys Res Commun* 329:646-52
- Chuah MK, Collen D, VandenDriessche T (2003) Biosafety of adenoviral vectors. *Curr Gene Ther* 3:527-43
- Chuah MK, Collen D, Vandendriessche T (2004) Preclinical and clinical gene therapy for haemophilia. *Haemophilia* 10 Suppl 4:119-25

- Chung T, Siol O, Dingermann T, Winckler T (2007) Protein interactions involved in tRNA gene-specific integration of Dictyostelium non-long terminal repeat retrotransposon TRE5-A. *Mol Cell Biol*
- Ciuffi A, Diamond TL, Hwang Y, Marshall HM, Bushman FD (2006) Modulating Target Site Selection During Human Immunodeficiency Virus DNA Integration In Vitro with an Engineered Tethering Factor. *Hum Gene Ther* 17:960-967
- Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, Shinn P, Ecker JR, et al. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med* 11:1287-9
- Collier LS, Carlson CM, Ravimohan S, Dupuy AJ, Largaespada DA (2005) Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature* 436:272-6
- Collins CH, Yokobayashi Y, Umeno D, Arnold FH (2003) Engineering proteins that bind, move, make and break DNA. *Curr Opin Biotechnol* 14:665
- Collins J, Forbes E, Anderson P (1989) The Tc3 family of transposable genetic elements in *Caenorhabditis elegans*. *Genetics* 121:47-55
- 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
- Cooley L, Kelley R, Spradling A (1988) Insertional mutagenesis of the *Drosophila* genome with single P elements. *Science* 239:1121-8
- Craig NL (1995) Unity in transposition reactions. *Science* 270:253-4
- Craig NL (1996) Transposon Tn7. *Curr Top Microbiol Immunol* 204:27-48
- Craigie R (1996) Quality control in Mu DNA transposition. *Cell* 85:137-40
- Cui Z, Geurts AM, Liu G, Kaufman CD, Hackett PB (2002) Structure-function analysis of the inverted terminal repeats of the *Sleeping Beauty* transposon. *J. Mol. Biol.* 318:1221-1235
- Curcio MJ, Derbyshire KM (2003) The outs and ins of transposition: from mu to kangaroo. *Nat Rev Mol Cell Biol* 4:865-77
- Czerny T, Schaffner G, Busslinger M (1993) DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site. *Genes Dev* 7:2048-61
- Davidson AE, Balciunas D, Mohn D, Shaffer J, Hermanson S, Sivasubbu S, Cliff MP, et al. (2003) Efficient gene delivery and gene expression in zebrafish using the *Sleeping Beauty* transposon. *Dev. Biol.* 263:191-202.
- Davies DR, Goryshin IY, Reznikoff WS, Rayment I (2000) Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* 289:77-85

- Deininger PL, Batzer MA (1999) Alu repeats and human disease. *Mol Genet Metab* 67:183-193
- Derbyshire KM, Hwang L, Grindley ND (1987) Genetic analysis of the interaction of the insertion sequence IS903 transposase with its terminal inverted repeats. *Proc Natl Acad Sci U S A* 84:8049-53
- Dewannieux M, Esnault C, Heidmann T (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* 35:41-8
- Dewannieux M, Heidmann T (2005) LINEs, SINEs and processed pseudogenes: parasitic strategies for genome modeling. *Cytogenet Genome Res* 110:35-48
- Ding S, Wu X, Li G, Han M, Zhuang Y, Xu T (2005) Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* 122:473-83
- 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. U.S.A.* 91:942-6.
- Dobbelstein M (2003) Viruses in therapy--royal road or dead end? *Virus Res.* 92:219-21.
- Drabek D, Zagoraiou L, deWit T, Langeveld A, Roumpaki C, Mamalaki C, Savakis C, et al. (2003) Transposition of the *Drosophila hydei* Minos transposon in the mouse germ line. *Genomics* 81:108-11.
- Driever W, Solnica-Krezel L, Schier AF, Neuhauss SC, Malicki J, Stemple DL, Stainier DY, et al. (1996) A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* 123:37-46
- Dupuy AJ, Akagi K, Largaespada DA, Copeland NG, Jenkins NA (2005) Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature* 436:221-6
- Dupuy AJ, Clark K, Carlson CM, Fritz S, Davidson AE, Markley KM, Finley K, et al. (2002) Mammalian germ-line transgenesis by transposition. *Proc. Natl. Acad. Sci. U.S.A.* 99:4495-9.
- Dupuy AJ, Fritz S, Largaespada DA (2001) Transposition and gene disruption in the male germline of the mouse. *Genesis* 30:82-88
- Dupuy AJ, Jenkins NA, Copeland NG (2006) Sleeping beauty: a novel cancer gene discovery tool. *Hum Mol Genet* 15 Spec No 1:R75-9
- 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* 266:1981-6

