OPEN ACCESS

Repository of the Max Delbrück Center for Molecular Medicine (MDC)
Berlin (Germany)
http://edoc.mdc-berlin.de/8811/

Targeted *Sleeping Beauty* transposition in human cells

Zoltán Ivics, Andrea Katzer, Eva E. Stüwe, Dora Fiedler, Siegne Knespel, and Zsuzsanna Izsvák

Published in final edited form as:
Molecular Therapy. 2007 Jun ; 15(6): 1137–1144 | doi: 10.1038/sj.mt.6300169
Nature Publishing Group (U.K.) ►

OPEN ACCESS

Final Draft

Targeted *Sleeping Beauty* transposition in human cells

Zoltán Ivics[1], Andrea Katzer[1], Eva E. Stüwe[1], Dora Fiedler[1], Siegne Knespel[1], and Zsuzsanna Izsvák[1,2]

[1] Max Delbrück Center for Molecular Medicine, Berlin, Germany
[2] Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences, Szeged, Hungary

ABSTRACT | Transposons are natural gene delivery vehicles. The *Sleeping Beauty* (SB) transposon shows efficient transposition and long-term transgene expression in the cells of vertebrates including humans. SB transposition into chromosomal DNA occurs in a fairly random manner. This is clearly not desirable in human gene therapeutic applications because there are potential genotoxic effects associated with transposon integration. In this study we set out to manipulate the selection of SB's target sites for targeted transposition into predetermined chromosomal regions. We evaluated experimental strategies based on engineered proteins composed of DNA-binding domains fused to (i) the transposase; (ii) another protein that binds to a specific DNA sequence within the transposable element; and (iii) another protein that interacts with the transposase. We demonstrated targeted transposition into endogenous matrix attachment regions (MARs) and a chromosomally integrated tetracycline response element (TRE) 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 an approximately 10^7-fold enrichment of transgene insertion at a desired locus. Our experiments provide proof-of-principle for targeted chromosomal transposition of an otherwise randomly integrating transposon. Targeted transposition can be a powerful technology for safe transgene integration in human therapeutic applications.

Introduction

DNA transposons have been developed as gene transfer vectors in invertebrate model organisms and, more recently, in vertebrates (reviewed in ref. 1). The *Sleeping Beauty* (SB) transposable element is a valuable tool for functional genomics in several model organisms [1], and shows promise for human gene therapeutic applications (see ref. 2 for a recent review). SB is a non-viral, integrating vector system that can provide long-term gene expression *in vivo*, coupled with a favorable safety profile as compared to widely used viral approaches (reviewed in ref. 2). SB is a two-component gene transfer vector system; it consists of a gene-of-interest cloned between the terminal inverted repeats (IRs) of SB that contain binding sites for the transposase [3] and the transposase polypeptide, the enzymatic factor of transposition.

As is the case with most other transposable elements, SB displays a certain degree of specificity in target-site utilization [4]. That is, SB exclusively integrates into TA dinucleotides, which are duplicated upon transposition and flank the integrated element. This is referred to as target-site duplication [3,4]. Even though SB transposition shows considerable specificity at the actual DNA sequence level, it can be considered fairly random on a genomic level [4]. Random genomic integration may limit transgene expression through silencing caused by position effects imposed by DNA methylation and/or the formation of a local heterochromatin structure. Although several studies have established that SB-mediated transposition can provide long-term expression *in vivo*, it is not currently known whether a fraction of the integrated transposons are subject to silencing. Random transposon integration also poses the potential risk of insertional activation or inactivation of cellular genes [5]. Even though SB-based vectors appear to be safer than retrovirus vectors [6], the integrating transposons are nevertheless potential mutagens. Thus, the ability to achieve directed integration into a specific chromosomal region will have a major positive impact on both the efficacy and the safety of integrating vectors in genetic engineering and therapy [7,8].

The process of integration-site selection of retroviruses and most transposons is primarily determined by the transposase or integrase (IN) itself, and regions within the catalytic core DDE domains are involved in mediating interactions with the target DNA [9,10]. Consequently, manipulation of target-site selection of retroviruses has been approached mainly by fusing IN to specific DNA-binding proteins including the phage λ repressor [11] and the *Escherichia coli* LexA protein [12,13]. Such engineered IN proteins have shown biased integration patterns near targeted sequences *in vitro*, but these methods have proven difficult to adapt to cellular applications [14]. Neither the LexA operator nor the λ operator sequences represent physiological targets in human DNA. One class of DNA-binding proteins that offers both specificity and flexibility in site-targeting(conferring site specificity) is the synthetic zinc finger (ZF) proteins [15]. Fusion proteins consisting of human immunodeficiency virus type-1 IN and E2C, a synthetic ZF protein recognizing a unique, 18-base pair (bp) sequence in the promoter region of the human *erb-B2* gene [16], were found to bias retroviral integration near the E2C-binding site *in vitro* [17]. Successful experimental retargeting of the yeast retrotransposon Ty5 suggests an alternative approach to controlling transposon integration, wherein transposase/IN proteins are modified to carry small peptide motifs or domains that recognize proteins bound to chromosomal target sites [18].

