**Sleeping Beauty Transposition**

ZOLTÁN IVICS\(^1\) and ZSUZSANNA IZSVÁK\(^2\)

\(^1\)Division of Medical Biotechnology, Paul Ehrlich Institute, Langen, Germany;  
\(^2\)Max Delbrück Center for Molecular Medicine, Berlin, Germany

**ABSTRACT** Sleeping Beauty (SB) is a synthetic transposon that was constructed based on sequences of transpositionally inactive elements isolated from fish genomes. SB is a Tc1/mariner superfamily transposon following a cut-and-paste transpositional reaction, during which the element-encoded transposase interacts with its binding sites in the terminal inverted repeats of the transposon, promotes the assembly of a synaptic complex, catalyzes excision of the element out of its donor site, and integrates the excised transposon into a new location in target DNA. SB transposition is dependent on cellular host factors. Transcriptional control of transposase expression is regulated by the HMG2L1 transcription factor. Synaptic complex assembly is promoted by the HMGB1 protein and regulated by chromatin structure. SB transposition is highly dependent on the nonhomologous end joining (NHEJ) pathway of double-strand DNA break repair that generates a transposon footprint at the excision site. Through its association with the Miz-1 transcription factor, the SB transposase downregulates cyclin D1 expression that results in a slowdown of the cell-cycle in the G1 phase, where NHEJ is preferentially active. Transposon integration occurs at TA dinucleotides in the target DNA, which are duplicated at the flanks of the integrated transposon. SB shows a random genome-wide insertion profile in mammalian cells when launched from episomal vectors and “local hopping” when launched from chromosomal donor sites. Some of the excised transposons undergo a self-destructive autointegration reaction, which can partially explain why longer elements transpose less efficiently. SB became an important molecular tool for transgenesis, insertional mutagenesis, and gene therapy.

**Sleeping Beauty KISSED BACK TO LIFE – A SHORT HISTORY**

Members of the Tc1/mariner superfamily are probably the most widespread DNA transposons in nature. However, these elements appear to be transpositionally inactive in vertebrates due to the accumulation of mutations. In an attempt to isolate potentially active copies, we surveyed a number of fish genomes for the presence of Tc1-like elements from 11 different species. In summary, all the Tc1-like elements that we \(^2\) and others \(^3, 4\) described from the different fish species were defective copies carrying inactivating mutations that accumulated over long evolutionary times. Nevertheless, careful sequence analysis allowed us to predict a consensus sequence that would likely represent an active archetypal sequence. We have engineered this sequence by eliminating the inactivating mutations from the transposase open reading frame. The resurrected synthetic transposon was named Sleeping Beauty (SB), in analogy of the Grimm brothers’ famous fairy tale. SB can be identical or closely related to an ancient transposon that once successfully invaded several fish genomes, in part by horizontal transmission between species. The resurrection of SB was the first demonstration that ancient transposable elements can be brought back to life. Before this work was published in 1997, there was no indication that any DNA-based transposon was active in vertebrates. SB not only represents the first DNA-based transposon ever shown to be active in cells of vertebrates, but the first functional gene ever reconstructed from inactive, ancient genetic material, for which an active, naturally occurring copy either does not exist or has not yet been isolated.
STRUCTURAL AND FUNCTIONAL COMPONENTS OF THE Sleeping Beauty TRANSPOSON

The SB transposon has a simple structure. In its natural form, it consists of a single gene encoding the transposase polypeptide, the enzymatic factor of transposition, which is flanked by terminal inverted repeats (IRs) containing binding sites for the transposase [Fig. 1(A)] (3). The transposase gene can be physically separated from the IRs, and replaced by other DNA sequences [Fig. 1(B)]. This is because the transposase can mobilize transposons in trans, as long as they retain the IRs. SB transposes through a conservative, cut-and-paste mechanism, during which the transposable element is excised from its original location by the transposase, and is integrated into a new location.

THE TRANSPOSASE

The DNA-binding domain

The overall domain structure of the transposase is conserved in the entire Tc1/mariner superfamily (1). Specific substrate recognition is mediated by an N-terminal, bipartite DNA-binding domain of the transposase [Fig. 2] (6, 7, 8). This DNA-binding domain has been proposed to consist of two helix-turn-helix (HTH) motifs, similar to the paired domain of some transcription factors in both amino acid sequence and structure (2, 8, 9). The modular paired domain has evolved versatility in binding to a range of different DNA sequences through various combinations of its subdomains (PAI+RED) (10). The origin of the paired domain is not clear, but phylogenetic analyses indicate that it might have been derived from an ancestral transposase (11).

The first of these HTH motifs has been crystallized in complex with double-stranded DNA corresponding to the termini of Tc3 transposons in Caenorhabditis elegans (12). The crystal structure indeed showed a HTH fold, and a dimer of transposase subunits bringing together the two DNA ends. The recently described NMR solution structure of the PAI subdomain of the SB transposase identified amino acid residues located in the second and third alpha helices forming the HTH motif to be involved in binding to DNA [Fig. 3] (13).