- Ehrhardt A, Engler JA, Xu H, Cherry AM, Kay MA (2006) Molecular analysis of chromosomal rearrangements in mammalian cells after phiC31-mediated integration. *Hum Gene Ther* 17:1077-94
- Ehrhardt A, Kay MA (2002) A new adenoviral helper-dependent vector results in long-term therapeutic levels of human coagulation factor IX at low doses in vivo. *Blood* 99:3923-30
- Emmons SW, Yesner L, Ruan KS, Katzenberg D (1983) Evidence for a transposon in *Caenorhabditis elegans*. *Cell* 32:55-65
- Engels W (1989). P elements in *Drosophila melanogaster*. Mobile DNA I. B. D. M and H. M. E. Washington D.C., American Society of Microbiology Washington D.C.: 437-483.
- Engels WR, Johnson-Schlitz DM, Eggleston WB, Sved J (1990) High-frequency P element loss in *Drosophila* is homolog dependent. *Cell* 62:515-25
- Essner JJ, McIvor RS, Hackett PB (2005) Awakening gene therapy with Sleeping Beauty transposons. *Curr Opin Pharmacol* 5:513-9
- Fadool JM, Hartl DL, Dowling JE (1998) Transposition of the mariner element from *Drosophila mauritiana* in zebrafish. *Proc Natl Acad Sci U S A* 95:5182-6
- Fanning T, Singer M (1987) The LINE-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retrovirus proteins. *Nucleic Acids Res* 15:2251-60
- Fedoroff N (2002). Control of mobile DNA. Mobile DNA II. N. L. Craig, R. Craigie, M. Gellert and A. M. Lambowitz. Washington DC, ASM Press: 836-869.
- Feng JA, Johnson RC, Dickerson RE (1994) Hin recombinase bound to DNA: the origin of specificity in major and minor groove interactions. *Science* 263:348-55
- Fischer SE, Wienholds E, Plasterk RH (2001) Regulated transposition of a fish transposon in the mouse germ line. *Proc. Natl. Acad. Sci. U.S.A.* 98:6759-64
- Franz G, Loukeris TG, Dialektaki G, Thompson CR, Savakis C (1994) Mobile *Minos* elements from *Drosophila hydei* encode a two-exon transposase with similarity to the paired DNA-binding domain. *Proc. Natl. Acad. Sci. U.S.A.* 91:4746-50
- Franz G, Savakis C (1991) *Minos*, a new transposable element from *Drosophila hydei*, is a member of the Tc1-like family of transposons. *Nucleic Acids Res.* 19:6646
- Garcia-Fernandez J, Bayascas-Ramirez JR, Marfany G, Munoz-Marmol AM, Casali A, Baguna J, Salo E (1995) High copy number of highly similar mariner-like transposons in planarian (Platyhelminthe): evidence for a trans-phyla horizontal transfer. *Mol Biol Evol* 12:421-31

- Garrick D, Fiering S, Martin DI, Whitelaw E (1998) Repeat-induced gene silencing in mammals. *Nat. Genet.* 18:56-9
- Ge H, Si Y, Roeder RG (1998) Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation. *Embo J* 17:6723-9
- Gellert M (2002) V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu. Rev. Biochem.* 71:101-132
- Geurts AM, Collier LS, Geurts JL, Oseth LL, Bell ML, Mu D, Lucito R, et al. (2006) Gene mutations and genomic rearrangements in the mouse as a result of transposon mobilization from chromosomal concatemers. *PLoS Genet* 2:e156
- Geurts AM, Wilber A, Carlson CM, Lobitz PD, Clark KJ, Hackett PB, McIvor RS, et al. (2006) Conditional gene expression in the mouse using a Sleeping Beauty gene-trap transposon. *BMC Biotechnol* 6:30
- Ginsburg DS, Calos MP (2005) Site-specific integration with phiC31 integrase for prolonged expression of therapeutic genes. *Adv Genet* 54:179-87
- Glover DJ, Lipps HJ, Jans DA (2005) Towards safe, non-viral therapeutic gene expression in humans. *Nat Rev Genet* 6:299-310
- 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 U S A* 95:9150-4
- Gorin AA, Zhurkin VB, Olson WK (1995) B-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.* 247:34-48
- Gossen M, Freundlieb S, Bender G, Muller G, Hillen W, Bujard H (1995) Transcriptional activation by tetracyclines in mammalian cells. *Science* 268:1766-9
- Goulaouic H, Chow SA (1996) Directed integration of viral DNA mediated by fusion proteins consisting of human immunodeficiency virus type 1 integrase and Escherichia coli LexA protein. *J Virol* 70:37-46
- Grabher C, Henrich T, Sasado T, Arenz A, Wittbrodt J, Furutani-Seiki M (2003) Transposon-mediated enhancer trapping in medaka. *Gene* 322:57-66
- Haapa-Paananen S, Rita H, Savilahti H (2002) DNA transposition of bacteriophage Mu. A quantitative analysis of target site selection in vitro. *J. Biol. Chem.* 277:2843-2851
- Hacein-Bey-Abina S, Le Deist F, Carlier F, Bouneaud C, Hue C, De Villartay JP, Thrasher AJ, et al. (2002) Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy. *N Engl J Med* 346:1185-93

- Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, Lim A, et al. (2003) LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 302:415-9
- Hackett PB, Ekker SC, Largaespada DA, McIvor RS (2005) Sleeping beauty transposon-mediated gene therapy for prolonged expression. *Adv Genet* 54:189-232
- Haffter P, Granato M, Brand M, Mullins MC, Hammerschmidt M, Kane DA, Odenthal J, et al. (1996) The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* 123:1-36
- Hallet B, Rezsohazy R, Mahillon J, Delcour J (1994) IS231A insertion specificity: consensus sequence and DNA bending at the target site. *Mol. Microbiol.* 14:131-139.
- Hama C, Ali Z, Kornberg TB (1990) Region-specific recombination and expression are directed by portions of the *Drosophila* engrailed promoter. *Genes Dev* 4:1079-93
- Hamlet MR, Yergeau DA, Kuliyeve E, Takeda M, Taira M, Kawakami K, Mead PE (2006) Tol2 transposon-mediated transgenesis in *Xenopus tropicalis*. *Genesis* 44:438-45
- Haren L, Ton-Hoang B, Chandler M (1999) Integrating DNA: transposases and retroviral integrases. *Annu Rev Microbiol* 53:245-81
- Hartl DL (2000) Molecular melodies in high and low C. *Nat Rev Genet* 1:145-9
- Hartl DL, Lozovskaya ER, Nurminsky DI, Lohe AR (1997) What restricts the activity of *mariner*-like transposable elements? *Trends Genet.* 13:197-201
- Hauer B, Shapiro JA (1984) Control of Tn7 transposition. *Mol Gen Genet* 194:149-58
- He CX, Shi D, Wu WJ, Ding YF, Feng DM, Lu B, Chen HM, et al. (2004) Insulin expression in livers of diabetic mice mediated by hydrodynamics-based administration. *World J Gastroenterol* 10:567-72
- Hedges RW, Jacob AE (1974) Transposition of ampicillin resistance from RP4 to other replicons. *Mol Gen Genet* 132:31-40
- Heggestad AD, Notterpek L, Fletcher BS (2004) Transposon-based RNAi delivery system for generating knockdown cell lines. *Biochem Biophys Res Commun* 316:643-50
- Horie K, Kuroiwa A, Ikawa M, Okabe M, Kondoh G, Matsuda Y, Takeda J (2001) Efficient chromosomal transposition of a Tc1/*mariner*-like transposon *Sleeping Beauty* in mice. *Proc. Natl. Acad. Sci. U.S.A.* 98:9191-9196
- Horie K, Yusa K, Yae K, Odajima J, Fischer SE, Keng VW, Hayakawa T, et al. (2003) Characterization of *Sleeping Beauty* transposition and its application to genetic screening in mice. *Mol. Cell. Biol.* 23:9189-207.
- Houck MA, Clark JB, Peterson KR, Kidwell MG (1991) Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. *Science* 253:1125-8