We set out to evaluate molecular strategies for targeted transposition of the SB system. Three distinct strategies were evaluated, all designed to direct the transpositional complex to a given chromosomal region by specific DNA–protein interactions. An N-terminal fusion of the artificial Jazz ZF protein recognizing a 9-bp site in the promoter region of the human utrophin promoter [19] retained some transpositional activity, but targeted transposition by this fusion protein was not seen. We demonstrated targeted transposition into endogenous chromosomal matrix attachment regions (MARs) and a chromosomally integrated tetracycline response element (TRE) in cultured human cells, using targeting proteins that bind to the transposon DNA. As high as >10% of transposant cells were found to contain targeted insertion events by

OPEN ACCESS

the use of an approach based on protein–protein interactions between the SB transposase and an engineered targeting protein consisting of the N-terminal protein interaction domain of SB and the tetracycline repressor (TetR). We discuss potential applications and future development of targeted DNA transposition in molecular genetics and therapy.

Results

Transpositional activities of transposase fusions

Because the transposon system consists of two main functional components, namely, the transposon DNA and the transposase protein, the tethering of 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 (i) the transposase (Fig.1a), or (ii) a protein domain that binds to a site within the transposable element (Fig.1b), or (iii) a protein domain that makes contacts with the transposase through protein–protein interactions (Fig.1c).

Fusion proteins were engineered, containing the SB transposase and either the bacterial TetR that specifically binds the tetracycline operator sequence or the Jazz and E2C ZF peptides (Fig.2a). Transposon excision activity of the fusion proteins was tested using a polymerase chain reaction (PCR)-based excision assay [20]. Out of the four constructs tested, only the Jazz/SB fusion showed detectable activity in human HeLa cells, though at a clearly reduced efficiency compared to unfused transposase (Fig.2b, compare lanes 2 and 4). In line with the excision data, the Jazz/SB fusion was found to retain transpositional activity at approximately 10–15% of the wild-type level (Fig.2c). 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 utrophin locus, and no occurrence of the 9-bp binding site of Jazz within a 1-kilobase 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.

Targeted SB transposition into chromosomal MARs in cultured human cells

In view of 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.1b). We began our experiments with targeting into chromosomal regions that are represented many times in a complex genome, whose base composition is AT-rich, and for which interacting protein determinants are known and well characterized, so that SB can efficiently integrate into them. Scaffold attachment regions, also called 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 [21].

A fusion protein in which the SAF-box is fused to the C-terminus of the *E. coli* LexA protein was engineered (Fig.3a). 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 an SB transposon vector (pTzeo-322/LexOP in Fig.3a). Importantly, the Tzeo-322/LexOP transposon (either with or without the LexA protein) jumped as efficiently as the parental Tzeo-322 transposon lacking the LexA binding site (data not shown), thereby indicating that binding of the fusion protein to the transposon does not interfere with transposition.

We hypothesized that targeted transposition could occur by simultaneous interaction of LexA/NLS/SAF with both the transposon vector and chromosomal MARs, 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 analyzed 56 unique SB insertions recovered from cells expressing LexA/NLS/SAF and 57 from cells expressing LexA/NLS only. The results from each data set were categorized into six groups according to proximity to a predicted MAR. 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-kilobase distance from a predicted MAR, versus two events in the control group (Fig.3b). An output of the MAR-Wiz program showing the positions of the MAR peak and that of a transposon insertion is shown in **Supplementary Figure S1**. Some of the cloned chromosomal DNA fragments flanking the transposon insertions and predicted to contain MAR sequences were tested in an *in vitro* assay for binding by the SAF-box. As shown in **Figure 3c**, 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.

Targeted transposition into a specific locus in human cells

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 on the TetR-operator system [22]. Unlike the binding of the SAF-box to scaffold attachment regions, the interaction of TetR with its operator site is highly sequence-specific. We generated a transgenic HeLa-derived cell line containing a TRE (encompassing seven units of the tetracycline operator)-driven enhanced green fluorescent protein (EGFP) gene, which served as a chromosomal target for transposon integration (**Fig.4a**). The targeting fusion protein consisted of TetR, an NLS, and the LexA protein (TetR/NLS/LexA in **Fig.4a**). The ability of TetR/NLS/LexA

OPEN ACCESS

to bind to the TRE and the LexA operator-containing transposon DNA was tested in an electrophoretic mobility shift experiment using nuclear extracts of HeLa cells transfected with the fusion protein expression construct and radioctively labeled TRE and transposon probes (**Fig.4b**). Both the TRE and the transposon probes were shifted with TetR/NLS/LexA (**Fig.4b**, lanes 2 and 5), but not with a nuclear extract prepared from untransfected cells (**Fig.4b**, lanes 3 and 6).

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, and genomic DNA was prepared and subjected to PCR analysis with primers designed to amplify transposition events upstream of the EGFP gene (**Fig.4a**). PCR products consistent with transposon integration in both orientations with respect to the EGFP gene were obtained from transfections with pTzeo-322/LexOP but not with pTzeo-322, a transposon vector without a LexA binding site (**Fig.4c**), thereby 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 bp downstream of the TRE region (Sites 1 and 2 in **Fig.4d**).

