



Feng, X. and Bednarz, A.L. and Colloms, S.D. (2010) *Precise targeted integration by a chimaeric transposase zinc-finger fusion protein*. *Nucleic Acids Research*, 38 (4). pp. 1204-1216. ISSN 0305-1048

<http://eprints.gla.ac.uk/25684/>

Deposited on: 25 March 2010

Precise targeted integration by a chimaeric transposase zinc-finger fusion protein

Xiaofeng Feng, Amy L. Bednarz and Sean D. Colloms*

Faculty of Biomedical and Life Sciences, University of Glasgow, Bower Building, University Ave, Glasgow G12 8QQ, Scotland, UK

Received August 4, 2009; Revised October 29, 2009; Accepted October 30, 2009

ABSTRACT

Transposons of the Tc1/*mariner* family have been used to integrate foreign DNA stably into the genome of a large variety of different cell types and organisms. Integration is at TA dinucleotides located essentially at random throughout the genome, potentially leading to insertional mutagenesis, inappropriate activation of nearby genes, or poor expression of the transgene. Here, we show that fusion of the zinc-finger DNA-binding domain of Zif268 to the C-terminus of ISY100 transposase leads to highly specific integration into TA dinucleotides positioned 6–17 bp to one side of a Zif268 binding site. We show that the specificity of targeting can be changed using Zif268 variants that bind to sequences from the HIV-1 promoter, and demonstrate a bacterial genetic screen that can be used to select for increased levels of targeted transposition. A TA dinucleotide flanked by two Zif268 binding sites was efficiently targeted by our transposase-Zif268 fusion, suggesting the possibility of designer ‘Z-transposases’ that could deliver transgenic cargoes to chosen genomic locations.

INTRODUCTION

Various strategies have been used to integrate foreign DNA into the genomes of model and other organisms, including homologous recombination, site-specific recombination and double-strand break repair following site-specific DNA cleavage by engineered nucleases (1). These strategies integrate the foreign DNA into specific locations in the host genome, but can suffer from low efficiency or a lack of suitable integration sites. In contrast, transposon and retroviral integration systems act with high efficiency (2,3), but suffer from a

lack of control over the site of integration. Integration at inappropriate genomic locations can lead to activation or inactivation of important host genes with severe deleterious consequences (4). Alternatively, insertion into regions of heterochromatin can lead to poor expression of the foreign DNA.

Members of the Tc1/*mariner* family of transposable elements are widespread in nature, being found in both prokaryotes and eukaryotes (5). Naturally occurring active elements from the Tc1/*mariner* family include Mos1 *mariner* from *Drosophila mauritiana* (6), Tc1 and Tc3 from *Caenorhabditis elegans* (7,8), and ISY100 (ISTcSa) from *Synechocystis* sp. PCC6803 (9,10). Other active elements, such as *Sleeping Beauty* and *Himar1* have been reconstructed from the sequences of multiple inactive copies present in the genomes of fish and insects, respectively (11,12). Transposons from the Tc1/*mariner* family are highly active in cultured cells and model organisms, and there is much interest in their use as vectors for gene delivery (13).

One transposon from the Tc1/*mariner* family, the *Synechocystis* transposon ISY100, transposes efficiently both *in vitro*, and *in vivo* in *Escherichia coli* (9,10). ISY100 has a simple structure, consisting of a single transposase gene flanked by 24 bp imperfect terminal inverted repeats (IRs) that mark the transposon ends. Like other members of this family, ISY100 transposase has an N-terminal DNA-binding domain that specifically recognizes the transposon IRs, and a C-terminal DDE domain that catalyses the cutting and rejoining reactions involved in transposition. Transposition occurs by a cut and paste mechanism in which the transposon is cut from one site in the genome by double-strand breaks at both transposon ends, and then inserted into a new target site. Transposition is exclusively into TA dinucleotides, and insertions are flanked by duplications of this target TA.

Previous work has had some success in modifying the target specificity of retroviral integrases and eukaryotic transposases, by directly or indirectly tethering the transposase to specific sites in the DNA using

*To whom correspondence should be addressed. Tel: +0141 330 6236; Fax: +0141 330 4878; Email: s.colloms@bio.gla.ac.uk
Present addresses:

Xiaofeng Feng, National Institute of Biological Sciences, Beijing, China.

Amy L. Bednarz and Sean D. Colloms, Xermit, Colchester Drive, Glasgow, Scotland, UK.

sequence-specific DNA-binding domains (14). In most of these studies, multiple binding sites for the DNA-binding domain were required, and integration was observed in a relatively large stretch of DNA around the specific binding site. Here, we investigate the properties of a 'Z-transposase' made by fusing the zinc-finger DNA-binding domain from the mouse transcription factor Zif268 to the C-terminus of ISY100 transposase. We find that this fusion protein combines the TA target-specificity of the Tc1/*mariner* family of transposases with the DNA-binding specificity of Zif268 and promotes transposition in specific TAs located adjacent to a single Zif268 binding site. The target specificity of our Z-transposase can be changed by substituting the Zif268 DNA-binding with variants that specifically recognize sequences from the HIV-1 promoter. This system holds great promise for the development of gene transfer applications with precise control of the integration site.

MATERIALS AND METHODS

Donor and target plasmids

To make target plasmids for standard transposition assays, the sequences shown in Figure 1C, flanked by HindIII and EcoRI cohesive ends, were synthesized as oligonucleotides and inserted in the polylinker of the 5339 bp plasmid pH2 (10). Papillation target plasmids were based on the promoter probe vector pRS415 (15), which contains four copies of the *rrnB* transcription terminator upstream of an EcoRI–BamHI polylinker and the complete *lac* operon. The sequence between the BamHI site and codon 9 of *lacZ* was deleted from pRS415 by PCR, and papillation target sites with EcoRI and BamHI cohesive ends were inserted in the polylinker of the resulting vector.

Donor plasmids were based on pλISY100-kan-IRR30, which has 30 bp ISY100 IRR and IRL sequences flanking a kanamycin resistance determinant on the λ dv-based plasmid pCLIP18 (16). An active copy of the ISY100 transposase gene was inserted together with *lacI^f* and the *trc* promoter from pTrc99A (17) into pλISY100-kan-IRR30, upstream of the mini-transposon, as an SphI–XbaI fragment. To create the fusion proteins, an EagI site was added to the transposase gene, just before the stop codon. This adds three additional amino acids (YGR) to the C-terminus of transposase, with no detectable effect on transposition. Synthetic fragments encoding Zif268, or ZifA or ZifB variants of Zif268 (18), together with linker sequences L1, L2 or L3 were inserted between EagI and KpnI sites at the 3'-end of the transposase gene to produce the finished donor plasmids. To make the papillation donor plasmids, the mini-transposon was modified by the addition of the sequence 5'-gctagctcaacAGGAA acagctATG, containing the start codon and ribosome binding site from *lacZ* (uppercase) and an NheI restriction site (underlined) just upstream of IRR30. Further details and DNA sequences of all plasmids used in this study are available from SDC.

Transposition assays

To carry out *in vivo* transposition assays, target plasmids were introduced into the *recA*-deficient strain DH5α using standard CaCl₂-mediated transformation. The resulting strains were then transformed with donor plasmids using the same method. For standard transposition assays, transformants were selected on Luria–Bertani (LB) agar containing 50 µg/ml ampicillin, 25 µg/ml chloramphenicol, 25 µg/ml kanamycin, 50 µM ZnSO₄, and incubated for 60 h at 37°C. ~1000 transformant colonies were washed from the plate in LB broth, cells were harvested by centrifugation and plasmid DNA containing donor, target and transposition product, was purified using QIAprep spin miniprep kits (Qiagen). A total of 100 ng of this DNA was electroporated into DS964 (*recA*, λ lysogen) in which the donor plasmids cannot replicate. A small aliquot of the electroporation mixture was diluted 10⁴-fold, and spread on plates containing ampicillin to select for target plasmids. The rest was spread undiluted on plates containing kanamycin to select for transposition products. The transposition frequency was calculated by dividing the number of kanamycin resistant colonies per millilitre of transformation mix by the number of ampicillin-resistant colonies per millilitre.

Plasmid DNA was purified from individual kanamycin-resistant colonies and characterised by restriction digestion and DNA sequencing. To display the overall pattern of insertion sites, plasmid DNA was isolated from pools of ~100 kanamycin-resistant colonies, cut with BamHI and run on 1.2% agarose gels in 50 mM Tris–acetate 1 mM EDTA pH 8.2 running buffer. Gels were stained with ethidium bromide, DNA was visualised by 260 nm UV illumination and photographed with a Canon EOS D30 digital camera using a 480 nm band pass filter (Peca products).

For papillation assays, transformants containing donor and target plasmids were selected on LB-agar containing 50 µg/ml ampicillin, 25 µg/ml chloramphenicol, 25 µg/ml kanamycin, 50 µM ZnSO₄, 0.1% lactose and 64 µg/ml X-gal, so as to obtain ~1000 colonies per 9 cm plate. Transformants were incubated at 37°C, and photographed every day for 6 days on a fluorescent light box using a Canon EOS D30 digital camera with no filter. Papillae were counted in the resulting image files using the 'threshold' function in Adobe Photoshop followed by 'analyze particles' in ImageJ.