FIGURE 1 The SB transposon system. (A) Structure of the SB transposon. The central transposase gene (purple box) is flanked by terminal IRs (black arrows) that contain binding sites for the transposase (white arrows). (B) Gene transfer vector system based on SB. 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). Reprinted from Current Gene Therapy (161) with permission from the publisher. doi:10.1128/microbiolspec.MDNA3-0042-2014.f1
We found that a GRPR-like sequence (GRRR in SB) between the two HTH motifs is conserved in Tc1/mariner transposases (Fig. 2). The GRPR motif is similar to an AT-hook (6), and characteristic to homeodomain proteins (14). It mediates minor groove interactions with DNA in the case of the Hin invertase of Salmonella (15) and in the V(D)J recombination activating gene (RAG1) recombinase (see V(D)J chapter) (16).

Partially overlapping with the RED subdomain in the transposase is a bipartite nuclear localization signal (NLS) (Fig. 2), flanked by phosphorylation target sites of casein kinase II (2). 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.

The catalytic domain
The second major domain of the SB transposase has been referred to as the catalytic domain, because it is

---

**FIGURE 2** Structural and functional components of SB. On top, a schematic drawing of the transposon is shown. The terminal IR/DR (black arrows) contain two binding sites for the transposase (white arrows). The element contains a single gene encoding the transposase (purple box). The transposase has an N-terminal, bipartite, paired-like DNA-binding domain containing a GRRR AT-hook motif, an NLS, and a C-terminal catalytic domain. The DNA binding domain consists of a PAI and a RED subdomain containing helix-turn-helix DNA-binding motifs. The DDE amino acid triad is a characteristic signature of the catalytic domain that catalyzes the DNA cleavage and joining reactions. Reprinted from *Molecular Therapy* (164) with permission from the publisher. doi:10.1128/microbiolspec.MDNA3-0042-2014.f2

**FIGURE 3** Structures of the PAI subdomain of the SB transposase and the DNA-bound N-terminal DNA-binding subdomains of the Tc3 and Mos1 mariner transposases and the Pax5 transcription factor. Residues on the second and third alpha-helices of the SB PAI subdomain are directly involved in DNA-binding. Reprinted from *Protein Science* (13) with permission from the publisher. doi:10.1128/microbiolspec.MDNA3-0042-2014.f3
responsible for the DNA cleavage and joining reactions of transposition. The majority of known transposases, including *SB*, and retroviral integrases possess a well-conserved triad of amino acids, known as the aspartate–aspartate–glutamate, in short the DDE signature in their C-terminal catalytic domain (Fig. 2) (17, 18, 19, 20). These amino acids play an essential role in catalysis by coordinating, in general, two divalent cations necessary for activity. 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 nucleophile during trans-esterification and strand transfer (21). The biologically relevant cation required for the catalytic steps of transposition is thought to be Mg**+** (22, 23).

Site-directed mutagenesis of the DDE residues in the related Tc3 transposase confirmed that these three amino acids are indeed essential for all catalytic activities (24). Similarly, a DAE variant of the *SB* transposase is catalytically dead (25). Although the crystal structure of the *SB* transposase catalytic domain is yet to be solved, the available Mos1 *mariner* structure from *Drosophila mauritiana* (26) is assumed to be the closest to model the structure of the *SB* transposase. Intriguingly, crystallographic analyses of the catalytic domains of transposases and some other proteins whose functions are not obviously related to transposition, such as RNaseH (27) or RuvC (28), have revealed a remarkably similar overall fold. Besides similarities, there are important differences between *mariner* and *SB* transposases. First, while *SB* has the characteristic DDE motif in its catalytic domain, this motif in *mariner*s is DDD. Interestingly, a change of the exceptional third D of *mariner*, turning the DDD into the canonical DDE, inactivates the transposase (29), and the reciprocal substitution in the *SB* transposase (DDD) is also inactive (30). A conserved glycine-rich subdomain can be found within the catalytic domains of *SB* and Tc1-like transposases (3). This glycine-rich subdomain is not present in *mariner*s or retroviral integrases, and its function is yet to be determined. The emerging picture reinforces the notion of a common structural motif that catalyzes polynucleotidyl transfer reactions in diverse biological contexts (31, 32), and 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/D recombinases (18).

**Molecular evolution of the SB100X hyperactive Sleeping Beauty transposase**

In evolutionary terms, the *SB* transposon was a successful element with the ability to colonize several fish genomes millions of years ago (2). However, even successful transposons have not been selected for the highest possible activity. On the contrary, there is strong selective pressure to avoid insertional mutagenesis of essential genes of their host. In an attempt to derive hyperactive transposase variants for advanced genetic engineering, amino acid substitutions spanning almost the entire *SB* transposase polypeptide have been screened for eliciting a change in catalytic activity. These amino acid replacements were conducted either by systematic alanine-scanning (33), by “importing” single amino acids or small (2 to 7 aa) blocks of amino acids from related transposases (34, 35, 36), and by rational replacement of selected amino acid residues based on charge (35). These approaches generated transposase variants with (i) no change in activity; (ii) reduced activity, or (iii) a relatively modest increase of transposition activity. The vast majority of the mutations have a neutral or a negative effect on transposition activity. The inactivating mutations generally map to the evolutionary conserved domains of the transposase. Surprisingly, some combinations of hyperactive mutations were found to result in a significant reduction of activity. Nevertheless, a strategy of identifying those hyperactive variants that acted in an additive or synergistic manner combined with a high throughput genetic screening of ∼2,000 possible combinations yielded *SB* transposase variants with significantly enhanced activities (30). The most hyperactive version, *SB100X*, displays a ∼100-fold hyperactivity when compared to the originally resurrected transposase (Fig. 4) (30). The hyperactivity of *SB100X* cannot be explained by altered transposase stability, nor by increased binding to the transposon IRs; instead, the particular combination of mutations in *SB100X* appears to affect the folding properties of the transposase (30). The use of the *SB100X* system yielded robust gene transfer efficiencies into human hematopoietic progenitors (30, 37), mesenchymal stem cells, muscle stem/progenitor cells (myoblasts), iPSCs (38), and T cells (39). These cells are relevant targets for stem cell biology and for regenerative medicine and gene- and cell-based therapies of complex genetic diseases. Moreover, the *SB100X* transposase enables highly efficient germline transgenesis in relevant mammalian models, including mice, rats, rabbits, and pigs (40, 41, 42).

