Targeted gene insertion for molecular medicine

Katrin Voigt, Zsuzsanna Izsvák, and Zoltán Ivics

Final Draft

Targeted gene insertion for molecular medicine

Katrin Voigt[1], Zsuzsanna Izsvák[1,2], and Zoltán Ivics[1]

[1] Max Delbrück Center for Molecular Medicine, Robert-Rössle Strasse 10, 13092 Berlin, Germany
[2] Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences, 6726 Szeged, Hungary

---

ABSTRACT | Genomic insertion of a functional gene together with suitable transcriptional regulatory elements is often required for long-term therapeutical benefit in gene therapy for several genetic diseases. A variety of integrating vectors for gene delivery exist. Some of them exhibit random genomic integration, whereas others have integration preferences based on attributes of the targeted site, such as primary DNA sequence and physical structure of the DNA, or through tethering to certain DNA sequences by host-encoded cellular factors. Uncontrolled genomic insertion bears the risk of the transgene being silenced due to chromosomal position effects, and can lead to genotoxic effects due to mutagenesis of cellular genes. None of the vector systems currently used in either preclinical experiments or clinical trials displays sufficient preferences for target DNA sequences that would ensure appropriate and reliable expression of the transgene and simultaneously prevent hazardous side effects. We review in this paper the advantages and disadvantages of both viral and non-viral gene delivery technologies, discuss mechanisms of target site selection of integrating genetic elements (viruses and transposons), and suggest distinct molecular strategies for targeted gene delivery.

KEYWORDS | DNA-Binding; Zinc Finger; Transposon; Virus; Non-Viral Vectors; Recombinase.

Viral vectors and genotoxic effects of their genomic integration

About 23% of gene therapy clinical trials have used retroviral and lentiviral vectors based on the murine leukemia virus (MLV), the avian sarcoma-leukosis virus (ASLV), or the human immunodeficiency virus (HIV) (http://www.wiley.co.uk/genmed/clinical). Retroviral vectors are very efficient in gene delivery and in providing sustained expression of the transgene, but their use raises serious concerns with respect to safety (recombination in vivo), immunological complications [1] and insertional mutagenesis. MLV has been shown to have a strong tendency to insert into transcription start sites of genes [2], whereas HIV exhibits a bias toward insertions into transcription units but without bias to transcription start sites [3]. ASLV shows the weakest preference for insertion into active genes in this group, but still at a frequency higher than that of random integration [4]. Integration of the vector into a gene or its regulatory elements can knock out the gene, alter its spatio/temporal expression pattern, or lead to truncation of the gene product (Fig. 1). Such genotoxic effects can have devastating consequences for the cell and the whole organism, including the development of cancer [5].

Such unfortunate events were observed in clinical trials using an MLV-based vector for gene therapy for X-linked severe combined immunodeficiency (SCID-X1). Nine out of 11 patients could be cured upon ex vivo transfer of a gene construct encoding the γ chain of the common cytokine receptor (γc) into autologous CD34+ bone marrow cells [6]. However, several years after the gene therapy treatment, two patients developed T-cell leukemia. In both patients, development of the leukemia was due to insertion of the transgene close to the promoter region of the LIM domain only 2 (*LMO2*) gene [7], and deregulated cell proliferation driven by retrovirus enhancer activity on the *LMO2* promoter. Since then, the number of severe adverse events in this particular clinical trial has grown to four [8], and yet a new case has been reported in a separate SCID-X1 trial [9]. These incidents very drastically underscored the peril of insertional mutagenesis upon transgene integration. Despite these adverse events, it needs to be emphasized that a number of patients who received gammaretroviral vector-mediated gene therapy treatment profit from improvements or even cure of their disease. This includes successful gene therapy trials for adenosine deaminase deficiency-linked SCID [10].

Retroviral vectors, with the exception of HIV-based systems, require active cell division for transgene delivery to the nucleus. In contrast, adenovirus-based vectors are capable of infecting dividing as well as postmitotic cells. In postmitotic cells, adenovirus-based vectors persist in an episomal state within the host cell, thereby alleviating the problems associated with mutagenic chromosomal insertions. However, due to their episomal state, adenoviral vectors are eliminated from proliferating cells over time. Native adenoviruses have the ability to transfer genes to a range of cell types. Capsid modification can alter the tropism of the virus, allowing infection of different, defined tissue targets [11].