RESULTS

Target choice by wild-type transposase

Before attempting to modify the target specificity of ISY100 transposition, we first investigated the target specificity of the wild-type system. The sequences of 21 ISY100 insertion sites from the *Synechocystis* PCC6803 genome (19), three from the large *Synechocystis* plasmids pSYSG and pSYSYA (20), 19 insertion sites from the work of Urasaki *et al.* (9), and 58 different insertion sites generated by *in vivo* and *in vitro* transposition into the target plasmid pH2 (10) were analysed. All except two insertions, both from *in vitro*

transposition reactions into pH2 using pre-cleaved transposon, were in TA target sites. We could detect no other differences in the pattern of sequences flanking insertions from different sources, so all 101 insertion sites were grouped together, aligned according to the direction of ISY100 and compared. In addition to the TA target sequence, there was significant sequence preference for the next three nucleotides on either side (Figure 1A). The A/T-rich consensus insertion site can be summarised as 5'-ADWTAWHT (where W = A or T, D = not C and H = not G), with the strongest preference being for A at position -4 and T at position +4. This consensus target sequence is palindromic, suggesting that ISY100 inserts into preferred target sites with no orientation specificity and that the left and right transposon ends are equivalent in the integration reaction.

Design of donor and target plasmids for Z-transposition

To make Z-transposases, the DNA-binding domain from the mouse transcription factor Zif268 was fused to the C-terminus of ISY100 transposase (Figure 1B). Zif268 binds as a monomer with high affinity and specificity to a single asymmetric 9–10 bp DNA site. It recognises its binding site using three independently folded zinc-finger domains, each recognising 3 bp of DNA via an α -helix inserted into the DNA major groove (21). Zinc-finger DNA-binding proteins with novel specificities can be assembled from libraries of modified zinc-fingers selected to bind to different triplet sequences (22,23). Engineered zinc-finger proteins based on Zif268 and related proteins are now available that bind with high affinity and specificity to many different chosen sequences, so there is the prospect of using them to direct transposition to any genomic target of interest.

Two short glycine- and serine-rich flexible linkers (L1, [GSG]₃ and L2, [GSG]₄) were used to fuse the Zif268 DNA-binding domain to ISY100 transposase (Figure 1B). In case these linkers were not sufficiently long to allow the catalytic domain of transposase to reach TA dinucleotides adjacent to the Zif268 binding site, a longer linker (L3), consisting of amino acids 141–185 of Tn3 resolvase and an additional 8 amino acids (Figure 1B) was also used. To construct donor plasmids, the genes encoding these Z-transposases, or the control wild-type transposase, were placed adjacent to an ISY100 mini-transposon, consisting of ISY100 left and right IRs flanking a 1.3 kb kanamycin resistance determinant, on a λ *dv*-based plasmid vector (Figure 2A). These donor plasmids replicate normally in most *E. coli* strains, but cannot replicate in λ lysogens, or other strains expressing the λ repressor.

Transposase tethered to the DNA by Zif268 might be constrained so that it can catalyse transposition into TA targets at a fixed distance from the Zif268 binding site. To study this systematically, an array of nine copies of the sequence [TANN] (containing nine TAs at 4 bp intervals) was placed adjacent to a 10 bp Zif268 binding site. The design of the array ensures that all except the first and last TA conform to the simplified ISY100 consensus target site ANNTANNT. To obtain TAs at all possible

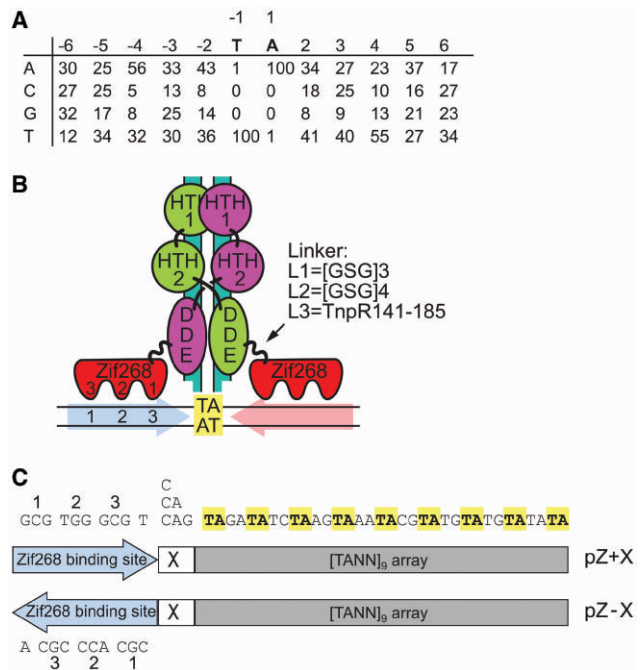


Figure 1. Design of targets and chimaeric Z-transposases. (A) Sequence summary of 101 different ISY100 integration sites in the *Synechocystis* PCC6803 genome and the target plasmid pAU5, pSEK80 and pH2. (B) ISY100 transposase is expected to act as a dimer (27); one monomer is coloured magenta, the other green. ISY100 transposase contains two N-terminal helix–turn–helix domains (HTH1 and HTH2), which recognize DNA sequences within the transposon IRs (turquoise) and a catalytic DDE domain which integrates the IRs into a target TA (yellow). The C-terminus of transposase was fused to the DNA-binding domain of Zif268 (red) via three different linkers: L1, L2 or L3. The three zinc-fingers of Zif268 are numbered 1 to 3, as are the three triplets of its recognition sequence. Note that finger 1 recognizes triplet 3, and finger 3 recognizes triplet 1. Target sites contain a single Zif268 binding site (blue arrow) adjacent to an array of target TAs. The dimeric nature of Z-transposase suggests how a second Zif268 binding site (pink arrow) could enhance the specificity or efficiency of targeting. (C) Target sites for chimaeric Z-transposase consist of a 10 bp Zif268 binding site separated from a [TANN]₉ array by 0, 1, 2 or 3 bp.

distances between 1 and 36 bp from the Zif268 binding site, the array was placed 0, 1, 2 or 3 bp from the Zif268 binding site. Since Zif268 might direct insertions to only one side of its asymmetric binding site, targets were constructed with both possible orientations of the Zif268 binding site. Target sites were placed on the 5.4 kb pH2, which was used as a target plasmid in our previous experiments with ISY100 (10). The resulting plasmids are here referred to as pZ±X, where ‘+’ or ‘-’ indicates the orientation of the Zif268 binding site, and X = 0, 1, 2 or 3 indicates the length of the spacer between the Zif268 binding site and the first TA of the array (Figure 1C). Two target plasmids lacking Zif268 binding sites were also constructed. These are identical to pZ+3 except that the Zif268 binding site (5'-GCGTGGGCGT) was changed to either 5'-GACTGGGGAG or 5'-GGAGC TCTCT, giving pB+3 and pF+3, respectively. The target plasmids all contain 340 TA dinucleotides outside of the [TANN]₉ array, and 10 inside.