- Huang X, Wilber AC, Bao L, Tuong D, Tolar J, Orchard PJ, Levine BL, et al. (2006) Stable gene transfer and expression in human primary T cells by the Sleeping Beauty transposon system. *Blood* 107:483-91
- Ichikawa H, Ikeda K, Amemura J, Ohtsubo E (1990) Two domains in the terminal inverted-repeat sequence of transposon Tn3. *Gene* 86:11-7
- Ivanov VI, Minchenkova LE (1995) The A-form of DNA: in search of the biological role. *Mol. Biol.* 28:1258-71
- Ivics Z, Hackett PB, Plasterk RH, Izsvak Z (1997) Molecular reconstruction of *Sleeping Beauty*, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* 91:501-10
- Ivics Z, Izsvak Z (2006) Transposons for gene therapy! *Curr Gene Ther* 6:593-607
- Ivics Z, Izsvák Z (2004) Transposable elements for transgenesis and insertional mutagenesis in vertebrates: a contemporary review of experimental strategies. *Methods Mol. Biol.* 260:255-76.
- Ivics Z, Izsvák Z, Minter A, Hackett PB (1996) Identification of functional domains and evolution of Tc1-like transposable elements. *Proc. Natl. Acad. Sci. U.S.A.* 93:5008-13.
- Izsvák Z, Ivics Z (2004) *Sleeping Beauty* transposition: biology and applications for molecular therapy. *Mol. Ther.* 9:147-56.
- Izsvák Z, Ivics Z, Hackett PB (1995) Characterization of a Tc1-like transposable element in zebrafish (*Danio rerio*). *Mol. Gen. Genet.* 247:312-22
- Izsvák Z, Ivics Z, Plasterk RH (2000) *Sleeping Beauty*, a wide host-range transposon vector for genetic transformation in vertebrates. *J. Mol. Biol.* 302:93-102
- Izsvák Z, Khare D, Behlke J, Heinemann U, Plasterk RH, Ivics Z (2002) Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in *Sleeping Beauty* transposition. *J. Biol. Chem.* 277:34581-34588
- Izsvák Z, Stüwe EE, Fiedler D, Katzer A, Jeggo PA, Ivics Z (2004) Healing the wounds inflicted by *Sleeping Beauty* transposition by double-strand break repair in mammalian somatic cells. *Mol. Cell* 13:279-290
- Jain C, Kleckner N (1993) Preferential cis action of IS10 transposase depends upon its mode of synthesis. *Mol Microbiol* 9:249-60
- Jones JM, Gellert M (2004) The taming of a transposon: V(D)J recombination and the immune system. *Immunol Rev* 200:233-48
- Jurka J, Klonowski P, Trifonov EN (1998) Mammalian retroposons integrate at kinkable DNA sites. *J. Biomol. Struct. Dyn.* 15:717-21

- Kafri T, Morgan D, Krah T, Sarvetnick N, Sherman L, Verma I (1998) Cellular immune response to adenoviral vector infected cells does not require de novo viral gene expression: implications for gene therapy. *Proc Natl Acad Sci U S A* 95:11377-82
- Kapitonov VV, Jurka J (2001) Rolling-circle transposons in eukaryotes. *Proc Natl Acad Sci U S A* 98:8714-9
- Kassis JA, Noll E, VanSickle EP, Odenwald WF, Perrimon N (1992) Altering the insertional specificity of a *Drosophila* transposable element. *Proc Natl Acad Sci U S A* 89:1919-23
- Katayanagi K, Miyagawa M, Matsushima M, Ishikawa M, Kanaya S, Ikehara M, Matsuzaki T, et al. (1990) Three-dimensional structure of ribonuclease H from *E. coli*. *Nature* 347:306-9
- Katz RA, Merkel G, Skalka AM (1996) Targeting of retroviral integrase by fusion to a heterologous DNA binding domain: in vitro activities and incorporation of a fusion protein into viral particles. *Virology* 217:178-90
- Kaufman CD, Izsvak Z, Katzer A, Ivics Z (2005) *Frog Prince* transposon-based RNAi vectors mediate efficient gene knockdown in human cells. *J. RNAi Gene Silenc.* 1:97-104
- 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
- Kawakami K, Noda T (2004) Transposition of the *Tol2* element, an *Ac*-like element from the Japanese medaka fish *Oryzias latipes*, in mouse embryonic stem cells. *Genetics* 166:895-9.
- Kawakami K, Shima A, Kawakami N (2000) Identification of a functional transposase of the *Tol2* element, an *Ac*-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage. *Proc Natl Acad Sci U S A* 97:11403-8
- Kawakami K, Takeda H, Kawakami N, Kobayashi M, Matsuda N, Mishina M (2004) A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. *Dev Cell* 7:133-44
- Kazazian HHJ (1999) An estimated frequency of endogenous insertional mutations in humans. *Nat Genet* 22:130
- Ke Z, Grossman GL, Cornel AJ, Collins FH (1996) *Quetzal*: A transposon of the Tc1 family in the mosquito *Anopheles albimanus*. *Genetica* 98:141-147
- Keng VW, Yae K, Hayakawa T, Mizuno S, Uno Y, Yusa K, Kokubu C, et al. (2005) Region-specific saturation germline mutagenesis in mice using the Sleeping Beauty transposon system. *Nat Methods* 2:763-9