**The inverted repeats of the Sleeping Beauty transposon**

Similarly to most transposon ends, the IRs of *SB* are composed of two functional parts. The 2 to 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.

The IRs of SB are 200 to 250 bp long and carry a pair of transposase-binding sites within the ends of each IR, characterized by short, 15 to 20 bp direct repeats (DRs) (Fig. 1). This special organization of IRs is termed IR/DR (1, 43), and can be found in numerous elements in the Tc1 transposon family, including the Minos, S, Paris and Bari elements in various Drosophila species (1, 44, 45, 46), Quetzal elements in mosquitos (47), at least three Tc1-like transposon subfamilies in fish (2), and Txr, Eagle, Froggy, and Jumpy transposons in Xenopus (48, 49). The spacing of about 200 bp between the outer and inner DRs 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 differs significantly from Tc1 or the mariner elements, which are simpler and have repeats of less than 100 bp and a single transposase binding site per repeat (8, 50). The four DRs of SB are not identical; the outer ones are longer by two base pairs. The IRs are not identical either; the left IR contains a sequence motif called “half-DR” (HDR), which resembles the 3′-half of the transposase binding sites (6). A construct containing two left IRs transposes more efficiently 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 (6). The multiple binding sites of the IR/DR elements likely impose control over the timing and specificity of the transposition reaction (see below).

THE MOLECULAR MECHANISM OF Sleeping Beauty TRANSPOSITION AND ITS REGULATION BY HOST-ENCODED FACTORS

The typical “cut-and-paste” transposition process of SB can be divided into at least four major steps: (1) binding of the transposase to its sites within the transposon IRs; (2) formation of a synaptic complex in which the two ends of the elements are paired and held together by transposase subunits; (3) excision from the donor site; (4) reintegration at a target site (Fig. 5).

The wide phylogenetic distribution of the Tc1/mariner family suggested no or weak host factor requirement of the transposition reaction. Supporting this assumption, the activity of Tc1 was reconstructed in vitro, and the reaction required only the presence of the transposon and the transposase (51). In addition, the nematode Tc3 element was demonstrated to jump in zebrafish; however, this trans-species transposition reaction was not efficient (52). Despite earlier assumptions, SB transposition turned out to be highly dependent on cellular host factors (described in detail in sections below) (53, 54, 55, 56) and became an excellent model system to study transposon–host interactions in higher eukaryotes (57).
SB can transpose in a wide range of vertebrate cells from fish to human, although the efficiency of transposition varies significantly (58), suggesting that differential interactions between the transposon and host-encoded factors may affect activity and eventually limit the host range. Indeed, transposition of SB seems to be restricted to vertebrates, with the exception of a chordate, Ciona intestinalis (59). Importantly, the identified host factors of SB are evolutionarily conserved in vertebrates and support SB transposition from fish to human. In summary, the regulation of SB transposition is mediated both by transposon- and host-encoded factors. Thus, the SB transposon has an intimate relationship with the host that likely modulates transposition at every step of the transposition reaction (Fig. 5).

Transcriptional regulation of the **Sleeping Beauty** transposon

Some of the 5'-untranslated regions (UTRs) upstream of the initiation codon of the transposase gene contain promoter motifs (60), suggesting that they might have functions associated with control of transposition activity. However, previous studies did not reveal an internal promoter in the Tc1 element; instead they showed that the elements are transcribed by read-through transcription from *C. elegans* genes (61). The left IR of SB is separated from the transposase coding sequence by a 160-bp stretch of DNA [Fig. 6(A)] with no apparent function in the transposition reaction (35). As measured by transient luciferase reporter assays, transcription driven by the 5'-UTR of SB is ~18-fold higher.