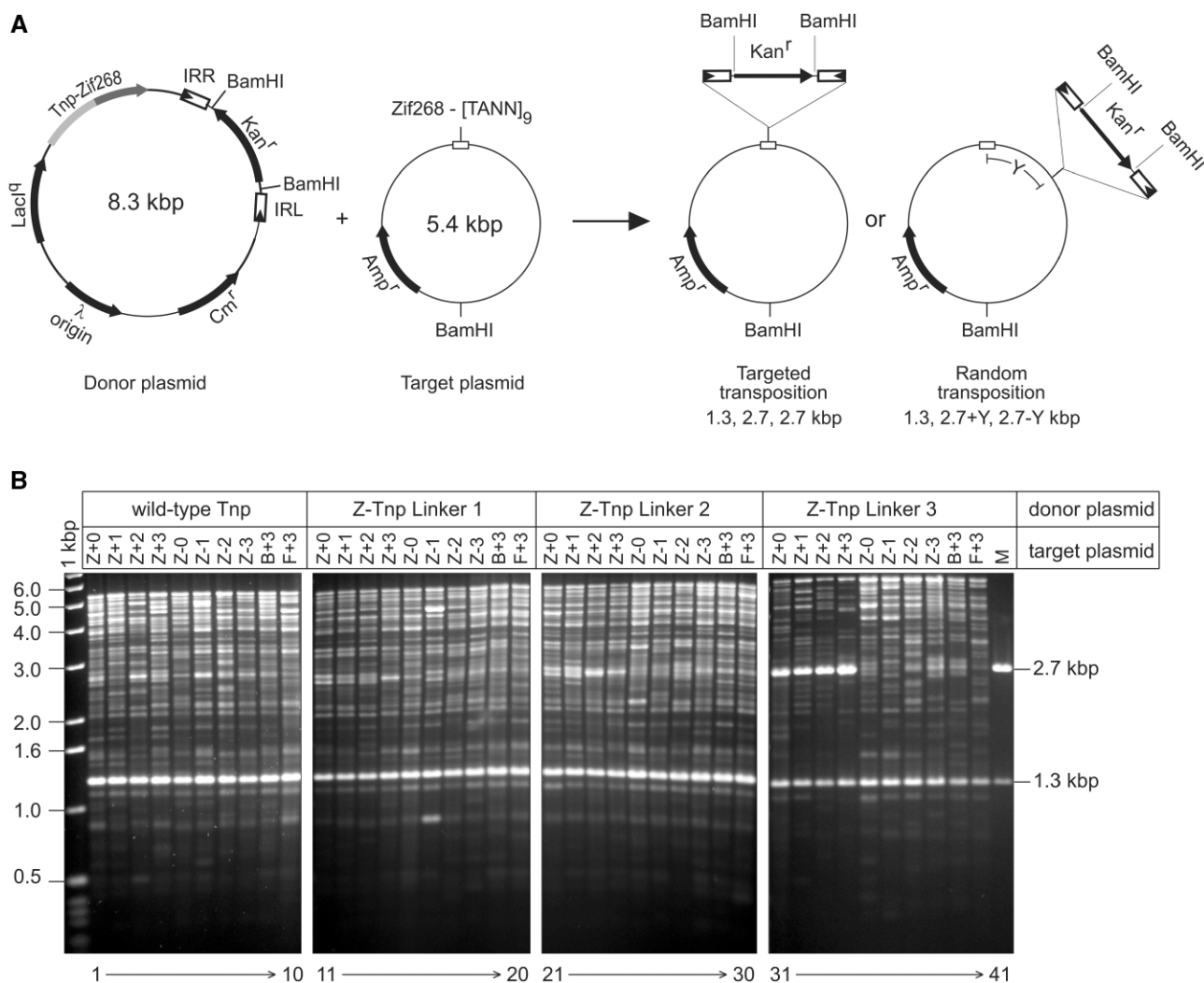


Figure 2. Targeted transposition of ISY100 by a chimaeric Z-transposase. **(A)** Structure of donor and target plasmids. Donor plasmids carry a mini-ISY100 transposon encoding resistance to kanamycin, and express transposase from a hybrid *trp-lac* promoter under the control of the plasmid-encoded *lacI^q* gene on a λ *dv* replicon. Target plasmids carry a [TANN]₉ array adjacent to a binding site for Zif268. On digestion with BamHI, transposition products give a 1.3 kb kanamycin resistance fragment and two other fragments that add up to 5.4 kb. If the transposon is in the [TANN]₉ target, a 2.7 kb doublet is produced. **(B)** Agarose gel showing BamHI-digested pooled transposition products from assays using the indicated donor and target plasmids. The sizes of the kan^r fragment (1.3 kb) and the doublet indicative of targeted transposition (2.7 kb) are shown. Lane 41 (M) contains BamHI-digested DNA from an isolated targeted transposition product.

Targeted transposition by Z-transposase

In vivo transposition assays were carried out using all four donor plasmids in all possible combinations with the 10 different target plasmids, making a total of 40 different transposition assays. After incubating donor and target plasmids together in a *recA*-deficient *E. coli* strain to allow transposition to occur, plasmid DNA, containing a mixture of donor and target plasmids and any transposition product, was purified from pooled colonies. Transposition products were isolated from this mixture by transformation into a λ lysogen strain and selecting for kanamycin-resistant colonies. Since the donor plasmids cannot replicate in this strain, kanamycin-resistant colonies are only obtained if the minitransposon has jumped from the donor to the target plasmid.

For each combination of donor and target plasmid, the transposition frequency (the fraction of target plasmids containing a transposon insertion) was calculated by dividing the number of kanamycin-resistant colonies by the number of ampicillin-resistant colonies. Averaged over all target plasmids, wild-type transposase gave a transposition frequency of 1.3×10^{-4} , while Z-transposases with linkers L1 ([GSG]₃) and L2 ([GSG]₄) gave average transposition frequencies ~ 3.5 -fold lower (Table 1). Z-transposase with the longer L3 linker gave transposition at still lower frequencies, ~ 9 -fold lower than wild-type (Table 1). For each donor plasmid, the different target plasmids were used with approximately equal efficiency, except that L3 Z-transposase gave higher transposition frequencies into target plasmids with the Zif268 binding site in the '+' orientation than into target plasmids

Table 1. Transposition frequencies ($\times 10^{-6}$) calculated as the ratio of $\text{kan}^r/\text{amp}^r$ colonies as described in the text

Target	Donor			
	Wild-type Tnp	Z-Tnp L1	Z-Tnp L2	Z-Tnp L3
Z+0	143 (± 18)	41 (± 17)	38 (± 20)	25 (± 14)
Z+1	97 (± 64)	44 (± 19)	68 (± 33)	16 (± 12)
Z+2	98 (± 37)	39 (± 13)	56 (± 33)	23 (± 14)
Z+3	115 (± 16)	31 (± 16)	46 (± 30)	17 (± 9)
Z-0	128 (± 42)	44 (± 31)	26 (± 13)	14 (± 6)
Z-1	162 (± 41)	52 (± 29)	45 (± 26)	13 (± 7)
Z-2	198 (± 42)	31 (± 13)	34 (± 19)	11 (± 4)
Z-3	125 (± 62)	24 (± 5)	37 (± 21)	12 (± 5)
B+3	134 (± 11)	36 (± 23)	25 (± 13)	12 (± 10)
F+3	136 (± 51)	14 (± 8)	20 (± 5)	8 (± 4)
Average	133 (± 46)	38 (± 24)	41 (± 26)	15 (± 10)

Average frequencies ($\pm 1\sigma$) were calculated from at least three independent experiments for each combination of donor and target plasmids.

without Zif268 binding sites, or with Zif268 binding sites in the ‘-’ orientation (Table 1).

The distribution of mini-transposon insertions in pools of ~ 100 transposition products was displayed by agarose gel electrophoresis of BamHI-digested plasmid DNA (Figure 2B). The mini-transposon contains two BamHI sites just inside the ISY100 IRs, while the 5.4 kb target plasmids contain a single BamHI site diametrically opposite to the [TANN]₉ target sequence (Figure 2A). All transposition products therefore yield a 1.3 kb fragment released from the mini-transposon and two fragments totalling 5.4 kb. When the insertion is in the [TANN]₉ target array, these two fragments are both 2.7 kb (Figure 2A).

As expected, wild-type transposase gave a 1.3 kb transposon fragment and a ladder of bands representing insertions at many different positions in all of the target plasmids (Figure 2B, lanes 1–8). L1 and L2 Z-transposases gave similar ladders of bands, with little or no indication of an increase of transposition close to the Zif268 binding site (Figure 2B, lanes 11–18 and 21–28). In target plasmids with the Zif268 binding site in the ‘-’ orientation (Figure 1B), L3 Z-transposase also gave insertions at positions distributed randomly around the target plasmid (Figure 2B, lanes 35–38). In striking contrast, transposition products made by L3 Z-transposase in targets with the ‘+’ orientation of the Zif268 binding site were strongly biased towards insertion adjacent to the Zif268 binding site (Figure 2B, lanes 31–34). The 2.7 kb fragment, diagnostic of insertions into the [TANN]₉ array, was approximately double the intensity of the 1.3 kb transposon-derived fragment in these lanes. As the 2.7 kb band is a doublet and is twice the size of the 1.3 kb fragment, the relative intensity of these bands suggests that about 50% of the insertions are in or very close to the [TANN]₉ target array.

To quantify these results more accurately, individual transposition products were characterised by restriction digestion. $\sim 4\%$ of insertions by wild-type transposase were in the [TANN]₉ array (Table 2), roughly consistent

Table 2. Fraction of transposon inserts within [TANN]₉ array

Target	Donor	
	Wild-type	L3 Z-Transposase
Z+0	0/24	12/24
Z+1	0/24	11/24
Z+2	3/24	12/24
Z+3	1/24	11/24
Z-0		2/22
Z-1		2/24
Z-2		4/24
Z-3		1/24
B+3		1/24
F+3		1/19

Approximately 24 transposition products were selected at random for each combination of donor and target plasmid. Restriction digestion and DNA sequencing were used to determine the number containing inserts within the [TANN]₉ array.

with the fraction of TAs in the target plasmid within the array (2.6%). In contrast, nearly 50% of insertions catalysed by L3 Z-transposase in pZ+X target plasmids were in the TA target array (Table 2), ~ 19 times more than expected by chance. When the Zif268 binding site was in the opposite orientation (in pZ-X targets) only $\sim 9\%$ of insertions catalysed by L3 Z-transposase were in the TA array (Table 2).