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.

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 fragment of the SB transposase to encode a transposase interaction function [23]. 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 [23]. 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 [23]. Therefore, we chose to pursue a strategy to generate fusion proteins of DNA-binding proteins with N-57, in the hope that these fusions would retain their ability 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 employed, except that the targeting fusion protein consisted of TetR and N-57 (**Fig.5a**). 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 were 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.5b**). It therefore follows that 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 170 bp (Site 3) and 423 bp (Site 4) upstream of the TRE region (**Fig.5b** and **d**). 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.5c**, transposon insertions into the TRE region can be recovered only in the absence of doxycycline. Collectively, the results presented in **Fig.5b** and **c** 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 6 independent transfections, from which 3 sites, Site 1, Site 4 and Site 5, were hit multiple times independently (**Fig.5d**).

How efficiently 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 twenty-four clones contained targeted transposition events into Site 1 (**Fig.5e**, upper gel). In an independent experiment, 4 out of 48 individual transgenic cell clones were found to have targeted transposition events (**Fig.5e**, lower gel). That is, on an 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 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).

Discussion

Here we demonstrate targeted chromosomal insertion of the SB 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 fused to (i) the SB transposase itself, or (ii) a protein domain that binds the transposon DNA, or (iii) a protein domain that interacts with the SB transposase (**Fig.1**). Fusions of the bacterial IS30 transposase with the λ repressor and with the DNA-binding domain of the transcription factor Gli1, using plasmid targets, showed altered insertion profiles in *E. coli* and in zebrafish embryos, respectively [24]. 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 [25]. We have assessed the feasibility of fusing DNA-binding

OPEN ACCESS

proteins directly to the SB transposase polypeptide (**Fig.2a**). Out of four fusions, only one, namely, the artificial three-finger protein Jazz added to the N-terminus of SB, showed detectable transposition activity (**Fig.2b** and **c**). 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. It is possible that specificity for a 9-bp DNA sequence of Jazz is far too low for targeted transposition. Future efforts to improve this technology will therefore have to focus on the identification of highly specific DNA-binding domains, and on a systematic evaluation of protein spacer sequences linking the two fusion partners in order to allow rational designing 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 containing binding sites of the respective fusion partners, thereby bringing them into close proximity (**Fig.1b**). 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 [26,27]. This phenomenon, called transposon "homing", tends to be region-specific [27] with transposon integrations distributed over several kilobase pair regions near the targeted loci.

Potential SB targeting by such a mechanism was assessed by engineering a LexA operator into a benign site within an SB transposon vector (**Fig.3a**). Targeted transposition events into endogenous chromosomal MAR sequences as well as a chromosomally integrated TRE were recovered by employing targeting fusion proteins containing LexA and either the SAF-box, a protein domain that binds to chromosomal MARs, or the TetR (**Fig.3** and **4**). 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; however, 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.1c**), was also evaluated. Several, naturally occurring transposable elements have evolved strategies for targeted insertion into defined chromosomal sites or regions, and the mechanisms of targeted insertions often rely on protein–protein interactions between a transposon-encoded factor and a cellular, DNA-binding host factor. For example, the *E. coli* transposon *Tn7* integrates into the chromosomal site attTn7. This targeted insertion is mediated through sequence-specific DNA binding of the transposase subunit called TnsD at attTn7, and interaction with another subunit called TnsC [28]. There are several retrotransposable elements in yeast that have evolved strategies for targeted integration. For example, the integration of Ty1 and Ty3 occurs upstream of genes transcribed by RNA polymerase III, and this process is mediated through interaction of IN with components of transcription factor III [29,30]. The Ty5 element integrates into heterochromatin, mediated by interaction of the IN with the heterochromatin protein Sir4p [31]. Finally, human immunodeficiency virus integration is particularly favored in active transcription units, and the LEDGF/p75 transcription factor was implicated in directing human immunodeficiency virus integration-site selection via a tethering interaction with IN [32]. A recent study showed increased integration by human immunodeficiency virus IN near λ repressor binding sites in the presence of λ repressor-LEDGF/p75 fusions *in vitro* [33]. Altogether, these observations suggest a general model wherein interactions between transposase/IN and DNA-bound proteins mediate insertional target choice.

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.5**). 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 bp downstream of TRE was identified, targeted insertion events occurred within a 2.6-kilobase window around the targeted TRE (**Fig.5d**). 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 [11,13]. 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, and these must 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 at which 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.5d**), 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 this study that, in addition to the targeted insertion events, transposition also occurs into numerous non-targeted sites. Since the efficiency of targeted insertion of naturally occurring transposons can approach 100%, the question arises as to how to improve the success rate of SB targeting. A yet-to-be-explored possibility for increasing targeting efficiencies is to use varying ratios of targeting protein to transposase. A further factor to consider 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. Non-specific binding of the transposase to human chromosomal DNA likely competes with specific binding to a desired targeted sequence; this would limit the probability of a targeted transposition event, and consequently the frequency of occurrence. Such limitation could, in principle, be circumvented by interfering with the

OPEN ACCESS

target DNA binding function of the transposase. It is not immediately evident, and remains to be seen, 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.5e**). Because the estimated theoretical frequency of hitting any TA in the human genome by random transposition is approximately 1 in a total of 10^8 transposition events, this technology represents enrichment of transgene insertion by a factor of approximately 10^7 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 be lower 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, thereby suggesting that targeting efficiency is determined by the amount of targeting protein tethered to the target DNA [18]. Technologies for site-directed transgene integration could potentially have practical relevance in at least three areas of molecular genetics and therapy. First, targeted transposition could provide the means for target-selected chromosomal engineering in experimental model systems, for which 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 the 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 in 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*.