One substantial problem with the use of adenovirus-based gene therapy vectors is their immunogenicity. Jesse Gelsinger, a patient suffering from partial ornithine transcarbamylase deficiency, was the first person to die from the experimental technique of gene therapy. Soon after receiving adenovirus-based gene therapy, he developed acute respiratory distress syndrome, and died of multiple organ failure [12]. In first- and second-generation adenoviral vectors that are deficient in regions of the adenoviral genome that are transcribed during early stages of infection, residual expression from remaining adenoviral genes can trigger a cytotoxic T-lymphocyte (CTL) immune response toward infected cells, which can lead to elimination of transduced cells and thus expression of the therapeutic gene [13]. Adenoviral vectors also have the capacity to induce an adaptive humoral immune response against the vector capsid, which can lead to elimination of readministered vectors due to circulating neutralizing antibodies [14]. Furthermore, systemic delivery of adenoviral vectors can lead to activation of an innate immune response even against third-generation, or so-called gutless or helper-dependent, adenoviral vectors that are completely devoid of all viral genes [15]. The problems associated with episomal DNA also apply to adenovirus-based vectors: stability of the gutless vector particle in vivo, as well as a

decrease in the expression of the therapeutic transgene can be observed over time [16].

Adeno-associated virus (AAV) is a single-stranded DNA virus that depends on the protein machinery of a helper virus such as adenovirus or Herpes Simplex Virus (HSV) to enter its lytic cycle [17]. In the absence of helper virus, the rep proteins encoded by AAV catalyze chromosomal integration and formation of a provirus. AAV has limited capacity in terms of cargo size, and thus in recombinant AAV (rAAV) vectors the viral genes including *rep* are removed to make room for genes of interest and elements necessary for their expression. Thus, even though rAAV vectors are able to transduce a wide variety of cells, they lack chromosomal integration. As a result, transgene expression from rAAV predominantly results from episomal vector DNA, though some integration of viral episomes can occur, dragging along the problems described above [18].

Taken together, potential genotoxic effects elicited by integrating viral vector systems, immunological complications associated with virus readministration and loss of therapeutic transgene expression for episomal vectors give rise to serious problems bearing great risk for patients undergoing gene therapy. Targeted integration of the therapeutic gene to a "safe" site in the human genome would prevent possible hazards to the host cell and organism due to the problems mentioned above.

Nonviral vectors

Due to safety concerns regarding viral vectors, there is much interest in developing nonviral gene delivery technologies. Nonviral vector systems collectively cover physical methods, i.e., hydrodynamic pressure techniques, electroporation, ballistic delivery or microinjection and complexing of nucleic acids with cationic polymers such as lipofection, cationic peptides, polyethylenimine (PEI), or receptor-mediated delivery for introducing therapeutic nucleic acids into cells (reviewed in [19]).

The major challenge with any nonviral gene delivery method is to provide efficient entry into the cell, escape of the endosomal and lysosomal compartment and transport into the nucleus. Though generally showing lower immunogenicity and toxicity than viruses, easy and cost-effective production and no strict size limitation for the therapeutic cargo, nonviral vectors are difficult to deliver at a reasonable efficiency. Integration of DNA by homologous recombination, though a highly sequence-specific process, generally takes place at frequencies <0.1% [20], making it unsuitable for transgene delivery in clinical settings. However, frequencies of homologous recombination-based gene repair can be boosted by delivering a homologous DNA template by an AAV vector. Such an approach allows high-fidelity targeted gene repair at frequencies of up to 1% in human fibroblasts [21].

In the absence of genomic integration, DNA introduced into a cell can persist in an episomal state. Episomal DNA is eventually lost in dividing cells, and expression often decreases in quiescent tissues over time, probably due to transgene silencing. Exclusion of bacterial sequences and inclusion of certain elements of the Epstein–Barr Virus (EBV) such as nuclear antigen I and human scaffold or matrix attachment regions (S/MARs) can prolong persistence of the episome, and thereby ensure effective expression levels [22, 23]. However, EBV nuclear antigen I has been suggested to contribute to the oncogenic potential of EBV [24] and shown to inhibit apoptosis independently from other viral genes [25], thereby presenting a risk of cellular transformation and the development of cancers. An alternative is to construct and apply human artificial chromosomes (HACs) [26]. However, experimental manipulation and full characterization of HACs may prove difficult due to their large size and their content of sometimes highly repetitive sequences.