Targeted transposition was dependent on the presence of a functional Zif268 binding site adjacent to the [TANN]₉ array. Changing the Zif268 binding site in pZ+3 from 5'-GCGTGGGCGT to either 5'-GACTGGG GAG or 5'-GGAGCTCTCT, in target plasmids pB+3 and pF+3 respectively, abolished targeted transposition by L3 Z-transposase (Figure 2B, lanes 39 and 40), but had no discernible effect on transposition by wild-type, L1 Z-transposase, or L2 Z-transposases (Figure 2B, lanes 9, 10, 19, 20, 29 and 30).

All 40 transposition assays (Table 1) were repeated independently at least three times. In all cases, the same results were obtained. L3 Z-transposase catalysed insertion at high frequency adjacent to the Zif268 binding site only when it was in the ‘+’ orientation, whereas wild-type and Z-transposases with L1 and L2 linkers catalysed transposition at many different positions distributed randomly throughout the target plasmids. In transposition assays with wild-type transposase, L1 Z-transposase, and L2 Z-transposase, some pairs of bands adding up to 5.4 kb were more intense than others (Figure 2B). Where these bands are present in all lanes, they presumably reflect preferred integration sites in the target plasmid. Where they appear only in some lanes (Figure 2B, lanes 16 and 25), these intense bands are thought to reflect transposition events in one transformant colony soon after the introduction of the donor plasmid into cells containing target plasmid. These early events are amplified during growth of the colonies in which they occurred, and are therefore over-represented in the pooled transposition products. In contrast to L3 Z-transposase with pZ+X targets, which always gave a

strong band at the 2.7 kb position, these early transposition events were different in independent replicates of the same experiment, reflecting the stochastic nature of the assay.

In the absence of a Zif268 binding site, or when the Zif268 binding site was in the ‘-’ orientation, there was no evidence that any secondary sites were used at high frequency by L3 Z-transposase. More than 30 different integration sites used by L3 Z-transposase can be identified from examination of BamHI-digested DNA (Figure 2, lanes 35–40) and these sites seem similar to those used by wild-type transposase (Figure 2, lanes 1–10).

Specific TAs in the [TANN]₉ array are targeted

Individual transposition products containing inserts in or close to the [TANN]₉ target array were further characterised by DNA sequencing. This confirmed the location of insertions in the [TANN]₉ array, and allowed their exact positions to be established (Figure 3). The majority of insertions catalysed by L3 Z-transposase in the [TANN]₉ arrays of pZ+2 and pZ+3 were in the second TA of the array (24 out of 42 insertions in pZ+2 and 34 out of 46 in pZ+3), so that approximately one-third of all insertions in these plasmids were in a single TA 6–7 bp from the Zif268 binding site. When the [TANN]₉ array was moved one or two base-pairs closer to

the Zif268 binding site, almost no insertions were in the second TA, and instead insertions were concentrated in the fourth and seventh (pZ+0) or the fifth and seventh (pZ+1) TAs (Figure 3). Because of the way the experiment was carried out, the individual insertions described earlier are not necessarily all independent. However, the results were derived from at least four independent assays for each combination of donor and target plasmids, and each separate assay gave a similar distribution of insertion site preference. When the Zif268 binding site was in the ‘-’ orientation, only a small number of insertions in the [TANN]₉ array were obtained, but again there appeared to be a preference for the second and seventh TAs (Figure 3).

To see if these insertion sites were preferred because of their position relative to the Zif268 binding site, or because their local sequence context made them preferred targets for the ISY100 catalytic domain, sites of insertions in the [TANN]₉ array catalysed by wild-type transposase were also determined (Figure 3). Every TA in the [TANN]₉ array, except the first and the fifth, was used at least once by wild-type transposase, but the two most common integration sites were the second and seventh (Figure 3). Therefore, although the numbers are small, wild-type transposase appears to have a preference for the second and seventh TAs in the array, and this

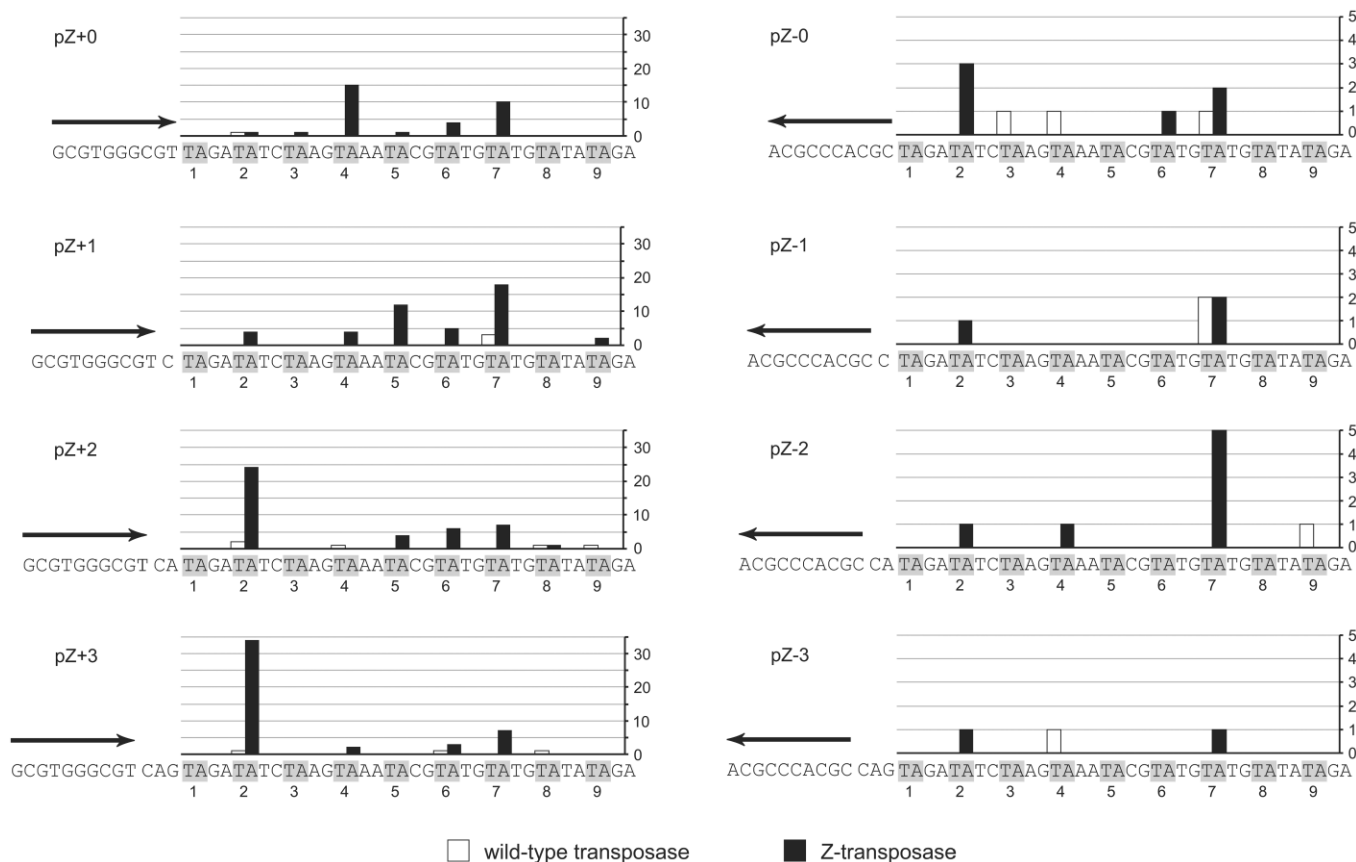


Figure 3. Distribution of insertion in the [TANN]₉ array. Histograms show the number of insertions catalysed by wild-type and L3 Z-transposase into each TA in the array. The target sequences are shown below each histogram. Arrows represent Zif268 binding sites. Note the different orientation of the Zif268 binding site and the different vertical scales in left and right panels.

may explain at least part of the target site preference by L3-Z-transposase.

A papillation assay for targeted transposition

Next, a papillation assay was designed so that targeted transposition could be followed in individual transformant colonies. A target plasmid (pZ+3_{pap}) was constructed with a Zif268 binding site 3 bp away from a [TANN]₄ array (the same spacing as in pZ+3) upstream of a promoterless copy of the *lacZ* gene lacking its first eight codons (Figure 4A). The donor plasmid (pZ-DONOR_{pap}) carried the gene for L3 Z-transposase and a modified ISY100 mini-transposon, containing a ribosome binding site and ATG translational start reading out through the right end of the minitransposon (Figure 4A). Donor and target plasmids were designed such that transposition of the modified mini-ISY100 in the correct orientation into the second TA of the [TANN]₄ target array (the preferred TA in pZ+3) forms a transcriptional and translational fusion, leading to expression of functional *lacZ*.

Assays were carried out by introducing the donor plasmid into cells already containing the target plasmid, and selecting transformant colonies on solid media containing X-gal and lactose. Transformant colonies are expected to be white because there is no *lacZ* expression. However, any cell in the colony which becomes Lac⁺, due to transposition in the correct orientation into the second TA of the [TANN]₄ array, will have a growth advantage on the lactose-containing media and will produce a blue minicolony (or papilla) on the otherwise colourless colony. The number of blue papillae on a colony therefore gives an indication of the rate of targeted transposition within that colony.