Materials and Methods

Plasmids

The LexA gene was PCR-amplified from plasmid pEG202, and inserted into the pFV4a expression vector containing the carp β-actin promoter [34]. A double-stranded oligo encoding the simian virus 40 T antigen NLS was inserted downstream of LexA, followed by an in-frame insertion of 44 amino acids (excluding the initiator methionine) of the SAF-box. The sequence containing the LexA operator site (**Fig.3a**) was inserted into pT/zeo-322 [35] using PCR mutagenesis. The TetR/NLS/LexA and TetR/NLS/N-57 fusions were constructed by inserting LexA or 56 amino acids (excluding the initiator methionine) of the SB transposase into the TetR expression vector pUHD141-1. The TetR/SB fusion was constructed by replacing a restriction fragment of FV4a-SB with a corresponding fragment of TetR/NLS/N-57. In both of these fusions, the TetR domain and the transposase domain are connected by an NLS/glycine-bridge linker with the following amino acid composition: PKKKRKLAGGGGGGGGG. The SB/E2C fusion was made by inserting the SB transposase gene together with an E2C gene fragment into FV4a, whereas the E2C/SB and Jazz/SB fusion were constructed by inserting E2C or Jazz in place of TetR in FV4a-TetR/SB. In these three constructs, the ZF domains and the transposase domain are connected by a linker consisting of 10 glycine residues. Primer sequences for all PCRs and additional cloning details are provided in **Supplementary Materials and Methods**.

Cell culture and transfections

Transposition assays in HeLa cells were carried out as described [3]. Typically, 10^5 cells were transfected with 90 ng of each the transposon donor plasmid pT/neo or pT/zeo-322 or pTzeo-322/LexOP, the transposase-expressing vector pCMV-SB10 [3], and a plasmid expressing a targeting fusion protein, using Fugene6 transfection reagent (Roche, Basel, Switzerland). Selection with doxycycline was done at 1 µg/ml as recommended by the manufacturer (Clontech, Mountain View, CA). The transgenic HeLa-derived cell line was generated by cotransfection of pTRE-d2EGFP (Clontech, Mountain View, CA) and a simian virus 40-Hygro selectable marker, and selection with hygromycin. One resulting clone that responded well to transactivation by the pTet-Off plasmid (Clontech) was retained for all further experiments. The 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 2 days post-transfection directly into 96-well tissue culture plates such that all wells contained at least one transgenic cell. This ensures that the chance for subclonal propagation of transgenic clones within a transfected cell population is kept at a minimum.

Mapping of transposon insertion sites

Cloning of transposon insertion sites from human genomic DNA was done using splinkerette PCR or transposon rescue as described [3,4]. 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 the 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.

Electrophoretic mobility shift assay

The probes used were an *Eco*RI fragment of pTzeo-322/LexOP containing the LexA operator site within the left inverted repeat of SB and an *Xho*I/*Sac*II fragment of pTRE-d2EGFP containing the TRE region, in assay conditions described in [3]. Reactions contained labeled probe, 1 µg poly[dI][dC], 100 pg labeled fragment and 1 µl nuclear protein extract.

Pull-down SAF-binding assay

A DNA-binding assay was done essentially as described [21]. 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). Radiolabeled 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 unlabeled unspecific competitor in 1,000-fold excess. Unbound DNA was removed by washing the complex extensively with the binding buffer. DNA binding was quantified by scintillation counting.

Acknowledgments

Materials for SAF-binding were kindly provided by Frank O. Fackelmeyer (Heinrich Pette Institute, Hamburg, Germany). The E2C ZF gene was obtained from Carlos F. Barbas (Scripps Research Institute, La Jolla, CA), and the Jazz ZF gene was provided by Claudio Passananti (Institute of Molecular Biology and Pathology, Rome, Italy). The LexA gene was kindly provided by Roger Brent (Molecular Sciences Institute, Berkeley, CA). We are grateful to Manfred Gossen (Max Delbrück Center, Berlin, Germany) for Tet-system vectors and for valuable discussions. Work in the authors' laboratory is supported by European Union FP6 grant INTHER LSHB-CT-2005018961, and a grant from the Deutsche Forschungsgemeinschaft SPP1230 "Mechanisms of gene vector entry and persistence".