**FIGURE 5** Mechanism and regulation of SB transposition. The transposable element consists of a gene encoding a transposase (orange box) bracketed by terminal IRs (solid black arrows) that contain binding sites of the transposase (white arrows) and flanking donor DNA (blue boxes). Transcriptional control elements in the 5'-UTR of the transposon drive transcription (arrow) of the transposase gene. The transposase (purple spheres) binds to its sites within the transposon IRs. Excision takes place in a synaptic complex, and separates the transposon from the donor DNA. The excised element integrates into a TA site in the target DNA (green box) that is duplicated and flanks the newly integrated transposon. On the right, the various steps of transposition are shown. On the left, mechanisms and host factors regulating each step of the transposition reaction are indicated. Reprinted from (57) with permission from the publisher. doi:10.1128/microbiolspec.MDNA3-0042-2014.f5
FIGURE 6 The UTRs of the SB transposon exhibit moderate, directional promoter activities. (A) Transcriptional activities residing within the SB transposon. On top, a schematic drawing of the transposon is shown. The terminal IRs contain two binding sites for the transposase (white arrows). The element contains a single gene encoding the transposase (purple box). Relative promoter activities as determined by transient luciferase assays in HeLa cells. Activity of a minimal promoter (TATA-box) control was arbitrarily set to value 1. Transposon sequences flanking the transposase gene were placed in front of a luciferase reporter gene in two possible orientations (in the case of the 5′-UTR, the luciferase gene precisely replaces the transposase coding region). Blue box: left IR of SB; green box: right IR of SB; beige box: left IR of Frog Prince; black lines connecting the IRs and the luciferase gene represent transposon sequences directly upstream of the transposase coding regions. The 5′-UTR of SB can drive transposase expression at a level sufficient for the detection of chromosomal transposition events in cultured cells. A neo-tagged SB transposon plasmid was cotransfected together with an SB expression construct, in which the transposase is expressed from the 5′-UTR of the transposon or with an empty cloning vector. The difference in numbers of G418-resistant cell colonies is evidence for transposition. (B) A model for transcriptional regulation of the SB transposase gene. In the wild-type, natural transposon, the central transposase gene (purple box) is flanked by UTRs that include the left and right inverted repeats (IRs, blue and green boxes, respectively) that contain binding sites for the transposase (white arrows). Arrows indicate the direction of transcription that is initiated within the UTRs. HMG2L1 upregulates, whereas SB transposase downregulates transcription from the 5′-UTR. Reprinted from Molecular Therapy (54) with permission from the publisher. doi:10.1128/microbiolspec.MDNA3-0042-2014.f6
than transcription of a promoter-less sequence, ∼4.6-fold higher than transcription driven by a TATA-box minimal promoter, and ∼2.5-fold higher than transcription driven by the 5′-UTR of the closely related Frog Prince (FP) transposon [Fig. 6(A)] (54). The 5′-UTR drives expression of the SB transposase at a level sufficient to detect SB transposition in a colony-forming transposition assay in HeLa cells [Fig. 6(A)]. The right IR can also drive expression towards the inside of the element, but at a lower efficiency than the 5′-UTR [Fig. 6(A)]. Convergent transcription of SB transposons raises the possibility of the formation of transposon-specific double-stranded RNA molecules that may serve as triggers of transposition regulation by RNA interference (54, 55). Recent NMR data confirmed differential binding of the DNA-binding domain of the SB transposase to the inner and outer DRs (13). The PAI subdomain also binds to the HDR motif within the left IR of SB and mediates protein–protein interactions with other transposase subunits (6). 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 (6). 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 (6). Zebrafish Tdr1 elements are closely related to SB, but are not mobilized by the 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.

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 (6). Specificity of DNA-binding is predominantly determined by base-specific interactions mediated by the PAI subdomain (6).

All four binding sites within the IR/DR structure are required for SB transposition (58). The paired-like DNA-binding domain forms tetramers in complex with transposase binding sites (6). The inner DRs are more strongly bound by the transposase than the outer DRs (53, 65). A cellular host factor, HMG2L1 (alias HMGXB4), was identified as a physical interacting partner of the SB transposase and shown to upregulate transcription from the 5′-UTR of SB by 10- to 15-fold (54) [Fig. 6(B)]. HMG2L1 is an HMG-box DNA-binding domain-containing protein that shares structural similarity with lymphocyte enhancer-binding factor 1 (LEF-1), sex-determining region Y (SRY), and SRY-related HMG-box protein 4 (SOX4) transcription factors (64). HMG-box transcription factors specifically bind their target DNA through their HMG-box domains, and regulate transcription of target genes (64). Indeed, in addition to its interaction with the transposase, HMG2L1 also interacts with the SB transposon DNA in vivo, as shown by chromatin immunoprecipitation (ChIP) experiments (54). Interestingly, co-expression of the SB transposase with HMG2L1 has a repressing effect on transcription by the 5′-UTR [Fig. 6(B)]. Thus, transposase expression in the context of the naturally occurring SB transposon is subject to negative feedback regulation, with the transposase acting as a transcriptional repressor. This model postulates a sensitive balance in the regulation of transposase expression that is calibrated by transposase concentrations in the cell: low concentrations allow more transposase to be made, whereas high concentrations lead to shutting off transposase expression.

Ordered assembly of synaptic complexes and the role of HMGB1 in Sleeping Beauty transposition