When this assay was first carried out with pZ-DONOR_{pap} and pZ+3_{pap}, only 10–20 blue papillae were seen per plate of several thousand colonies, and no more than one papilla was seen per colony. Transposase is expressed from a hybrid *trc-lac* promoter on pZ-DONOR_{pap}, and is repressed by the lactose repressor, encoded by *lacI* on the donor plasmid (Figure 4A). Lactose present in the papillation plates is expected to induce transposase expression, and may lead to ‘overproduction inhibition’ of transposition, as has been seen for other members of the Tc1/*mariner*/IS630 family (24,25). To try to obtain a higher level of transposition in the papillation assay, the region of the donor plasmid containing *lacI*, the *trp-lac* promoter and the Z-transposase gene was subjected to random mutagenesis by PCR, and donor plasmids yielding increased numbers of blue papillae were selected. Two such plasmids were isolated and characterized. These mutant donor plasmids reproducibly gave 10–100 blue papillae per colony, increasing in number over a period of several days at 37°C. Both of these donor plasmids contained mutations in the *lacI* gene (D275Y and S193L), at positions known to give a lactose insensitive (LacI^S) phenotype (26), such that the promoter driving expression of Z-transposase will be repressed even in the presence of lactose. One mutant donor plasmid (pZ-DONOR_{pap}* carrying the mutation D275Y) was chosen for use in further experiments.

Restriction digestion and sequencing of a number of transposition products from blue papillae produced by pZ-DONOR_{pap}* and pZ+3_{pap} confirmed their expected structure, with transposon insertions in the correct orientation in the second TA of the [TANN]₄ array.

Targeted transposition is most efficient 7–17 bp from the Zif268 binding site

Sequencing of targeted transposition events in the [TANN]₉ array showed that some TAs were preferred over others (Figure 3). This appeared to be due to a combination of preferential integration at certain distances from the Zif268 binding site, and an inherent preference of the catalytic domain for particular TAs in the array. To gain more insight into this, the papillation assay was used to investigate the effect of varying the distance between the Zif268 binding site and the [TANN]₄ array. The second TA, into which the mini-ISY100 must insert to make a functional *lacZ* fusion, is 7 bp away from the Zif268 binding site in the original papillation target plasmid (pZ+3_{pap}; Figure 4B). Sequences devoid of TA dinucleotides were inserted between the Zif268 binding site and the first TA of the [TANN]₄ array to move the target TA 8, 10, 12, 15, 17, 22 and 27 bp from the Zif268 binding site in target plasmids pZ+4_{pap}, pZ+6_{pap}... pZ+23_{pap} (Figure 4B). A papillation target plasmid (pZ-3_{pap}) with the same spacing as pZ+3_{pap} but with the Zif268 binding site in the opposite orientation was also constructed. Papillation assays were carried out using these new target plasmids and the pZ-DONOR_{pap}* donor plasmid (Figure 4C), and the results were quantitated by counting the number of blue papillae per plate at different time points (Figure 4D). The results confirmed that targeted transposition only occurs efficiently with the ‘+’ orientation of the Zif268 binding site, and that efficiency of targeted transposition falls off as the target TA is moved to 22 bp or more from the Zif268 binding site. The relatively low targeted transposition efficiency when the target TA is 15 bp from the Zif268 binding site in pZ+11_{pap} suggests there may also be helical phase effects.

Because transposases of the Tc1/*mariner* family act as dimers (27), and targeted integration was observed to just one side of the Zif268 binding site, we reasoned that two Zif268 binding sites pointing towards each other might direct transposition to a TA target between them (Figure 1B). A target site (ZZ) with two Zif268 binding sites, one with same spacing as in pZ+8 and the other with the same spacing as in pZ+3, flanking a [TANN]₃ array in a head-to-head arrangement was therefore constructed (Figure 4B). As predicted, this target was an efficient substrate for targeted transposition in the papillation assay (Figure 4C and D).

The target specificity of Z-transposase can be changed

Two engineered Zif268 variants that bind specifically to 10 bp sequences from the promoter region of HIV-1 (18) were used to alter the target specificity of Z-transposase. Zif268-A binds specifically to the sequence 5'-AGGGAG GCGT and Zif268-B binds specifically to 5'-GACTGGG

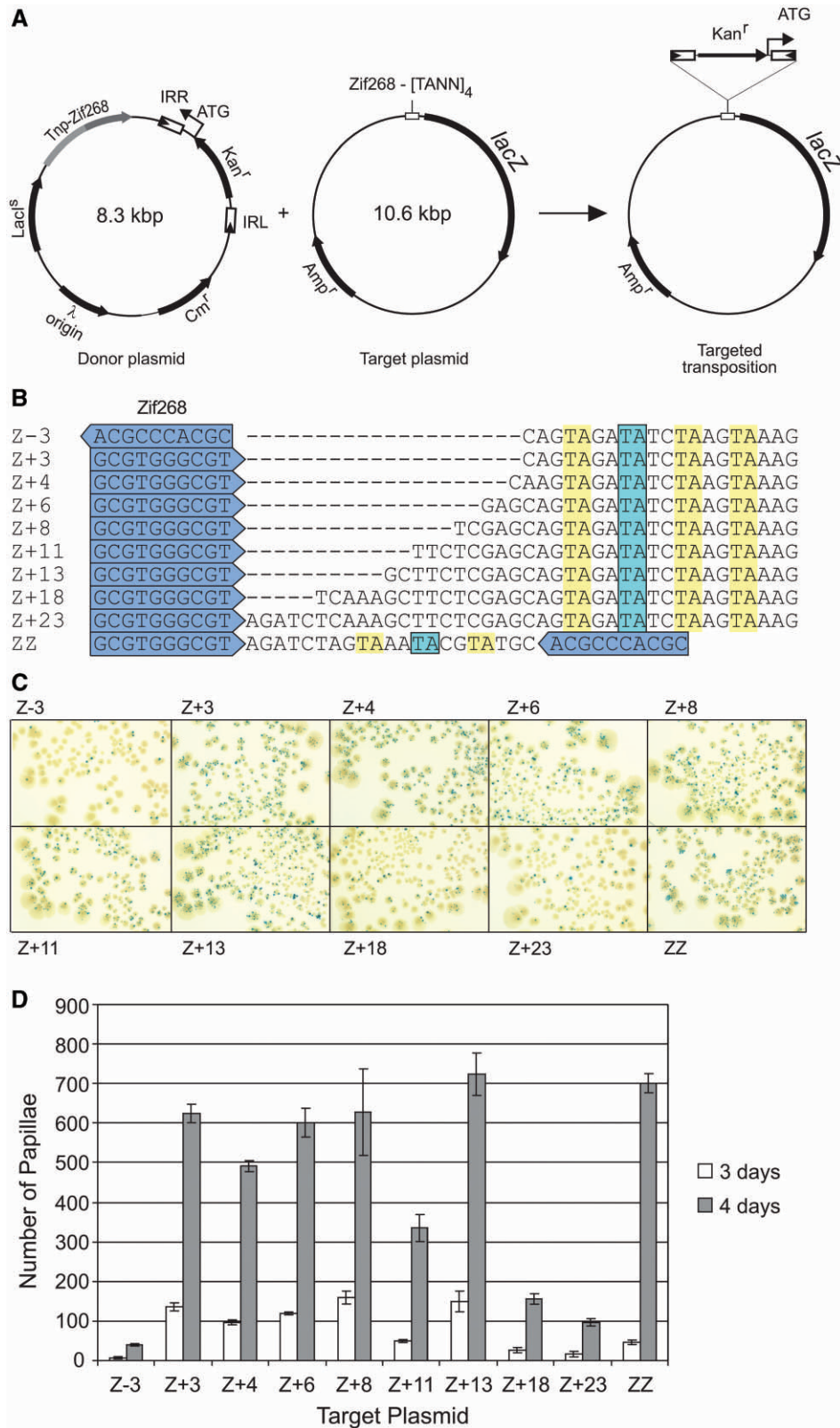


Figure 4. Papillation assays for targeted transposon. (A) Transposition from the donor plasmid to the second TA in the [TANN]₄ array upstream of *lacZ* in the target plasmid creates a functional *lacZ* translational fusion. (B) Sequences of the different target sites used. The Zif268 binding sites are highlighted by blue arrows; the four TA dinucleotides in the [TANN]₄ array are hi-lighted in yellow or cyan for the second TA, into which the transposon must insert to form a functional *lacZ* fusion. The ZZ target contains two Zif268 binding sites flanking a [TANN]₃ array (C) Papillation assays with Z-transposase donor and the different target plasmids. (D) The histogram shows the number of papillae per plate of ~1000 colonies after 3 and 4 days of incubation at 37°C. The values shown are the mean from two independent replicates; error bars represent the range between high and low data points.