Transposons as integrating, nonviral gene delivery vectors

Transposable elements represent nonviral vector systems that possess the capacity to stably integrate into the genome, and thus provide long-lasting expression of transgene constructs in cells. The synthetic fish and amphibian transposons *Sleeping Beauty* (SB) [27] and *Frog Prince* [28], respectively, are members of the Tc1/*mariner* superfamily belonging to so-called DNA transposons that transpose via a cut-and-paste mechanism from one DNA molecule to another. These transposon systems are made up of two components: the transposon carrying a gene of interest and the transposase, the enzymatic factor of the transposition process (Fig. 2a). During transposition, the transposable element stably integrates into a recipient DNA molecule (Fig. 2b). Since, unlike viruses, transposons are not infectious, they have to be actively delivered into the cell.

Various methods for non-viral DNA delivery including hydrodynamic injection, electroporation, microinjection, and complexing of the transposon components with PEI have been tested in conjunction with transposable element vectors (reviewed in [29]). Alternatively, transposon vectors can be delivered into cells by coupling the integration machinery of the transposable element to the cell infection machinery of a virus. Transposon-virus hybrid vectors delivering the components of the SB transposon system into cells by infection of adenovirus [30] or herpes simplex virus [31] have been developed.

Sleeping Beauty is the most thoroughly studied vertebrate transposon to date, and it has shown highly efficient transposition in different somatic tissues of a wide range of vertebrate species including humans, as well as in the germline of fish, frogs, mice, and rats (reviewed in [32]). SB has been shown to provide long-term transgene expression in preclinical animal models (see [29] for a recent review). SB inserts into TA dinucleotides, and shows additional target site preferences based on physical properties of the DNA rather than on primary DNA sequence [33, 34]. A slight bias toward integration into genes and their upstream regulatory sequences can be observed [35]; this tendency, however, is not as pronounced as seen for viral vectors, and no insertion preference was seen for transcribed genes. The safety profile of SB transposon-based vectors is further improved by recent findings that they are fairly inert in their transcriptional activities, and that insulator elements can successfully be incorporated in the next generation of transposon vectors [36].

In this context, a clear distinction between an SB vector used for gene therapy and an SB vector engineered for the purpose of somatic, gain-of-function mutagenesis has

to be made. The SB vectors used in genetic screens in the mouse for oncogene discovery were specifically developed to contain strong, viral enhancers and splice donor signals to purposefully overexpress genes that happen to be located near the transposon insertion sites [37]. This is clearly not the case in a typical SB vector used for gene therapeutic purposes that would carry a therapeutic expression cassette (Fig. 2a). Indeed, it is important to note that no dominant adverse effects associated with SB vector integration have been observed in experimental animals, not even in a cancer-predisposed genetic background [38].

The *piggyBac* (PB) element, a DNA transposon isolated from the cabbage looper moth, also has a potential as a vector in gene therapy [39]. PB has shown transpositional activity in mouse and human cells. PB is able to mobilize a cargo of up to 14 kb without loss of efficiency [40]; it integrates into TTAA sequences, preferentially in introns of transcriptionally active regions [39].

The *Tol2* element from medaka fish is the only known, naturally occurring, active DNA transposon of vertebrate origin [41]. Its cargo capacity covers at least 10 kb; it has been widely used as a tool for transgenesis in zebrafish [42] and was shown to be active in a preclinical experimental setting in the mouse liver [43]. So far, neither sequence- nor DNA structure-specific insertion preference has been observed for *Tol2*. Its potential as a vector in gene therapy remains to be further investigated.

Naturally occuring specificity in target site selection of integrating genetic elements

Site-selectivity in viral integration

As discussed above, most viral vectors show an integration bias toward transcriptionally active regions in the genome. Because no sequence-specific integration preference of the retroviral/lentiviral integrase (IN) protein itself has been observed, biased genomic integration can be due to the interaction of the viral components with certain host proteins or recognition of different chromatin states of the chromosomes during integration [4]. For example, in contrast to MLV, the integration pattern of HIV does not correspond to the genomic distribution of DNaseI hypersensitivity sites that are associated with open chromatin found in regions upstream of genes and in active transcription units [44]. Instead, the bias of HIV toward integration into active cellular transcription units was proposed to be due to tethering interactions with cellular proteins rather than to chromatin accessibility. In particular, the cellular lens epithelium-derived growth factor (LEDGF)/p75 was shown to influence HIV target site selection [45]. LEDGF/p75 acts as a transcriptional co-activator, and interacts with components of the basal transcription machinery [46]. LEDGF/p75 binds tightly to HIV IN and drives IN into the nucleus when both proteins are produced at high levels [47]. LEDGF/p75 is conserved among vertebrate species, indicating that insertion site selection of HIV could be maintained among vertebrates [48]. Cells in which LEDGF/p75 expression is knocked down to <10% by RNAi are still capable of production of infectious HIV, indicating that LEDGF/p75 is dispensable for virus replication [45, 47], but showed reduced integration into transcribed units as compared to normal control cells.