A uniform requirement among transposition reactions is the formation of a nucleoprotein complex, before transposon excision 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. 5). The necessary factors that are required for synaptic complex assembly of SB include the complete 1Rs 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
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 the Mu transposase, where the two specificities of binding to the enhancer and to the recombination sites are encoded in two distinct domains (66), the paired-like region of SB transposase combines these two functions in a single protein domain. As described above, the SB transposase preferentially binds the inner DRs within the transposon IRs. This suggests that the unequal strengths of transposase binding and the positions of the DRs within the IRs 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 high mobility group protein HMGB1 is required for efficient SB transposition in mammalian cells (53). HMG family members have been shown to also be required for V(D)J recombination (67, 68). HMGB1 is an abundant, nonhistone, nuclear protein associated with eukaryotic chromatin and has the ability to bend DNA (69). SB transposition is significantly reduced in HMGB1-deficient mouse cells (53). This effect was complemented by expressing HMGB1 and HMGB2, but not with the more distantly related HMG A1 protein. Overexpression of HMGB1 in wild-type cells enhanced transposition, indicating that HMGB1 is a limiting factor of transposition (53). HMGs have low affinity for standard B-form DNA, and interactor proteins need to guide them to certain sites (69). SB transposase was found to interact with HMGB1 in vivo and to form a ternary complex with the transposase and transposon DNA (Fig. 7) (53), 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 (53). 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 IRs has a critical importance. It was also found that HMGB1 enhances transposase binding to both DRs, but its effect is significantly more pronounced at the inner sites (53). 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 (65). This ordered assembly process probably controls that cleavage at the outer sites occurs only if all the previous requirements have been fulfilled.

In summary, the IR/DR-type organization of IRs introduces a higher-level regulation into the transposition process. These elements might have evolved novel “built in regulatory checkpoints” to enforce synopsis prior to catalysis, thereby ensuring a higher level of accuracy and fidelity during the transposition process compared to transposons with simply structured IRs (53, 56, 70). 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. Such strictly regulated assembly of catalytically primed complexes could suppress unpaired reaction products or promiscuous synapses of distant ends of the transposon (70).
Regulation of transposon excision by DNA CpG methylation

CpG methylation of chromosomal DNA leads to formation of heterochromatin and is known to decrease or inhibit transpositional activity of diverse transposons (71). Surprisingly, CpG methylation of the SB transposon was found to enhance transpositional activity in mouse embryonic stem (ES) cells (72). It was subsequently found that the enhancing effect of CpG methylation is not restricted to SB but, rather, seems to be an intrinsic feature associated with the characteristic IR/DR structure of the SB, Frog Prince, and Minos elements (73). At which step(s) of cut-and-paste transposition does the effect of CpG methylation manifest? Because (i) CpG methylation has no effect on SB transposase binding, (ii) CpG methylation induces the formation of a condensed chromatin structure (heterochromatin), and (iii) DNA compaction by protamine enhances transposition, a model was proposed in which CpG methylation and subsequent chromatin condensation aids synaptic complex formation. Because heterochromatin formation results in tight packaging of DNA and histones, DNA sites that are usually far away from each other – for example, the two transposase binding sites in the IRs – might be brought closer together (Fig. 8). The physical proximity of the inner and outer binding sites might assist the formation of transposase dimers as soon as they bind, thereby facilitating the formation of a catalytically active synaptic complex. Thus, similarly to the effect of HMGB1, conformational changes of the excising transposon may greatly influence

FIGURE 8 A model for the enhancing effect of a compact chromatin structure on SB transposition. Euchromatin contains DNA wrapped around nucleosomes in a “beads-along-a-string”-like conformation (upper panel). Transposase subunits bound within the transposon IRs are separated by 166 bp DNA. Heterochromatin (lower panel), characterized by DNA CpG methylation and specific histone tail modifications, e.g., trimethylated lysine 9 of histone H3, features a higher histone : DNA ratio. Positioning of a nucleosome between the transposase binding sites (small orange arrows) shortens the distance between these sites, thereby facilitating the formation of a transposase dimer per IR and subsequent assembly of the synaptic complex. Reprinted from Mobile DNA (73) with permission from the publisher. doi:10.1128/mobileDNA3-0042-2014.f8
the efficiency of transposition. Assuming that the transposase source is provided by a transcriptionally active element located in euchromatin, host-cell-induced CpG methylation/heterochromatin-based silencing of transposons can be offset by a higher transposition efficiency out of condensed chromatins, thereby constituting a potential mechanism of SB and other similar-structured transposons to escape CpG methylation-mediated silencing imposed by the host.

**Excision of the Sleeping Beauty transposon and double-strand break repair of excision sites**

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 (74). Every DNA strand cleavage in all transposition reactions is a transposase-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 (75), which begin with a single-strand nick that generates a free 3′-OH group. In the case of the first strand cleavage the nucleophile is H₂O (74). During cut-and-paste transposition, nicking of the element is followed by the cleavage of the complementary DNA strand. To catalyze second strand cleavage, DDE recombinases developed versatile strategies (76). Most DDE transposases use a single active site to cleave both strands of DNA at the transposon end via a DNA–hairpin intermediate (77, 78, 79, 80, 81). For example, in V(D)J recombination the single-strand nick is converted into a double-stranded break (DSB) by a transestrafication reaction in which the free 3′-OH attacks the opposite strand, thereby creating a hairpin intermediate at the donor site (82, 83). Tn5 and Tn10 transposons also transpose via a hairpin intermediate, with the difference that the hairpin is on the transposon and not on the flanking DNA (84, 85). However, mariner (86) and SB (55) transposition does not proceed through a hairpin intermediate, and the exact mechanism of second-strand cleavage remains unknown. Thus, in the absence of a hairpin intermediate, mariner and SB transposases likely cleave the two strands of the DNA at each transposon end by sequential hydrolysis reactions.