GAG, with reported dissociation constants of 1.2 nM and 1.0 nM, respectively (Figure 5A). The Zif268 binding site in the papillation target plasmid pZ+3_{pap} was changed to the A and B binding sites to give pA+3_{pap} and pB+3_{pap} target plasmids, respectively, and the Zif268 DNA-binding domain in the papillation donor plasmid pDONOR-Z_{pap}* was changed to the A and B variants to give two new donor plasmids (pDONOR-A_{pap} and pDONOR-B_{pap}). Papillation assays were then carried out with all combinations of donor and target plasmids (Figure 5B). The original p-DONOR-Z_{pap}* donor gave the expected high number of papillae with the pZ+3_{pap} target plasmid but much lower numbers of papillae per

colony with the altered target plasmids. Although the level of targeted transposition was somewhat lower than with Z-transposase carrying the native Zif268 DNA-binding domain, A-transposase and B-transposase preferentially catalysed targeted transposition into their cognate target sites (Figure 5B and C), demonstrating that the specificity of targeted transposition can be switched using altered specificity zinc-finger variants.

DISCUSSION

In this study, we set out to create a targeted DNA integration system by combining the DNA-binding specificity

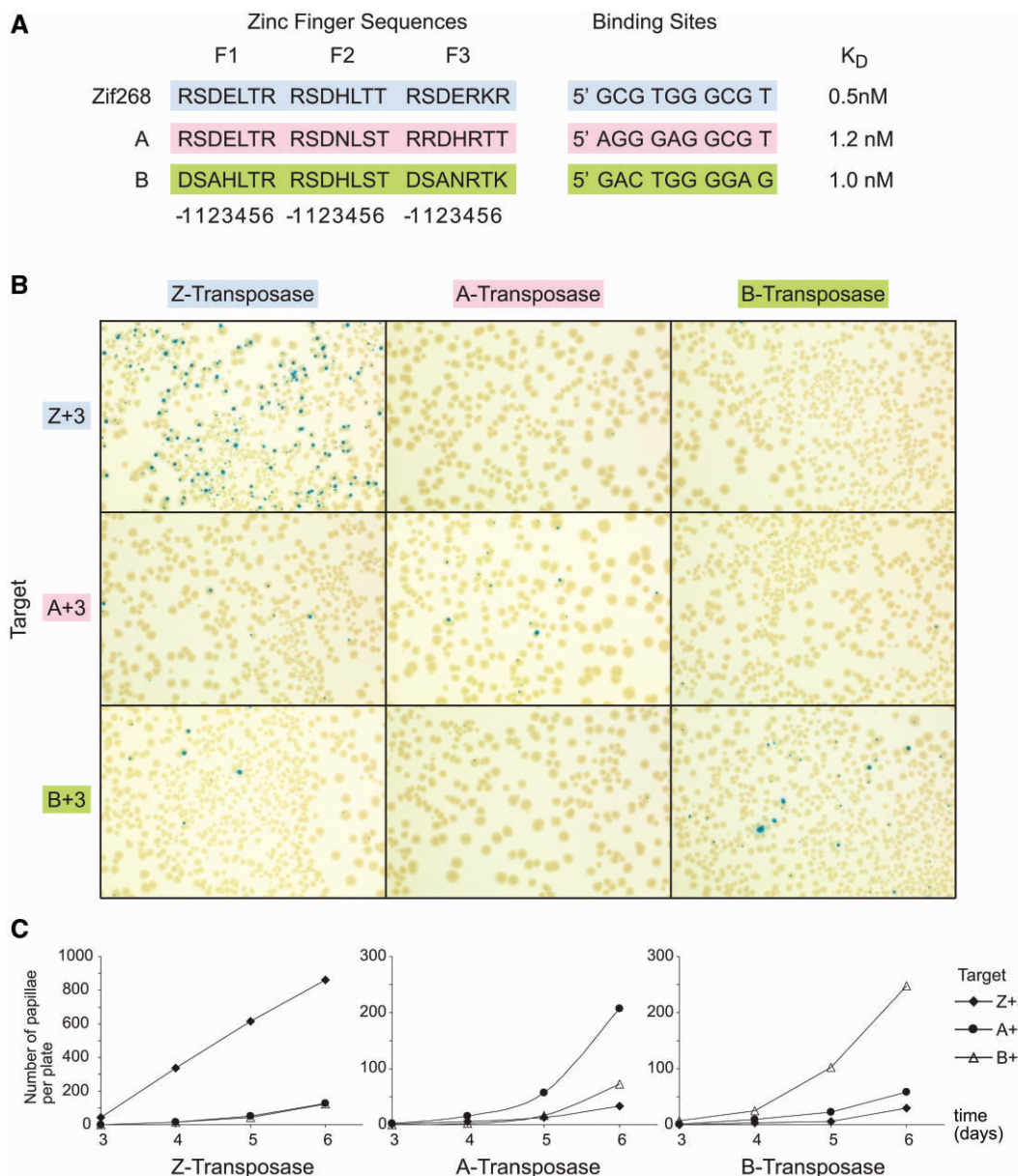


Figure 5. The specificity of targeted transposition can be changed by altering the recognition helices of the Zif268 DNA-binding domain. (A) Sequences of recognition helices from the three fingers F1, F2 and F3 of Zif268 and altered specificity mutants Zif-A and Zif-B are shown along with the DNA sequences they recognize and their reported dissociation constants (18). (B) Papillation assays using donor and target plasmids indicated. Plates were photographed after 5 days of incubation. (C) Charts showing the total number of papillae per plate at different time points for each combination of donor and target plasmid.

of a zinc-finger domain with the TA target specificity of the Tc1/*mariner* family transposon ISY100. Fusion of the Zif268 DNA-binding domain to ISY100 transposase via protein linker L3 reduced the overall level of transposition into target plasmids without a Zif268 binding site by a factor of ~ 10 (Table 1). When the target plasmid contained a Zif268 binding site in one orientation adjacent to a [TANN]₉ array, the transposition frequency was partially restored, and nearly 50% of insertions were in the [TANN]₉ array.

When the Zif268 binding site was in the opposite orientation, there was no increase in transposition and no targeted integration was observed. It appears that tethering the catalytic domain of transposase to the DNA via Zif268 and the L3 linker, positions it so that transposition takes place into TAs a short distance only to one side of the Zif268 binding site. Zif268 binds with a fixed orientation to its asymmetric binding site, and is expected to position the C-terminal catalytic domain of transposase closest to the third triplet of the Zif268 binding site (Figure 6), in full agreement with the observed orientation specificity of targeted transposition.

Although a [TANN]₉ array adjacent to a Zif268 binding site was used in some of the experiments reported here, shorter [TANN]₄ and [TANN]₃ arrays were sufficient in the papillation assay. We predict that any TA conforming to the preferred target site consensus will act as an efficient

target for Z-resolvase when placed at an appropriate distance from a Zif268 binding site.

Retroviral vectors have been used extensively to integrate DNA in gene therapy trials (28). Retroviral vectors insert at many different locations in the genome, with a bias towards active genes. Much work has gone into the modification of their target specificity to avoid the problems associated with integration at unwanted genomic locations (29). Retroviral integrases fused to the bacterial DNA-binding proteins LexA and the λ repressor, or to zinc-finger DNA-binding domains, directed increased levels of integration *in vitro* near to appropriate binding sites on the target DNA (30–35). However, considerable integration still occurred at sites distant from the desired target site, and integrase fusion proteins were poorly incorporated into active virus particles. One lab successfully incorporated an HIV-1 integrase zinc-finger fusion protein into infective virus particles and obtained a 10-fold increase in integration near to the genomic zinc-finger binding site in cultured human cells (35). A more recent approach has been to tether the HIV-1 integrase-binding protein LEDGF/p75 to specific sites by fusing it to a sequence-specific DNA-binding domain (36), and a similar approach has been taken with the yeast retrotransposon Ty5 (37).

Transposons have also been used with great success in gene modification applications, first in bacteria and invertebrates, and more recently in vertebrates (38). However, transposons also integrate at many different genomic sites with possible deleterious consequences. One widely used vertebrate transposon from the Tc1/*mariner* family, Sleeping Beauty (SB), transposes and can stably integrate transgenes in a wide variety of eukaryotic cells. Attempts to fuse DNA-binding domains to the C-terminus of SB transposase have led to total loss of activity (39–41). Fusions to the N-terminus of SB transposase retain 10–20% of wild-type activity, and one group reported site-directed transposition in a plasmid to plasmid transposition assay in human cells (41). An 8- to 11-fold increase in transposition into a 443 bp region around a cluster of tandemly repeated binding sites for the DNA-binding domain used was observed. However, transposition was not targeted to a single integration site in the plasmid assay and no targeted transposition was observed in genomic targets. The target specificity of the bacterial IS30 transposase has also been successfully modified by fusing it to two different DNA-binding domains (42), and similar experiments have also been reported for the eukaryotic Mos1 and *piggyBac* transposases (43). Two other approaches that have been tried with some success for SB are: (i) tethering the transposon DNA itself to a specific target using a bifunctional DNA-binding protein, or (ii) using a domain that binds tightly to transposase fused to a DNA-binding domain to tether the transposase to the desired target (39). This latter approach worked efficiently on a chromosomal target in human cells, with $\sim 10\%$ of insertions in a 2.6 kb target region (39).