Corresponding Author

Zoltán Ivics, Max Delbruck Center for Molecular Medicine, Robert Rössle Strasse 10, D-13092 Berlin, Germany, E-mail: zivics@mdc-berlin.de

References

1. Miskey, C, Izsvák, Z, Kawakami K and Ivics, Z (2005). DNA transposons in vertebrate functional genomics. Cell Mol Life Sci 62: 629–641.
2. Ivics, Z and Izsvák, Z (2006). Transposons for gene therapy!Curr Gene Ther 6: 593–607.
3. Ivics, Z, Hackett, PB, Plasterk, RH and Izsvák, Z (1997). Molecular reconstruction of sleeping beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell 91: 501–510.
4. Vigdal, TJ, Kaufman, CD, Izsvák, Z, Voytas DF and, 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.
5. Baum, C, von Kalle, C, Staal, FJ, Li, Z, Fehse, B, Schmidt, M *et al*. (2004). Chance or necessity? Insertional mutagenesis in gene therapy and its consequences. Mol Ther 9: 5–13.
6. Yant, SR, Wu, X, Huang, Y, Garrison, B, Burgess, SM and Kay, MA (2005). High-resolution genome-wide mapping of transposon integration in mammals. Mol Cell Biol 25: 2085–2094.
7. Glover, DJ, Lipps, HJ and Jans, DA (2005). Towards safe, non-viral therapeutic gene expression in humans. Nat Rev Genet 6: 299–310.
8. Bushman, F (2002). Targeting retroviral integration? Mol Ther 6: 570–571.
9. Harper, AL, Sudol, M and Katzman, M (2003). An amino acid in the central catalytic domain of three retroviral integrases that affects target site selection in nonviral DNA. J Virol 77: 3838–3845.
10. Appa, RS, Shin, CG, Lee, P and Chow, SA (2001). Role of the nonspecific DNA-binding region and alpha helices within the core domain of retroviral integrase in selecting target DNA sites for integration. J Biol Chem 276: 45848–45855.
11. Bushman, FD (1994). Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. Proc Natl Acad Sci USA 91: 9233–9237.
12. Katz, RA, Merkel, G and 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: 178–190.
13. Goulaouic, H and 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.
14. Bushman, FD (2003). Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. Cell 115: 135–138.
15. Beerli, RR and Barbas, CF 3rd (2002). Engineering polydactyl zinc-finger transcription factors. Nat Biotechnol 20: 135–141.
16. Beerli, RR, Segal, DJ, Dreier, B and 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 USA 95: 14628–14633.
17. Tan W, Zhu K, Segal DJ, Barbas CF 3rd and 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: 1301–1313.
18. Zhu, Y, Dai, J, Fuerst, PG and Voytas, DF (2003). Controlling integration specificity of a yeast retrotransposon. Proc Natl Acad Sci USA 100: 5891–5895.
19. Corbi, N, Libri, V, Fanciulli, M, Tinsley, JM, Davies, KE and Passananti, C (2000). The artificial zinc finger coding gene 'Jazz' binds the utrophin promoter and activates transcription. Gene Ther 7: 1076–1083.
20. Izsvák, Z, Stüwe, EE, Fiedler, D, Katzer, A, Jeggo, PA and 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.
21. Kipp M, Gohring F, Ostendorp T, van Drunen CM, van Driel R, Przybylski M *et al*. (2000). SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA. Mol Cell Biol 20: 7480–7489.
22. Gossen, M, Freundlieb, S, Bender, G, Muller, G, Hillen, W and Bujard, H (1995). Transcriptional activation by tetracyclines in mammalian cells. Science 268: 1766–1769.
23. Izsvák, Z, Khare, D, Behlke, J, Heinemann, U, Plasterk, RH and 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.
24. Szabo, M, Muller, F, Kiss, J, Balduf, C, Strahle, U and Olasz, F (2003). Transposition and targeting of the prokaryotic mobile element IS30 in zebrafish. FEBS Lett 550: 46–50.
25. Maragathavally, KJ, Kaminski, JM and Coates CJ (2006). Chimeric Mos1 and piggyBac transposases result in site-directed integration. FASEB J 20: 1880–1882.
26. Hama, C, Ali, Z and Kornberg, TB (1990). Region-specific recombination and expression are directed by portions of

the Drosophila engrailed promoter. Genes Dev 4: 1079–1093.

27. Kassis, JA, Noll, E, VanSickle, EP, Odenwald, WF and Perrimon, N (1992). Altering the insertional specificity of a Drosophila transposable element. Proc Natl Acad Sci USA 89: 1919–1923.

28. Bainton, RJ, Kubo, KM, Feng, JN and Craig, NL (1993). Tn7 transposition: target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified *in vitro* system. Cell 72: 931–943.

29. Kirchner, J, Connolly, CM and Sandmeyer, SB (1995). Requirement of RNA polymerase III transcription factors for *in vitro* position-specific integration of a retroviruslike element. Science 267: 1488–1491.

30. Bachman, N, Gelbart, ME, Tsukiyama, T and Boeke, JD (2005). TFIIIB subunit Bdp1p is required for periodic integration of the Ty1 retrotransposon and targeting of Isw2p to *S. cerevisiae* tDNAs. Genes Dev 19: 955–964.