Strand 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 the 3′-cleavage site in V(D)J recombination (82) and for the bacterial Tn5 (87) and Tn10 elements (85) (thereby generating blunt ended products). In the case of the Tc1/mariner elements the non-transferred strand is cleaved a few nucleotides within the transposon [two nucleotides for the Tc1 and Tc3 elements (24, 50) and three nucleotides inside the element in the case of mariner and SB (Fig. 9) (50, 88)]. Thus, SB transposon excision leaves behind three-nucleotide-long 3′-overhangs (Fig. 9), which are processed by the DNA repair mechanisms of the cells leaving a transposon “footprint“ at the transposon donor site (see below).

The DSBs generated by transposon excision are repaired either by the nonhomologous end joining pathway (NHEJ) or by homologous recombination (HR) (89, 90). The main factors that mediate NHEJ are a complex of DNA ligase IV and Xrcc4 and the DNA-dependent protein kinase (DNA-PK), a serine/threonine protein kinase (91, 92). DNA-PK consists of a catalytic subunit (DNA-PKcs) and a DNA-binding subunit termed Ku. The Ku heterodimer (composed of Ku70 and Ku80) binds to DNA ends and facilitates DSB repair by recruiting DNA-PKcs and additional factors such as the Xrcc4/DNA ligase IV complex to the site of damage.

The prominent pathway of repairing transposon excision sites in somatic mammalian cells is NHEJ, which generates transposon “footprints” that are identical to the first or last 2 to 4 nucleotides of the transposon in Tc1/mariner transposition (88, 93). NHEJ of the three-nucleotide-long 3′-overhangs left behind by SB excision generates a 3-bp footprint (Fig. 9) (55, 93). Factors of the NHEJ pathway of DSB repair, including Ku70 and DNA-PKcs, are required for SB transposition by repairing the transposon excision sites (55). NHEJ components have also been shown to be required for efficient retroelement integration and V(D)J recombination (92, 94). Ku70 physically interacts with the SB transposase (55), suggesting that it might be involved in shepherding excision site repair to NHEJ.

NHEJ and HR have overlapping roles in maintaining chromosomal integrity in vertebrate cells (95), and they can serve as alternative pathways for repair of the same DSB (96). Although the NHEJ pathway of DSB repair plays a dominant role in repair of transposon excision sites in somatic cells, the dependence of SB transposition on NHEJ factors is not absolute. In contrast to V(D)J recombination (92), HR can also be involved in excision site repair during SB transposition (55). Similarly, both NHEJ and HR play significant roles in the repair of DSBs generated by Tc1 excision in C. elegans (97), in P-element transposition in Drosophila (98, 99), and in Ty1 retrotransposon integration in yeast (100, 101).
These observations support the view that DSBs generated by radiation, V(D)J recombination, retroviral integration, and DNA transposition require overlapping but different factors for repair. The interplay between the repair factors and the recombination machineries probably determines how mechanistically similar processes can produce different products.

**Sleeping Beauty** transposase modulates cell-cycle progression through interaction with Miz-1

The Myc-interacting zinc finger protein 1 (Miz-1) transcription factor (102) was identified as an interactor of the SB transposase in a yeast two-hybrid screen (103). Miz-1 is a transcriptional regulator of the cyclin D1 gene (102), and it downregulates the cyclin D1 promoter resulting in slower cell growth. Decreased cellular levels of cyclin D1 prevent cells from entering the S phase, resulting in cell-cycle arrest in the G1 phase (104). Strikingly, through its physical association with Miz-1, the SB transposase seems to downregulate cyclin D1 expression in human cells, resulting in a cell-cycle slowdown (Fig. 10) (103).

The likely biological significance of our finding is that, by inducing a temporary G1 delay, the SB transposase potentiates the involvement of NHEJ to repair transposition-inflicted DNA damage (55). Indeed, a delay in the G1/S transition and S phase progression by cell-cycle checkpoints is thought to facilitate DNA repair to avoid replication and subsequent propagation of potentially hazardous mutations. In eukaryotic cells, the NHEJ and HR pathways are complementary but act at different stages of the cell-cycle: NHEJ is preferentially active in
the G1 and early S phases (105), whereas HR is active in the late S and G2 phases (95). Accordingly, there is increasing evidence for a correlation between the particular pathway used for the repair of transposon-induced DNA damage and the cell-cycle stage where recombination occurs. This is nicely illustrated by gene rearrangements through V(D)J recombination, which is tightly linked to the G1 phase of the cell-cycle and to NHEJ (92, 106). As described above, DSBs generated by SB transposition are preferentially repaired by the NHEJ pathway (55, 107), and the SB transposase physically interacts with the Ku DNA-binding subunit of DNA-PK, a key component of the NHEJ machinery (55). The data suggest a model in which SB transposase induces a cyclin D1-dependent G1 slowdown in proliferating cells through interaction with Miz-1, thereby ensuring that transposon-induced DNA damage is repaired by NHEJ. In nature, preferential use of NHEJ for the repair of transposon-induced DSBs might help avoid homologous recombination events between dispersed copies of transposable elements in the genome, thereby assisting the maintenance of genomic stability.

Other parasitic genetic elements, including HIV-1 (108), Herpes simplex virus (109), cytomegalovirus (110), Epstein-Barr virus (111), Kaposi’s sarcoma-associated herpesvirus (112), and mouse hepatitis virus (113), have also developed versatile strategies to perturb the cellular machinery to maximize their chance for survival and propagation. Thus, overriding the normal cell-cycle program seems to be a shared strategy of parasitic genetic elements.