- Kennedy AK, Guhathakurta A, Kleckner N, Haniford DB (1998) Tn10 transposition via a DNA hairpin intermediate. *Cell* 95:125-34
- Kidwell MG (1992) Horizontal transfer. *Curr. Opin. Genet. Dev.* 2:868-873
- Kidwell MG (1992) Horizontal transfer of P elements and other short inverted repeat transposons. *Genetica* 86:275-86
- Kidwell MG, Lisch DR (2001) Transposable elements, parasitic DNA, and genome evolution. *Int J Org Evol* 55:1-24
- Kidwell MG, Lisch DR (2002). Mobile DNA II. N. L. Craig, R. Craigie, M. Gellert and A. M. Lambowitz. Washington DC, ASM Press: 60.
- Kile BT, Hentges KE, Clark AT, Nakamura H, Salinger AP, Liu B, Box N, et al. (2003) Functional genetic analysis of mouse chromosome 11. *Nature* 425:81-6
- Kim A, Terzian C, Santamaria P, Pelisson A, Purd'homme N, Bucheton A (1994) Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 91:1285-9
- Kim DR, Dai Y, Mundy CL, Yang W, Oettinger MA (1999) Mutations of acidic residues in RAG1 define the active site of the V(D)J recombinase. *Genes Dev* 13:3070-80
- Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res* 8:464-78
- Kipp M, Gohring F, Ostendorp T, van Drunen CM, van Driel R, Przybylski M, Fackelmayer FO (2000) SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA. *Mol Cell Biol* 20:7480-9
- Kirchner J, Connolly CM, Sandmeyer SB (1995) Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. *Science* 267:1488-91
- Kitada K, Ishishita S, Tosaka K, Takahashi RI, Ueda M, Keng VW, Horie K, et al. (2007) Transposon-tagged mutagenesis in the rat. *Nat Methods*
- Kleckner N, Chalmers RM, Kwon D, Sakai J, Bolland S (1996) Tn10 and IS10 transposition and chromosome rearrangements: mechanism and regulation in vivo and in vitro. *Curr Top Microbiol Immunol* 204:49-82
- Klobutcher LA, Herrick G (1995) Consensus inverted terminal repeat sequence of *Paramecium* IESs: resemblance to termini of Tc1-related and *Euplotes* Tec transposons. *Nucleic Acids Res* 23:2006-13
- Koga A, Suzuki M, Inagaki H, Bessho Y, Hori H (1996) Transposable element in fish. *Nature* 383:30.

- Kuduvalli PN, Mitra R, Craig NL (2005) Site-specific Tn7 transposition into the human genome. *Nucleic Acids Res* 33:857-63
- Kuduvalli PN, Rao JE, Craig NL (2001) Target DNA structure plays a critical role in Tn7 transposition. *EMBO J.* 20:924-32
- 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* 12:2331-8
- Kunze R (1996) The maize transposable element activator (Ac). *Curr Top Microbiol Immunol* 204:161-94
- Lam WL, Seo P, Robison K, Virk S, Gilbert W (1996) Discovery of amphibian Tc1-like transposon families. *J. Mol. Biol.* 257:359-66
- Lampe DJ, Churchill ME, Robertson HM (1996) A purified *mariner* transposase is sufficient to mediate transposition *in vitro*. *EMBO J.* 15:5470-9
- Lampe DJ, Grant TE, Robertson HM (1998) Factors affecting transposition of the *Himar1 mariner* transposon *in vitro*. *Genetics* 149:179-87
- Lampe DJ, Walden KK, Robertson HM (2001) Loss of transposase-DNA interaction may underlie the divergence of mariner family transposable elements and the ability of more than one mariner to occupy the same genome. *Mol Biol Evol* 18:954-61
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860-921
- Landree MA, Wibbenmeyer JA, Roth DB (1999) Mutational analysis of RAG1 and RAG2 identifies three catalytic amino acids in RAG1 critical for both cleavage steps of V(D)J recombination. *Genes Dev* 13:3059-69
- Laski FA, Rio DC, Rubin GM (1986) Tissue specificity of *Drosophila* P element transposition is regulated at the level of mRNA splicing. *Cell* 44:7-19
- Lavoie BD, Chaconas G (1996) Transposition of phage Mu DNA. *Curr Top Microbiol Immunol* 204:83-102
- Lee GS, Neiditch MB, Sinden RR, Roth DB (2002) Targeted transposition by the V(D)J recombinase. *Mol. Cell. Biol.* 22:2068-2077.
- Leung PC, Teplow DB, Harshey RM (1989) Interaction of distinct domains in Mu transposase with Mu DNA ends and an internal transpositional enhancer. *Nature* 338:656-658