31. Xie, W, Gai, X, Zhu, Y, Zappulla, DC, Sternglanz, R and Voytas, DF (2001). Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. Mol Cell Biol 21: 6606–6614.

32. Ciuffi, A, Llano, M, Poeschla, E, Hoffmann, C, Leipzig, J, Shinn, P *et al*. (2005). A role for LEDGF/p75 in targeting HIV DNA integration. Nat Med 11: 1287–1289.

33. Ciuffi, A, Diamond, TL, Hwang, Y, Marshall, HM and 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.

34. Liu, ZJ, Moav, B, Faras, AJ, Guise, KS, Kapuscinski, AR and Hackett, PB (1990). Development of expression vectors for transgenic fish. Biotechnology 8: 1268–1272.

35. Izsvák, Z, Ivics, Z and Plasterk, RH (2000). Sleeping beauty, a wide host-range transposon vector for genetic transformation in vertebrates. J Mol Biol 302: 93–102.

OPEN ACCESS

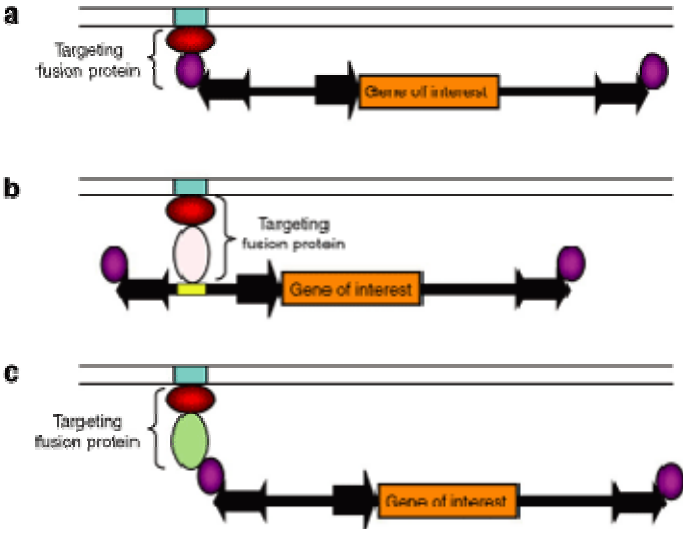

Figure 1. Experimental strategies for targeting *Sleeping Beauty* transposition. The common components of the targeting systems include a transposable element that contains the inverted repeats (IRs) (arrowheads) and a gene-of-interest equipped with a suitable promoter. The transposase (purple circle) 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.

OPEN ACCESS

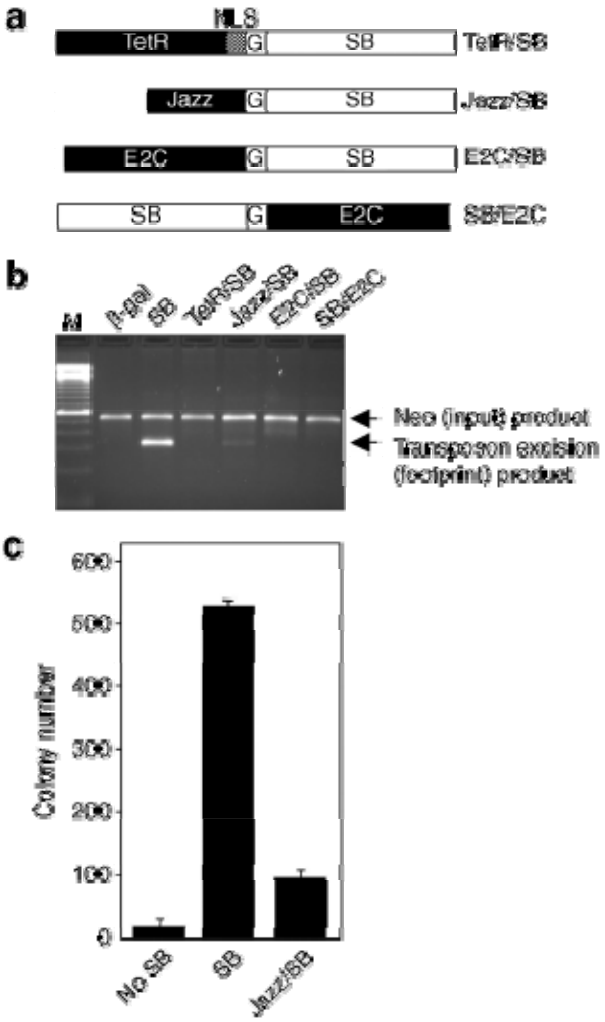

Figure 2. Design and transpositional activities of transposase fusions. (**a**) Schematic representation of the fusion proteins that consist of the Sleeping Beauty (SB) transposase fused to the tetracycline repressor (TetR), the Jazz or the E2C zinc finger (ZF) proteins. The TetR/SB fusion contains the simian virus 40 nuclear localization signal (NLS) in addition to the NLS naturally present in SB. All fusions contain a glycine-bridge (G) consisting of 10 consecutive glycine residues to provide a flexible linker between the fusion partners. (**b**) A fusion protein consisting of the SB transposase and the Jazz ZF 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 polymerase chain reaction (PCR) that amplifies a footprint product. PCR-amplification of the neo marker inside the transfected transposon donor serves as a loading control. (**c**) The Jazz/SB fusion protein is active in SB transposition. HeLa cells were cotransfected with a neo-marked transposon plasmid and vectors expressing β-galactosidase or SB transposase or Jazz/SB, and selected for transposition events by antibiotic selection with G-418. The numbers of G-418-resistant colonies are indicated.