Transposon integration: target site selection properties of Sleeping Beauty

The second step of the transposition reaction is the transfer of the exposed 3′-OH transposon tip to the target DNA molecule by transesterification (Fig. 9). Similarly to the initial DNA cut, the strand transfer is executed 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 strands, the other end joining to a displaced position of the other target strand. Due to this staggered fashion of the strand joining reaction, and because the inserted element has 3′-overhangs, the integration is flanked by single-stranded gaps (Fig. 9). DNA repair at these gaps restores the terminal nucleotides of the inserted transposon and generates a characteristic duplication of the target site flanking the element that is called target site duplication (TSD) (Fig. 9). SB transposition almost exclusively occurs at TA dinucleotides (5, 114, 115, 116, 117, 118), and SB integrants therefore are flanked by TA TSDs (Fig. 9), which are molecular hallmarks of SB transposition. At very low frequencies, non-TA target sites were also found at insertions generated by SB100X (115, 119). About 75% of SB transposon excision events are coupled to chromosomal integration (93) and no extrachromosomal, excised molecules are readily detectable. The ~25% loss of excised transposons might be due to challenge of productive transposition by suicidal autointegration, i.e., when the transposon integrates into itself (see below).

The genome-wide insertion pattern of most transposons is non-random, showing characteristic preferences for insertion sites at the primary DNA sequence level and ‘hotspots’ and ‘cold regions’ on a genome-wide scale. Sequences responsible for target site selection of the bacterial Tn10 transposon and retroviruses have been mapped to the core catalytic domain of the transposase or integrase, respectively (120, 121). However, despite the implication that the conserved catalytic DDE domain is responsible for locating the target site, no common pattern of integration emerges on the primary DNA sequence level. SB displays considerable specificity in target site selection at the primary DNA sequence level: in addition to the highly preferred TA.

**FIGURE 10** The SB transposase modulates cell-cycle progression through interaction with Miz-1. The SB transposase, through its interaction with Miz-1, downregulates cyclin D1 expression, which results in an inhibition of the G1/S transition of the cell-cycle. Reprinted from PNAS (103) with permission from the publisher. doi:10.1128/microbiolspec.MDNA3-0042-2014.f10

ASMscience.org/MicrobiolSpectrum
a palindromic AT-repeat consensus sequence in an AT-rich sequence context [Fig. 11(A)] with bendability and hydrogen bonding potential was found to constitute preferred target sites (122). It was shown that a characteristic deformation of the DNA sequence may be a recognition signal for target selection (118). This

FIGURE 11  Genomic insertion preferences of SB. (A) Consensus sequence of SB insertion sites. Seqlogo analysis and nucleotide probability plot of SB insertion sites in HeLa cells. Twenty base pairs upstream and downstream of the TA target sites were analyzed. The y-axis represents the strength of the information, with 2 bits being the maximum for a DNA sequence. (B) Relative frequencies of insertions into genes by retroviruses and transposons. The top portions of the graphs indicate an over-representation of genic insertions as compared to random. Part (A) reprinted from Molecular Therapy (126) with permission from the publisher. Part (B) reprinted from BioEssays (170) with permission from the publisher. doi:10.1128/microbiolspec.MDNA3-0042-2014.f11
deformation, and the likelihood a particular TA will be targeted by SB, can be computationally predicted (118), which may allow a theoretical assessment of the likelihood of transposon insertions in particular genomic regions (123). These results indicate that a combination of particular physical properties generates a spatial optimum of the DNA for transposase interaction. This pattern of structural preference is conserved in the Tc1/mariner family and in other relatively randomly integrating transposons in the DDE recombinase family (122). 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 (124). 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 to efficiently interact with such sequences specifies a subset within these sites where insertions occur.

In contrast to the considerable specificity at the primary DNA sequence level, SB integration can be considered fairly random on the genomic level (114, 115, 116, 117, 119, 122, 125). Roughly one-third of SB insertions in mouse and human cells occur in transcribed regions [Fig. 11(B)], and because genes cover about one-third of the genome, such frequency suggests neither preference for nor disfavoring of insertion into genes. SB shows no pronounced preference for inserting into transcription units or transcriptional regulatory regions of genes, and the vast majority of those insertions that occur in genes are located in introns (114, 115, 116, 117, 119, 122, 125). The transcriptional status of targeted genes apparently does not influence the integration profile of SB (125). This is in marked contrast to target site distributions of several other transposons including Tol2 [Fig. 11(B)] (117, 126), TcBuster (115), SPIN (115), and piggyBac [Fig. 11(B)] (115, 117, 126, 127, 128). These all show significant difference from random insertion with respect to favored integration into genes and near chromatin marks characteristic of active transcription units (e.g., H3K27 acetylation and H3K4 monomethylation) and disfavored integration near marks characteristic of inactive chromatin (e.g., H3K27 trimethylation).