- Lewinski MK, Yamashita M, Emerman M, Ciuffi A, Marshall H, Crawford G, Collins F, et al. (2006) Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog* 2:e60
- Li S, Ma Z (2001) Nonviral gene therapy. *Curr Gene Ther* 1:201-26
- Liao GC, Rehm EJ, Rubin GM (2000) Insertion site preferences of the P transposable element in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* 97:3347-51
- Linden RM, Winocour E, Berns KI (1996) The recombination signals for adeno-associated virus site-specific integration. *Proc Natl Acad Sci U S A* 93:7966-72
- Liu J, Jeppesen I, Nielsen K, Jensen TG (2006) Phi c31 integrase induces chromosomal aberrations in primary human fibroblasts. *Gene Ther* 13:1188-90
- Liu L, Mah C, Fletcher BS (2006) Sustained FVIII Expression and Phenotypic Correction of Hemophilia A in Neonatal Mice Using an Endothelial-Targeted Sleeping Beauty Transposon. *Mol Ther* 13:1006-15
- Liu L, Sanz S, Heggstad AD, Antharam V, Notterpek L, Fletcher BS (2004) Endothelial targeting of the Sleeping Beauty transposon within lung. *Mol Ther* 10:97-105
- Liu ZJ, Moav B, Faras AJ, Guise KS, Kapuscinski AR, Hackett PB (1990) Development of expression vectors for transgenic fish. *Biotechnology* 8:1268-72.
- Llano M, Vanegas M, Fregoso O, Saenz D, Chung S, Peretz M, Poeschla EM (2004) LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. *J Virol* 78:9524-37
- Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky SL, et al. (2004). Molecular Cell Biology. New York, Freeman.
- Lohe AR, Moriyama EN, Lidholm DA, Hartl DL (1995) Horizontal transmission, vertical inactivation, and stochastic loss of *mariner*-like transposable elements. *Mol. Biol. Evol.* 12:62-72
- Lohe AR, Timmons C, Beerman I, Lozovskaya ER, Hartl DL (2000) Self-inflicted wounds, template-directed gap repair and a recombination hotspot. Effects of the mariner transposase. *Genetics* 154:647-56
- Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, et al. (2007) Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol* 25:1298-306
- Loomis WF, Welker D, Hughes J, Maghakian D, Kuspa A (1995) Integrated maps of the chromosomes in *Dictyostelium discoideum*. *Genetics* 141:147-57

- Loonstra A, Vooijs M, Beverloo HB, Allak BA, van Drunen E, Kanaar R, Berns A, et al. (2001) Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. *Proc Natl Acad Sci U S A* 98:9209-14
- Lu B, Geurts AM, Poirier C, Petit DC, Harrison W, Overbeek PA, Bishop CE (2007) Generation of rat mutants using a coat color-tagged Sleeping Beauty transposon system. *Mamm Genome* 18:338-46
- Luan DD, Korman MH, Jakubczak JL, Eickbush TH (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72:595-605
- Luo G, Ivics Z, Izsvák Z, Bradley A (1998) Chromosomal transposition of a Tc1/*mariner*-like element in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* 95:10769-73
- Machida C, Machida Y (1989) Regulation of IS1 transposition by the insA gene product. *J Mol Biol* 208:567-74
- Mahillon J, Chandler M (1998) Insertion sequences. *Microbiol Mol Biol Rev* 62:725-74
- Makalowski W (2003) Genomics. Not junk after all. *Science* 300:1246-7
- Malik HS, Burke WD, Eickbush TH (1999) The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol* 16:793-805
- Mandell JG, Barbas CF, 3rd (2006) Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* 34:W516-23
- Manno CS, Arruda VR, Pierce GF, Glader B, Ragni M, Rasko J, Ozelo MC, et al. (2006) Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. *Nat Med* 12:342-7
- Maragathavally KJ, Kaminski JM, Coates CJ (2006) Chimeric Mos1 and piggyBac transposases result in site-directed integration. *Faseb J* 20:1880-2
- Mates L, Izsvák Z, Ivics Z (2007) Technology transfer from worms and flies to vertebrates: transposition-based genome manipulations and their future perspectives. *Genome Biol* 8 Suppl 1:S1
- McCarty DM, Young SM, Jr., Samulski RJ (2004) Integration of adeno-associated virus (AAV) and recombinant AAV vectors. *Annu Rev Genet* 38:819-45
- McClure MA (1991) Evolution of retroposons by acquisition or deletion of retrovirus-like genes. *Mol Biol Evol* 8:835-56
- Medhora M, Maruyama K, Hartl DL (1991) Molecular and functional analysis of the mariner mutator element Mos1 in *Drosophila*. *Genetics* 128:311-8