ISY100 is from the same Tc1/*mariner* family as SB, but while fusions to the C-terminus of SB transposase completely inactivated it, C-terminal fusions of the

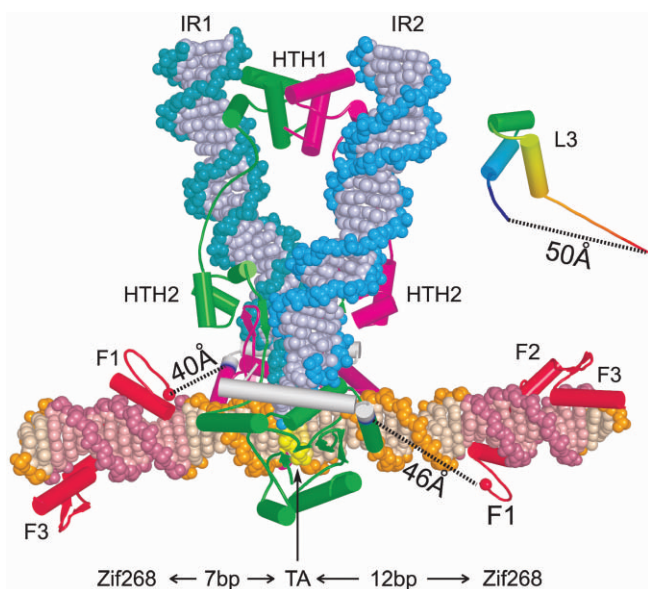


Figure 6. Molecular model of Z-transposase bound to the ZZ target site. A dimer of Mos1 transposase (magenta and green cylinders; PDB ID: 3HOT) bound to two transposon IRs (blue and turquoise) in a paired-end complex is modeled onto the ZZ target site as B-DNA to represent the ISY100 target-capture complex. The Zif268 binding sites are coloured pink, the target TA is yellow and the rest of target DNA is orange. In this view, the catalytic domain from the green monomer sits entirely in front of the target DNA, while that of the magenta monomer is behind. The C-terminal helical region of Mos1 transposase absent in ISY100 transposase is shown in grey. Zif268 (red) is bound to its two recognition sequences 7 and 12 bp from the target TA. Linker L3 (inset rainbow cylinders; from PDB ID: 1GDT) would have to stretch 40 Å and 46 Å (dotted lines) from the predicted C-termini of ISY100 transposase (blue) to the N-terminus of Zif268 (red sphere).

Zif268 DNA-binding domain to ISY100 transposase (described above) retained 10–30% of wild-type activity (Table 1). Examination of the recently published structure of the Mos1 paired end complex (27) shows that the C-terminus of Mos1 is close to the binding site for transposon ends, and the clamp-loop linker interaction that holds the dimeric complex together. Fusion of a large protein domain to the C-termini of Mos1 and SB transposases is therefore likely to interfere with binding to the transposon ends and disrupt formation of the paired-end complex required for catalysis. A structure-based sequence alignment of Mos1, SB and ISY100 transposases shows that while SB and Mos1 have similar C-termini, ISY100 lacks a C-terminal α -helical region of 32 amino acids present in Mos1 (10). Fusion of a protein domain to the C-terminus of ISY100 transposase therefore seems much less likely to interfere with DNA binding and dimerization of ISY100 transposase (Figure 6). We did not attempt to fuse a DNA-binding domain to the N-terminus of ISY100 transposase. However, because the N-terminus is further from the catalytic target-binding domain of transposase than the C-terminus, it seems likely that an N-terminal fusion would not target transposition as specifically as the C-terminal fusion studied here.

Targeted transposition required linker L3, consisting of residues 141–185 of Tn3 resolvase. These residues form an autonomously folding three-helix DNA-binding domain that binds specifically to the sequence TGTC CG (44,45). However, this sequence is not present in the ISY100 IRs or in any of the target sequences used in our experiments. Molecular modelling of target capture complexes formed by Z-transposase suggest that the linker between the C-terminus of transposase and the N-terminus of the Zif268 DNA-binding domain will have to stretch 30–60 Å to allow transposase to reach target TAs 7–17 bp from the Zif268 binding site (Figure 6). Linker L3 could stretch \sim 50 Å if the Tn3 resolvase DNA-binding domain remains folded and we propose that this linker is required simply to stretch the long distances required. Consistent with our results, molecular modelling also suggests that steric clashes between Zif268 and transposase will prevent integration at TAs closer than about 6 bp from the Zif268 binding site, and that linker L3 will have to stretch right round the DNA helix in the poorly favoured pZ+11_{pap} target plasmid.

L3 Z-transposase carried out a background level of transposition at sites outside the [TANN]₉ array, in plasmids both with or without the Zif268 binding site. This was most likely catalysed by Z-transposase molecules interacting with target sites using solely the target binding activity of the transposase catalytic domain. Mutations in the transposase target-binding region, that render transposase unable to bind target DNA without assistance from the zinc-finger DNA-binding domain, might abolish this undesirable off-target transposition.

We describe here a papillation assay that reports the level of targeted transposition by Z-transposase into a specific TA in a single bacterial colony. This assay could be used to select Z-transposase mutants with increased

levels of targeted transposition, and could be combined with an assay to select for loss of ‘off-target’ transposition to obtain Z-transposase mutants with increased targeting specificity. The overall activity of Z-transposase might also be increased by the incorporation of ‘hyperactivity’ mutations, identified in wild-type transposase using an assay for untargeted transposition.

To be useful for gene delivery in eukaryotic systems, a targeted transposition system must be capable of delivering its cargo to a single site in a genome of over 10⁹ bp. A 9 bp zinc-finger binding site is insufficient to specify a unique site in a genome of this size, but two 9 bp binding sites should suffice. We have observed efficient transposition into a TA located between two Zif268 binding sites in IR (Figure 4C). From the dimeric structure of transposase, it should be possible to select Z-transposase variants that are dependent on zinc-finger-mediated binding to a pair of appropriately spaced sequence motifs for the assembly of an active dimer. By changing the recognition helices of the zinc-finger domains, these Z-transposases could be designed to target many different chosen single sites in a large eukaryotic genome.

Z-transposase with the wild-type Zif268 domain gave a higher level of targeted transposition than those with Zif268 domains selected to bind to sequences from the HIV-1 promoter (Figure 5). One possible explanation for this is that wild-type Zif268 domain binds to its site with a higher affinity than either ZifA or ZifB variants (18). Therefore, another approach to increasing the specificity and efficiency of targeted transposition might be to use longer, tighter binding zinc-finger proteins, such as the synthetic polydactyl zinc-finger protein E2C. This DNA-binding domain contains six zinc-finger domains and binds with high affinity and specificity to an 18 bp sequence that occurs just once in the human genome (46).

We have yet to show that ISY100 transposase will function in eukaryotic cells. Nevertheless, there is a reasonable possibility that ISY100 transposase (or activated mutants thereof) will function in eukaryotes because: (i) ISY100 transposition requires no host specific proteins, (ii) other members of the same family (such as *Minos*, *SB* and derivatives of *Himar1*) work efficiently in vertebrate cells (47) and (iii) other bacterial recombinases, such as Cre and ϕ C31 integrase, work efficiently in eukaryotic cells and have found many useful applications. Even if ISY100 derived Z-transposases do not function in eukaryotic cells, the lessons learned from them using our tractable genetic and biochemical techniques should be applicable to other related systems.

With 50% of insertions in a 20 bp region adjacent to the Zif268 binding site, and up to 70% of these in a single TA dinucleotide, our ISY100 Z-transposase is one of the most specifically targeted transposition systems developed to date. The specificity of targeting can be altered using zinc-finger DNA-binding domains that have been selected to bind to different DNA sequences, and we believe that this system holds great promise for the development of a precisely targeted integration system that will be useful for a number of applications.

ACKNOWLEDGEMENTS

We thank Aram Akopian, Chris Proudfoot and Marshall Stark for the kind gift of DNA fragments containing Zif268 and linker sequences. We thank Marshall Stark for constructive comments on the manuscript and Elizabeth Kilbride for technical assistance.

FUNDING

Wellcome Trust (Senior Fellowship 057651 to S.D.C.; Prize Studentship 066001 to X.F.). Funding for open access charge: University of Glasgow Wellcome Trust Open Access Fund.

Conflict of interest statement. None declared.