The random genomic distribution of de novo SB insertions can be observed when the transposon DNA is introduced into the nucleus by extrachromosomal gene delivery, including plasmid vectors (114, 115, 116, 117, 119, 122, 125), integration-deficient lentiviral vectors (IDLVs) (114), adenovirus vectors (129), herpesvirus vectors (130), and adeno-associated vectors (131). In these cases, transposition takes place from the extrachromosomal vector into the genome. However, target site selection properties of SB when launched from a chromosomal site are markedly different and are governed by “local hopping”. Local hopping describes a phenomenon of chromosomal transposition in which transposons have a preference for landing into cis-linked sites in the vicinity of the donor locus. Local hopping seems to be a shared feature of “cut-and-paste” transposons. However, the actual extent of hopping to linked chromosomal sites and the interval of local transposition varies. For example, the P-element transposon of Drosophila prefers to insert within ~100 kb of the donor site at a rate ~50-fold higher than in regions outside that interval (132). Similarly, 30 to 80% of SB transposition events were found to re-insert locally on either side of the transposon donor locus (93, 133, 134, 135, 136, 137, 138, 139). In contrast to the P-element, SB seems to have a much larger local transposition interval (in the megabase range), but the targeted window for local reinsertion appears to be dependent on the donor locus and can range from <4 Mb (140) to 5 to 15 Mb (134). Local hopping of SB was also observed within 3 Mb of the donor locus in Xenopus (141).

The local hopping feature not only differs between different transposons, but a given transposon may show great variations in local hops in different hosts and in different donor loci even in the same host. For example, local hopping of the Ac element in tomato seems overall to be less prevalent than in maize (142, 143), and there are species-specific differences in its tendency for local hopping out of different transposon donor loci (144). This variation in local hopping of the same element could possibly be explained by varying affinity of the transposase for unknown, chromatin-associated factors in different hosts (145).

Self-disruptive autointegration of Sleeping Beauty

In the process of productive transposition, the excised DNA molecule integrates into a new genomic location. However, in principle, the excised transposon molecule could reinsert, in a self-disruptive process, into its own genome. This suicidal transposition event is called auto-integration, which has been observed with the bacterial systems Tn10 (146) and Mu (147), with Ty1 retrotransposons in yeast (148), and with retroviruses (149). Apparently, the SB transposon is also subject to auto-integration (56), thereby compromising integration of a fraction of excised transposon molecules.
The conserved size of Tc1-like elements is between 1.6 and 1.8 kb. The efficacy of transposition usually correlates negatively with increasing size of the transposon (58, 150, 151, 152), and such an effect was also observed with the bacterial transposons IS1 (153) and Tn10 (154) and with piggyBac (56). Since larger transposons contain more potential target sites, they could be particularly attractive targets for autointegration. Indeed, increasing size was found to sensitize both SB and piggyBac transposition for autointegration (56). However, the competition between autointegration and productive transposition is unlikely to be the only factor responsible for sensitivity to size as transposon excision, a step prior to integration, is already affected by the size of the SB transposon (56).

A host-encoded protein, barrier-to-autointegration factor (BAF or BANF1) has been identified by its ability to protect retroviruses from autointegration (155, 156, 157, 158, 159). Intriguingly, BANF1 also inhibited transposon autointegration of SB and was detected in higher-order protein complexes containing the transposase in human cells (56). Thus, the SB transposon seems to be able to recruit phylogenetically conserved cellular factors such as BANF1 that protects against self-disruption in a new environment (a human cell is a naïve host for SB). In fact, BANF1 might be an ideal cellular factor for

**FIGURE 12** Broad applicability of SB transposon-based gene vectors in vertebrate genetics. Reprinted from Mobile DNA (168) with permission from the publisher. doi:10.1128/microbiolspec.MDNA3-0042-2014.f12

- **Cell culture**

- **Transgenesis**

- **Functional genomics**

- **Gene therapy**

    **Sleeping Beauty Vector System Moves Toward Human Trials in the United States**
integrating elements in higher eukaryotes. As in retroviral integration (155, 156, 157, 158), BANF1 may compact the transposon genome to be a less accessible target for autointegration, thereby promoting the chromosomal integration step.

Sleeping Beauty AS A GENETIC TOOL

Transposons can be viewed as natural DNA transfer vehicles that, similar to integrating viruses, are capable of efficient genomic insertion. Transposition can be controlled by conditionally providing the transposase component of the transposition reaction. Thus, a DNA of interest (be it a fluorescent marker, an shRNA expression cassette, a mutagenic gene trap, or a therapeutic gene construct) cloned between the IR sequences of a transposon-based vector can be utilized for stable genomic insertion in a regulated and highly efficient manner. This methodological paradigm opened up a number of avenues for genome manipulations in vertebrates, including transgenesis for the generation of transgenic cells in tissue culture, the production of germline-transgenic animals for basic and applied research, forward genetic screens for functional gene annotation in model species, and therapy of genetic disorders in humans (Fig. 12). SB was the first transposon ever shown to be capable of gene transfer in vertebrate cells, and recent results confirm that SB supports a full spectrum of genetic engineering including transgenesis, insertional mutagenesis, and therapeutic somatic gene transfer, both ex vivo and in vivo. The first clinical application of the SB system will help to validate both the safety and efficacy of this approach. Applications of the SB system fall outside the scope of this chapter, and readers are referred to recent review articles (160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172).

ACKNOWLEDGMENTS

The authors thank past and present members of the “Transposition and Genome Engineering” group at the Paul Ehrlich Institute and the “Mobile DNA” group at the Max Delbrück Center for Molecular Medicine for their dedicated work and contribution. We are grateful to Irina Nesmelova for providing a color version of Fig. 3.

REFERENCES


Sleeping Beauty Transposition


20.

ASMscience.org/MicrobiolSpectrum