- Merriman PJ, Grimes CD, Ambroziak J, Hackett DA, Skinner P, Simmons MJ (1995) *S* elements: a family of Tc1-like transposons in the genome of *Drosophila melanogaster*. *Genetics* 141:1425-38
- Miller DG, Rutledge EA, Russell DW (2002) Chromosomal effects of adeno-associated virus vector integration. *Nat. Genet.* 30:147-8.
- Miskey C, Izsvak Z, Kawakami K, Ivics Z (2005) DNA transposons in vertebrate functional genomics. *Cell Mol Life Sci* 62:629-41
- Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC, Ecker JR, et al. (2004) Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol* 2:E234
- Mizuuchi K (1992) Polynucleotidyl transfer reactions in transpositional DNA recombination. *J Biol Chem* 267:21273-6
- Mizuuchi K (1997) Polynucleotidyl transfer reactions in site-specific DNA recombination. *Genes Cells* 2:1-12
- Montini E, Held PK, Noll M, Morcinek N, Al-Dhalimy M, Finegold M, Yant SR, et al. (2002) In vivo correction of murine tyrosinemia type I by DNA-mediated transposition. *Mol Ther* 6:759-69
- Moran JV, Gilbert N (2002). Mammalian LINE-1 retrotransposons and related elements. Mobile DNA II. N. L. Craig, R. Craigie, M. Gellert and A. M. Lambowitz. Washington DC, ASM Press: 836-869.
- Moschetti R, Caggese C, Barsanti P, Caizzi R (1998) Intra- and interspecies variation among *Bari-1* elements of the *melanogaster* species group. *Genetics* 150:239-50
- Muller HP, Varmus HE (1994) DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. *EMBO J.* 13:4704-14
- Nakai H, Montini E, Fuess S, Storm TA, Grompe M, Kay MA (2003) AAV serotype 2 vectors preferentially integrate into active genes in mice. *Nat Genet* 34:297-302
- Nasevicius A, Ekker SC (2000) Effective targeted gene 'knockdown' in zebrafish. *Nat Genet* 26:216-20
- Niidome T, Huang L (2002) Gene therapy progress and prospects: nonviral vectors. *Gene Ther.* 9:1647-52.
- Nowak R (1994) Mining treasures from 'junk DNA'. *Science* 263:608-10
- Ohlfest JR, Demorest ZL, Motooka Y, Vengco I, Oh S, Chen E, Scappaticci FA, et al. (2005) Combinatorial antiangiogenic gene therapy by nonviral gene transfer using the sleeping beauty transposon causes tumor regression and improves survival in mice bearing intracranial human glioblastoma. *Mol Ther* 12:778-88

- Ohlfest JR, Frandsen JL, Fritz S, Lobitz PD, Perkinson SG, Clark KJ, Nelsestuen G, et al. (2005) Phenotypic correction and long-term expression of factor VIII in hemophilic mice by immunotolerization and nonviral gene transfer using the Sleeping Beauty transposon system. *Blood* 105:2691-8
- Ohno S (1972) So much "junk" DNA in our genome. *Brookhaven Symp Biol* 23:366-70
- Olson WK, Gorin AA, Lu XJ, Hock LM, Zhurkin VB (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. U.S.A.* 95:11163-8
- Orgel LE, Crick FH (1980) Selfish DNA: the ultimate parasite. *Nature* 284:604-7
- Ortiz-Urda S, Thyagarajan B, Keene DR, Lin Q, Fang M, Calos MP, Khavari PA (2002) Stable nonviral genetic correction of inherited human skin disease. *Nat Med* 8:1166-70
- Parinov S, Kondrichin I, Korzh V, Emelyanov A (2004) Tol2 transposon-mediated enhancer trap to identify developmentally regulated zebrafish genes in vivo. *Dev Dyn* 231:449-59
- Pellizzari L, Tell G, Damante G (1999) Co-operation between the PAI and RED subdomains of Pax-8 in the interaction with the thyroglobulin promoter. *Biochem J* 337 (Pt 2):253-62
- Peters JE, Craig NL (2001) Tn7: smarter than we thought. *Nat Rev Mol Cell Biol* 2:806-14
- Pickeral OK, Makalowski W, Boguski MS, Boeke JD (2000) Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Res* 10:411-5
- Petrokovski S, Henikoff S (1997) A helix-turn-helix DNA-binding motif predicted for transposases of DNA transposons. *Mol. Gen. Genet.* 254:689-95
- Plasterk RH (1991) The origin of footprints of the Tc1 transposon of *Caenorhabditis elegans*. *Embo J* 10:1919-25
- Plasterk RH (1996) The Tc1/mariner transposon family. *Curr Top Microbiol Immunol* 204:125-43
- Plasterk RH, Izsvak Z, Ivics Z (1999) Resident aliens: the Tc1/*mariner* superfamily of transposable elements. *Trends Genet.* 15:326-32
- Pribil PA, Haniford DB (2000) Substrate recognition and induced DNA deformation by transposase at the target-capture stage of Tn10 transposition. *J. Mol. Biol.* 303:145-159.
- Pruss D, Reeves R, Bushman FD, Wolffe AP (1994) The influence of DNA and nucleosome structure on integration events directed by HIV integrase. *J. Biol. Chem.* 269:25031-25041

- Raz E, van Luenen HG, Schaerringer B, Plasterk RH, Driever W (1998) Transposition of the nematode *Caenorhabditis elegans* Tc3 element in the zebrafish *Danio rerio*. *Curr. Biol.* 8:82-8
- Richet E, Abcarian P, Nash H (1986) Synapsis of attachment sites during lambda integrative recombination involves capture of a naked DNA by a protein-DNA complex. *Cell* 46:1011-1021
- Rio DC, Barnes G, Laski FA, Rine J, Rubin GM (1988) Evidence for *Drosophila* P element transposase activity in mammalian cells and yeast. *J. Mol. Biol.* 200:411-5
- Robertson HM (1993) The *mariner* transposable element is widespread in insects. *Nature* 362:241-245
- Robertson HM (1995) The Tc1-*mariner* superfamily of transposons in animals. *J. Insect Physiol.* 41:99-105
- Robertson HM, Zuppano KL (1997) Molecular evolution of an ancient *mariner* transposon, *Hsmar1*, in the human genome. *Gene* 205:203-217
- Rubin GM, Kidwell MG, Bingham PM (1982) The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. *Cell* 29:987-94
- Russell WL, Kelly EM, Hunsicker PR, Bangham JW, Maddux SC, Phipps EL (1979) Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. *Proc. Natl. Acad. Sci. U.S.A.* 76:5818-9
- Sarkar I, Hauber I, Hauber J, Buchholz F (2007) HIV-1 proviral DNA excision using an evolved recombinase. *Science* 316:1912-5
- Sassaman DM, Dombroski BA, Moran JV, Kimberland ML, Naas TP, DeBerardinis RJ, Gabriel A, et al. (1997) Many human L1 elements are capable of retrotransposition. *Nat Genet* 16:37-43
- Sauer B, Henderson N (1990) Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. *New Biol* 2:441-9
- Scherdin U, Rhodes K, Breindl M (1990) Transcriptionally active genome regions are preferred targets for retrovirus integration. *J. Virol.* 64:907-912
- Schiedner G, Morral N, Parks RJ, Wu Y, Koopmans SC, Langston C, Graham FL, et al. (1998) Genomic DNA transfer with a high-capacity adenovirus vector results in improved in vivo gene expression and decreased toxicity. *Nat Genet* 18:180-3
- Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F (2002) HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110:521-9.