OPEN ACCESS

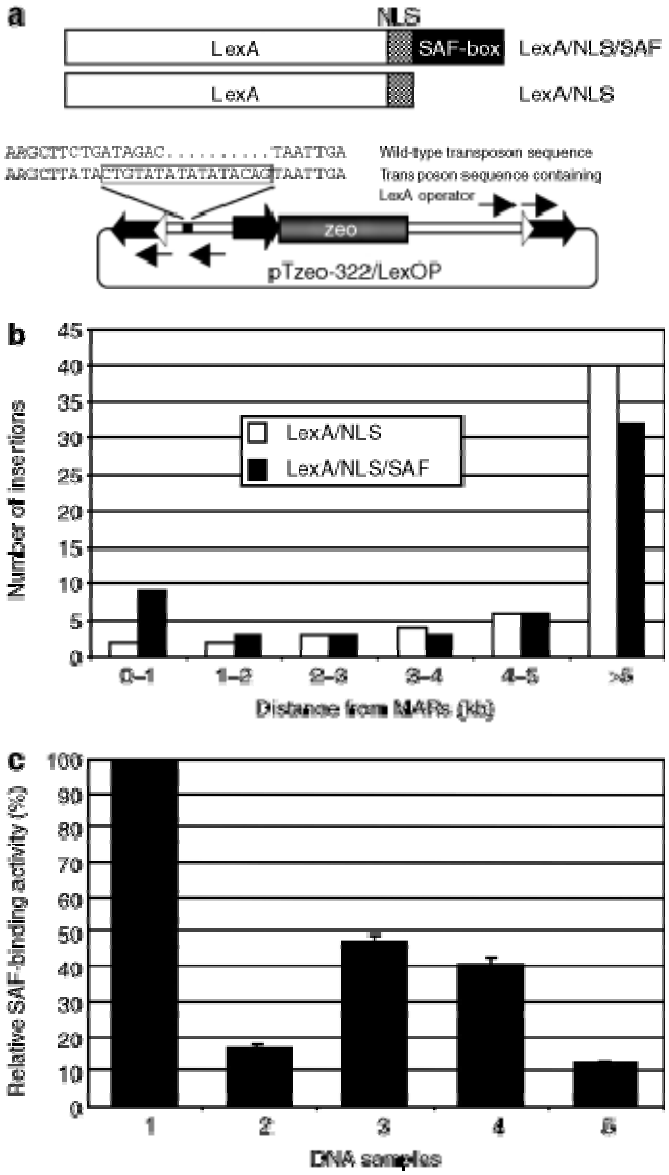

Figure 3. Transposon targeting into chromosomal matrix attachment regions (MARs). (**a**) Components of the targeting system. The targeting fusion protein consists of the bacterial LexA DNA-binding protein, a nuclear localization signal (NLS) and the scaffold attachment factor (SAF)-box. A control protein lacks the SAF-box. The transposable element contains a 16-base pair binding site for the LexA protein. Arrows indicate the approximate positions of nested polymerase chain reaction (PCR) primers that were used to identify targeted transposition events. (**b**) Transposon insertions were recovered from HeLa cells transfected with the components of the targeting system, and human chromosomal DNA flanking the insertion sites was analyzed with respect to proximity to chromosomal MARs. Distances were categorized, and the numbers of insertions obtained in the presence and absence of the targeting fusion protein in each category are shown. The two datasets show a statistically significant (P= 0.024) difference in the distribution of transposon insertions. (**c**) In vitro binding of Sleeping Beauty (SB) target sites by the SAF-box. Resin-bound SAF-box peptide was incubated with radiolabeled (1) MII DNA (positive control); (2) pUC19 plasmid DNA (negative control), and (3–5) transposon target sites, and the 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%.