REFERENCES

- Sorrell,D.A. and Kolb,A.F. (2005) Targeted modification of mammalian genomes. *Biotechnol. Adv.*, **23**, 431–469.
- Ivics,Z., Li,M.A., Mates,L., Boeke,J.D., Nagy,A., Bradley,A. and Izsvak,Z. (2009) Transposon-mediated genome manipulation in vertebrates. *Nat. Methods*, **6**, 415–422.
- Cockrell,A.S. and Kafri,T. (2007) Gene delivery by lentivirus vectors. *Mol. Biotechnol.*, **36**, 184–204.
- Hacein-Bey-Abina,S., Von Kalle,C., Schmidt,M., McCormack,M.P., Wulffraat,N., Leboulch,P., Lim,A., Osborne,C.S., Pawliuk,R., Morillon,E. *et al.* (2003) LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science*, **302**, 415–419.
- Plasterk,R.H., Izsvak,Z. and Ivics,Z. (1999) Resident aliens: the Tc1/Mariner superfamily of transposable elements. *Trends Genet.*, **15**, 326–332.
- Bryan,G., Garza,D. and Hartl,D. (1990) Insertion and excision of the transposable element mariner in *Drosophila*. *Genetics*, **125**, 103–114.
- Emmons,S.W., Yesner,L., Ruan,K.S. and Katzenberg,D. (1983) Evidence for a transposon in *Caenorhabditis elegans*. *Cell*, **32**, 55–65.
- Collins,J., Forbes,E. and Anderson,P. (1989) The Tc3 family of transposable genetic elements in *Caenorhabditis elegans*. *Genetics*, **121**, 47–55.
- Urasaki,A., Sekine,Y. and Ohtsubo,E. (2002) Transposition of cyanobacterium insertion element ISY100 in *Escherichia coli*. *J. Bacteriol.*, **184**, 5104–5112.
- Feng,X. and Colloms,S.D. (2007) In vitro transposition of ISY100, a bacterial insertion sequence belonging to the Tc1/mariner family. *Mol. Microbiol.*, **65**, 1432–1443.
- Ivics,Z., Hackett,P.B., Plasterk,R.H. and Izsvak,Z. (1997) Molecular reconstruction of *Sleeping Beauty*, a Tc1-like transposon from fish, and its transposition in human cells. *Cell*, **91**, 501–510.
- Lampe,D.J., Churchill,M.E. and Robertson,H.M. (1996) A purified mariner transposase is sufficient to mediate transposition *in vitro*. *EMBO J.*, **15**, 5470–5479.
- Ivics,Z. and Izsvak,Z. (2004) Transposable elements for transgenesis and insertional mutagenesis in vertebrates: a contemporary review of experimental strategies. *Methods Mol. Biol.*, **260**, 255–276.
- Voigt,K., Izsvak,Z. and Ivics,Z. (2008) Targeted gene insertion for molecular medicine. *J. Mol. Med.*, **86**, 1205–1219.
- Simons,R.W., Houman,F. and Kleckner,N. (1987) Improved single and multicopy lac-based cloning vectors for protein and operon fusions. *Gene*, **53**, 85–96.
- Boyd,A.C. and Sherratt,D.J. (1995) The pCLIP plasmids: versatile cloning vectors based on the bacteriophage lambda origin of replication. *Gene*, **153**, 57–62.
- Amann,E., Ochs,B. and Abel,K.J. (1988) Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene*, **69**, 301–315.
- Isalan,M., Klug,A. and Choo,Y. (2001) A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat. Biotechnol.*, **19**, 656–660.
- Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y., Miyajima,N., Hirotsawa,M., Sugiura,M., Sasamoto,S. *et al.* (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.*, **3**, 109–136.
- Kaneko,T., Nakamura,Y., Sasamoto,S., Watanabe,A., Kohara,M., Matsumoto,M., Shimpo,S., Yamada,M. and Tabata,S. (2003) Structural analysis of four large plasmids harboring in a unicellular cyanobacterium, *Synechocystis* sp. PCC 6803. *DNA Res.*, **10**, 221–228.
- Pavletich,N.P. and Pabo,C.O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science*, **252**, 809–817.
- Pabo,C.O., Peisach,E. and Grant,R.A. (2001) Design and selection of novel Cys2His2 zinc finger proteins. *Annu. Rev. Biochem.*, **70**, 313–340.
- Maeder,M.L., Thibodeau-Beganny,S., Osiak,A., Wright,D.A., Anthony,R.M., Eichinger,M., Jiang,T., Foley,J.E., Winfrey,R.J., Townsend,J.A. *et al.* (2008) Rapid ‘open-source’ engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell*, **31**, 294–301.
- Lohe,A.R. and Hartl,D.L. (1996) Autoregulation of mariner transposase activity by overproduction and dominant-negative complementation. *Mol. Biol. Evol.*, **13**, 549–555.
- Lampe,D.J., Grant,T.E. and Robertson,H.M. (1998) Factors affecting transposition of the Himar1 mariner transposon *in vitro*. *Genetics*, **149**, 179–187.
- Suckow,J., Markiewicz,P., Kleina,L.G., Miller,J., Kisters-Woike,B. and Muller-Hill,B. (1996) Genetic studies of the Lac repressor. XV: 4000 single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure. *J. Mol. Biol.*, **261**, 509–523.
- Richardson,J.M., Colloms,S.D., Finnegan,D.J. and Walkinshaw,M.D. (2009) Molecular architecture of the Mos1 paired-end complex: the structural basis of DNA transposition in a eukaryote. *Cell*, **138**, 1096–1108.
- Thomas,C.E., Ehrhardt,A. and Kay,M.A. (2003) Progress and problems with the use of viral vectors for gene therapy. *Nat. Rev. Genet.*, **4**, 346–358.
- Daniel,R. and Smith,J.A. (2008) Integration site selection by retroviral vectors: molecular mechanism and clinical consequences. *Hum. Gene Ther.*, **19**, 557–568.
- Tan,W., Zhu,K., Segal,D.J., Barbas,C.F. 3rd and Chow,S.A. (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.
- Bushman,F.D. (1994) Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. *Proc. Natl Acad. Sci. USA*, **91**, 9233–9237.
- Bushman,F.D. and Miller,M.D. (1997) Tethering human immunodeficiency virus type 1 preintegration complexes to target DNA promotes integration at nearby sites. *J. Virol.*, **71**, 458–464.
- Goulaouic,H. and Chow,S.A. (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.
- Katz,R.A., Merkel,G. and Skalka,A.M. (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.
- Tan,W., Dong,Z., Wilkinson,T.A., Barbas,C.F. III and Chow,S.A. (2006) Human immunodeficiency virus type 1 incorporated with fusion proteins consisting of integrase and the designed polydactyl zinc finger protein E2C can bias integration of viral DNA into a predetermined chromosomal region in human cells. *J. Virol.*, **80**, 1939–1948.
- Ciuffi,A., Diamond,T.L., Hwang,Y., Marshall,H.M. and Bushman,F.D. (2006) Modulating target site selection during human immunodeficiency virus DNA integration in

- vitro with an engineered tethering factor. *Hum. Gene Ther.*, **17**, 960–967.
37. Zhu,Y., Dai,J., Fuerst,P.G. and Voytas,D.F. (2003) Controlling integration specificity of a yeast retrotransposon. *Proc. Natl Acad. Sci. USA*, **100**, 5891–5895.
38. Ivics,Z. and Izsvak,Z. (2006) Transposons for gene therapy! *Curr. Gene Ther.*, **6**, 593–607.
39. Ivics,Z., Katzer,A., Stuwe,E.E., Fiedler,D., Knospel,S. and Izsvak,Z. (2007) Targeted Sleeping Beauty transposition in human cells. *Mol. Ther.*, **15**, 1137–1144.
40. Wilson,M.H., Kaminski,J.M. and George,A.L. Jr. (2005) Functional zinc finger/sleeping beauty transposase chimeras exhibit attenuated overproduction inhibition. *FEBS Lett.*, **579**, 6205–6209.
41. Yant,S.R., Huang,Y., Akache,B. and Kay,M.A. (2007) Site-directed transposon integration in human cells. *Nucleic Acids Res.*, **35**, e50.
42. 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.
43. Maragathavally,K.J., Kaminski,J.M. and Coates,C.J. (2006) Chimeric Mos1 and piggyBac transposases result in site-directed integration. *FASEB J.*, **20**, 1880–1882.
44. Abdel-Meguid,S.S., Grindley,N.D., Templeton,N.S. and Steitz,T.A. (1984) Cleavage of the site-specific recombination protein gamma delta resolvase: the smaller of two fragments binds DNA specifically. *Proc. Natl Acad. Sci. USA*, **81**, 2001–2005.
45. Rimphanitchayakit,V. and Grindley,N.D. (1990) Saturation mutagenesis of the DNA site bound by the small carboxy-terminal domain of gamma delta resolvase. *EMBO J.*, **9**, 719–725.
46. Beerli,R.R., Segal,D.J., Dreier,B. and Barbas,C.F. III (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.
47. Ni,J., Clark,K.J., Fahrenkrug,S.C. and Ekker,S.C. (2008) Transposon tools hopping in vertebrates. *Brief Funct. Genomic Proteomic*, **7**, 444–453.