- Selinger DA, Chandler VL (1999) Major recent and independent changes in levels and patterns of expression have occurred at the b gene, a regulatory locus in maize. *Proc Natl Acad Sci U S A* 96:15007-12
- Sijen T, Plasterk RH (2003) Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature* 426:310-4
- Silva JC, Loreto EL, Clark JB (2004) Factors that affect the horizontal transfer of transposable elements. *Curr Issues Mol Biol* 6:57-71
- Sinn PL, Sauter SL, McCray PB, Jr. (2005) Gene therapy progress and prospects: development of improved lentiviral and retroviral vectors--design, biosafety, and production. *Gene Ther* 12:1089-98
- Sinzelle L, Pollet N, Bigot Y, Mazabraud A (2005) Characterization of multiple lineages of Tc1-like elements within the genome of the amphibian *Xenopus tropicalis*. *Gene* 349:187-96
- Sinzelle L, Vallin J, Coen L, Chesneau A, Du Pasquier D, Pollet N, Demeneix B, et al. (2006) Generation of transgenic *Xenopus laevis* using the Sleeping Beauty transposon system. *Transgenic Res.* (In press):
- Sivasubbu S, Balciunas D, Davidson AE, Pickart MA, Hermanson SB, Wangenstein KJ, Wolbrink DC, et al. (2006) Gene-breaking transposon mutagenesis reveals an essential role for histone H2afza in zebrafish larval development. *Mech Dev* 123:513-29
- Spanopoulou E, Zaitseva F, Wang FH, Santagata S, Baltimore D, Panayotou G (1996) The homeodomain region of Rag-1 reveals the parallel mechanisms of bacterial and V(D)J recombination. *Cell* 87:263-76
- Spradling AC, Stern DM, Kiss I, Roote J, Laverly T, Rubin GM (1995) Gene disruptions using P transposable elements: an integral component of the *Drosophila* genome project. *Proc Natl Acad Sci U S A* 92:10824-30
- Szabo M, Muller F, Kiss J, Balduf C, Strahle U, Olsz F (2003) Transposition and targeting of the prokaryotic mobile element IS30 in zebrafish. *FEBS Lett* 550:46-50
- Tan W, Dong Z, Wilkinson TA, Barbas CF, 3rd, Chow SA (2006) Human immunodeficiency virus type 1 incorporated with fusion proteins consisting of integrase and the designed polydactyl zinc finger protein E2C can bias integration of viral DNA into a predetermined chromosomal region in human cells. *J Virol* 80:1939-48
- Tan W, Zhu K, Segal DJ, Barbas CF, 3rd, Chow SA (2004) Fusion proteins consisting of human immunodeficiency virus type 1 integrase and the designed polydactyl zinc finger protein E2C direct integration of viral DNA into specific sites. *J Virol* 78:1301-13

- Thomas CE, Ehrhardt A, Kay MA (2003) Progress and problems with the use of viral vectors for gene therapy. *Nat. Rev. Genet.* 4:346-58
- Thorpe HM, Smith MC (1998) In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc Natl Acad Sci U S A* 95:5505-10
- Thorrez L, VandenDriessche T, Collen D, Chuah MK (2004) Preclinical gene therapy studies for hemophilia using adenoviral vectors. *Semin Thromb Hemost* 30:173-83
- Thrasher AJ, Gaspar HB (2007) Severe Adverse Event in Clinical Trial of Gene Therapy for X-SCID. [ASGT press release](#)
- Thyagarajan B, Guimaraes MJ, Groth AC, Calos MP (2000) Mammalian genomes contain active recombinase recognition sites. *Gene* 244:47-54
- Thyagarajan B, Olivares EC, Hollis RP, Ginsburg DS, Calos MP (2001) Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol Cell Biol* 21:3926-34
- Tower J, Karpen GH, Craig N, Spradling AC (1993) Preferential transposition of *Drosophila* P elements to nearby chromosomal sites. *Genetics* 133:347-59
- Trono D (1995) HIV accessory proteins: leading roles for the supporting cast. *Cell* 82:189-92
- Turlan C, Chandler M (2000) Playing second fiddle: second-strand processing and liberation of transposable elements from donor DNA. *Trends Microbiol.* 8:268-274
- Ullu E, Tschudi C (1984) Alu sequences are processed 7SL RNA genes. *Nature* 312:171-2
- Urcelay E, Ward P, Wiener SM, Safer B, Kotin RM (1995) Asymmetric replication in vitro from a human sequence element is dependent on adeno-associated virus Rep protein. *J Virol* 69:2038-46
- van Gent DC, Mizuuchi K, Gellert M (1996) Similarities between initiation of V(D)J recombination and retroviral integration. *Science* 271:1592-4
- van Luenen HG, Colloms SD, Plasterk RH (1994) The mechanism of transposition of Tc3 in *C. elegans*. *Cell* 79:293-301
- van Luenen HG, Plasterk RH (1994) Target site choice of the related transposable elements Tc1 and Tc3 of *Caenorhabditis elegans*. *Nucleic Acids Res.* 22:262-9
- VandenDriessche T, Collen D, Chuah MK (2003) Biosafety of onco-retroviral vectors. *Curr Gene Ther* 3:501-15
- Vastenhouw NL, Plasterk RH (2004) RNAi protects the *Caenorhabditis elegans* germline against transposition. *Trends Genet* 20:314-9
- Verma IM, Somia N (1997) Gene therapy - Promises, problems and prospects. *Nature* 389:239-242