OPEN ACCESS

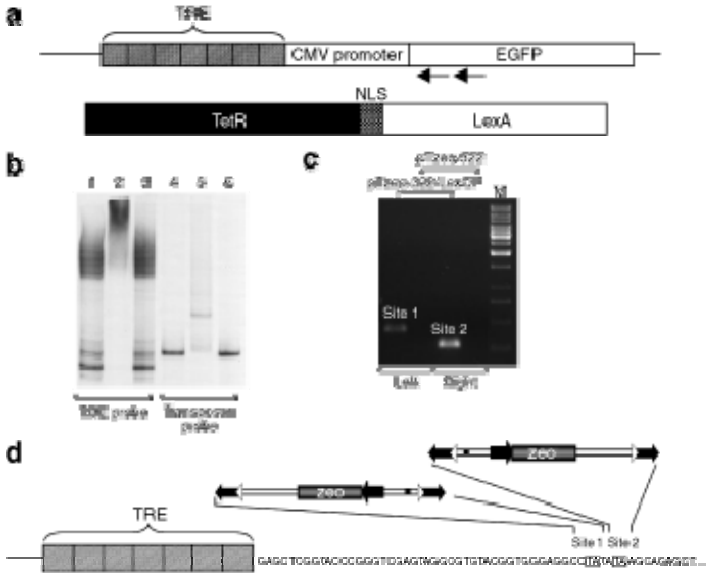

Figure 4. Transposon targeting into a chromosomally integrated tetracycline response element (TRE). (**a**) Components of the targeting system. The target DNA is a chromosomally integrated TRE upstream of the cytomegalovirus (CMV) minimal promoter and the enhanced green fluorescent protein (EGFP) gene in transgenic human HeLa cells. Arrows indicate the approximate positions of nested polymerase chain reaction (PCR) primers that were used for identifying 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 binding site engineered into a transposon vector, as shown in **Fig.3a**. (**b**) 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. (**c**) Cells were cotransfected with the components of the transposon targeting system, and genomic DNA of pooled transformant cells was subjected to PCR using primers specific to the transposable element (**Fig.3a**) and to the EGFP gene in the target DNA (**Fig.4a**). 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 inverted repeat of the transposon. M: size marker. (**d**) Targeted transposition close to the TRE. PCR products in **Fig.4c** 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).

OPEN ACCESS

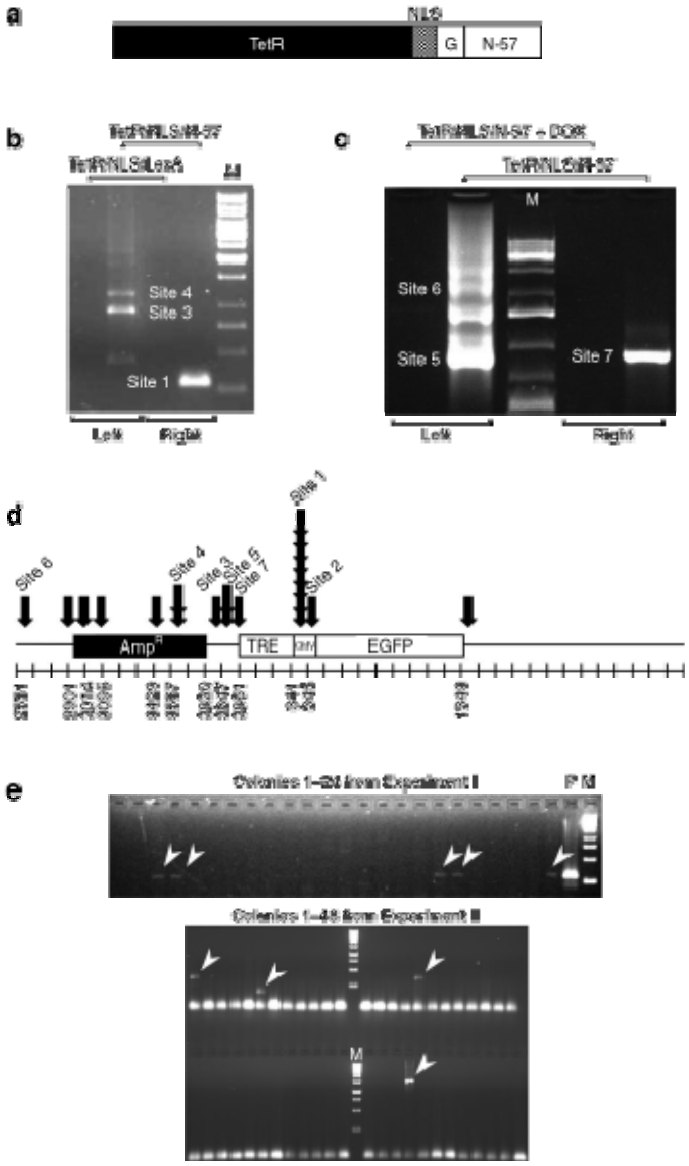

Figure 5. Transposon targeting using a strategy based on protein–protein interactions between a targeting fusion protein and the Sleeping Beauty (SB) transposase. (**a**) The targeting fusion protein consists of the tetracycline repressor (TetR) that binds to the tetracycline response element (TRE), a nuclear localization signal (NLS), a glycine-bridge (G) and the N-terminal protein interaction domain of the SB transposase (N-57). (**b**) Cells were cotransfected with the components of the transposon targeting system, and genomic DNA of pooled transformant cells was subjected to polymerase chain reaction (PCR) as described in **Fig.4c**. 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 inverted repeat (IR) of the transposon. M: size marker. (**c**) 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. (**d**) Mapping of targeted SB insertions (arrows) with respect to the TRE-EGFP target isolated from six independent experiments is shown. Multiple arrows represent independent insertions into the same site. Positions of the insertions are indicated below; the numbers correspond to the base pair numbering of the pTRE-d2EGFP plasmid (Clontech). (**e**) 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.