- Vigdal TJ, Kaufman CD, Izsvák Z, Voytas DF, Ivics Z (2002) Common physical properties of DNA affecting target site selection of *Sleeping Beauty* and other Tc1/*mariner* transposable elements. *J. Mol. Biol.* 323:441-452
- Vos JC, Plasterk RH (1994) Tc1 transposase of *Caenorhabditis elegans* is an endonuclease with a bipartite DNA binding domain. *EMBO J.* 13:6125-32
- Walisko O, Schorn A, Rolfs F, Devaraj A, Miskey C, Izsvák Z, Ivics Z (2008) Transcriptional activities of the Sleeping Beauty transposon and shielding its genetic cargo with insulators. *Mol Ther* 16:359-69
- Weichenrieder O, Wild K, Strub K, Cusack S (2000) Structure and assembly of the Alu domain of the mammalian signal recognition particle. *Nature* 408:167-73
- 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
- Weitzman MD, Kyostio SR, Kotin RM, Owens RA (1994) Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. *Proc Natl Acad Sci U S A* 91:5808-12
- Wessler SR (1996) Turned on by stress. Plant retrotransposons. *Curr Biol* 6:959-61
- Wilber A, Frandsen JL, Geurts JL, Largaespada DA, Hackett PB, McIvor RS (2006) RNA as a source of transposase for sleeping beauty-mediated gene insertion and expression in somatic cells and tissues. *Mol Ther* 13:625-30
- Wilson MH, Kaminski JM, George AL, Jr. (2005) Functional zinc finger/sleeping beauty transposase chimeras exhibit attenuated overproduction inhibition. *FEBS Lett* 579:6205-9
- Winckler T, Dingermann T, Glockner G (2002) Dictyostelium mobile elements: strategies to amplify in a compact genome. *Cell Mol Life Sci* 59:2097-111
- Winckler T, Szafranski K, Glockner G (2005) Transfer RNA gene-targeted integration: an adaptation of retrotransposable elements to survive in the compact Dictyostelium discoideum genome. *Cytogenet Genome Res* 110:288-98
- Winkler FK, Banner DW, Oefner C, Tsernoglou D, Brown RS, Heathman SP, Bryan RK, et al. (1993) The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.* 12:1781-1795
- Wu SC, Meir YJ, Coates CJ, Handler AM, Pelczar P, Moisyadi S, Kaminski JM (2006) piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. *Proc Natl Acad Sci U S A* 103:15008-13

- Wu X, Li Y, Crise B, Burgess SM (2003) Transcription start regions in the human genome are favored targets for MLV integration. *Science* 300:1749-51.
- Xie W, Gai X, Zhu Y, Zappulla DC, Sternglanz R, Voytas DF (2001) Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. *Mol Cell Biol* 21:6606-14
- Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J* 9:3353-62
- Yant SR, Ehrhardt A, Mikkelsen JG, Meuse L, Pham T, Kay MA (2002) Transposition from a gutless adeno-transposon vector stabilizes transgene expression in vivo. *Nat Biotechnol* 20:999-1005
- Yant SR, Huang Y, Akache B, Kay MA (2007) Site-directed transposon integration in human cells. *Nucleic Acids Res* 35:e50
- Yant SR, Meuse L, Chiu W, Ivics Z, Izsvak Z, Kay MA (2000) Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. *Nat. Genet.* 25:35-41
- Yant SR, Wu X, Huang Y, Garrison B, Burgess SM, Kay MA (2005) High-resolution genome-wide mapping of transposon integration in mammals. *Mol Cell Biol* 25:2085-94
- Yergeau DA, Mead PE (2007) Manipulating the *Xenopus* genome with transposable elements. *Genome Biol* 8 Suppl 1:S11
- Yieh L, Hatzis H, Kassavetis G, Sandmeyer SB (2002) Mutational analysis of the transcription factor IIIB-DNA target of Ty3 retroelement integration. *J Biol Chem* 277:25920-8
- Yu Y, Bradley A (2001) Engineering chromosomal rearrangements in mice. *Nat Rev Genet* 2:780-90
- Zagoraoui L, Drabek D, Alexaki S, Guy JA, Klinakis AG, Langeveld A, Skavdis G, et al. (2001) *In vivo* transposition of *Minos*, a *Drosophila* mobile element, in mammalian tissues. *Proc. Natl. Acad. Sci. U.S.A.* 98:11474-11478
- Zaiss AK, Muruve DA (2005) Immune responses to adeno-associated virus vectors. *Curr Gene Ther* 5:323-31
- Zambrowicz BP, Friedrich GA (1998) Comprehensive mammalian genetics: history and future prospects of gene trapping in the mouse. *Int. J. Dev. Biol.* 42:1025-36
- Zayed H, Izsvak Z, Khare D, Heinemann U, Ivics Z (2003) The DNA-bending protein HMGB1 is a cellular cofactor of *Sleeping Beauty* transposition. *Nucleic Acids Res.* 31:2313-22

- Zhu Y, Dai J, Fuerst PG, Voytas DF (2003) Controlling integration specificity of a yeast retrotransposon. *Proc Natl Acad Sci U S A* 100:5891-5
- Zou S, Ke N, Kim JM, Voytas DF (1996) The *Saccharomyces* retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci. *Genes Dev.* 10:634-645
- Zwaal RR, Broeks A, van Meurs J, Groenen JT, Plasterk RH (1993) Target-selected gene inactivation in *Caenorhabditis elegans* by using a frozen transposon insertion mutant bank. *Proc Natl Acad Sci U S A* 90:7431-5