AAV shows strict sequence specificity for integration (Table 1). In the absence of a helper virus, two of the four rep proteins termed rep78 and rep68 encoded by AAV catalyze integration at a single locus named *AAVS1* on human chromosome 19. The exact mechanism of site-specific integration of AAV is still unknown. The viral components involved in targeted DNA integration include the inverted terminal repeats (ITRs) and either the rep68 or the rep78 protein. The ITR spans the terminal 145 nt of the AAV genome and contains a rep-binding element (RBE) and a terminal resolution site (trs). An RBE and a trs-like site can also be found in the *AAVS1* locus in the human genome, and this region is required for site-specific integration of AAV into the human genome [49]. A replicative recombination mechanism has been suggested for site-specific integration of AAV. By binding to both the genomic as well as viral DNA, rep68/rep78 brings the viral genome to close proximity to the *AAVS1* locus [50]. Rep68/rep78 bound to the RBE at *AAVS1* introduces a nick at the trs, and initiates unidirectional DNA synthesis [51]. Rep68 bound to the RBE in the AAV genome also introduces a nick at the viral trs, and viral DNA is integrated into the *AAVS1* locus by template strand switches during unidirectional DNA synthesis [49].

Combining favorable traits of two vector systems could result in a powerful hybrid vector. A hybrid vector composed of a HSV amplicon and AAV components for site-specific integration showed genomic integrations in 10–30% of infected cells, 50% of which occurred at the *AAVS1* locus [52]. Using a hybrid vector combining the integration machinery of AAV with an adenoviral vector, site-specific integration frequencies of up to 2% were accomplished in a transgenic mouse model [53]. However, persistence of the rep protein leads to chromosomal instability and to mobilization of the transgene [54], greatly undesired effects in gene therapy.

Site-specific recombinases

Sequence-specific DNA integration is also mediated by some recombinases (Table 1). Two groups of recombinases can be distinguished: the serine and tyrosine recombinases that differ in the mechanisms by which they catalyze recombination. The structural domains of serine recombinases are often spatially separated as opposed to tyrosine recombinases whose domains are interwoven. Cre is a type I topoisomerase from bacteriophage P1 that mediates recombination of DNA between *loxP* sites. Cre has been shown to be active in eukaryotic, including human, cells and is widely used for genome engineering in mice [55]. DNA flanked by *loxP* sites in a direct orientation will be excised and integrated into a *loxP* site previously placed into the human genome [56]. Recombination at pseudo *loxP* sites (endogenous human DNA sequences that show similarity to *loxP*) in the human genome occurs with a fourfold lower efficiency than for wild-type *loxP* sites [57]. A directed evolution approach was employed to create a new site-specific Cre recombinase. The newly created recombinase, termed Tre, recombines sequences in the LTRs of integrated HIV proviruses, resulting in excision of the HIV provirus from genomic DNA [58].

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

mouse embryonic stem cells [59]. Both Cre and FLP are bidirectional recombinases that catalyze DNA excision and integration, but favoring the excision reaction. This feature leads to inefficient integration and expression of transgene constructs. Furthermore, genotoxic effects including chromosomal rearrangements and growth inhibition observed for Cre recombinase when expressed persistently at high levels make it a possible hazard to genome integrity [60].

The same holds true for a site-specific IN from the *Streptomyces* phage ΦC31 [61] that catalyzes recombination between so-called attachment (*att*) sites. The *attP* site is found in the ΦC31 genome, whereas *attB* is located in the host *Streptomyces* genome. ΦC31-mediated integration in human as well as mouse cells frequently occurs into pseudo *att* sites such as *psA* in human or *mpsA* in the mouse genome [62, 63]. *PsA* shares 44% identity with *attP* [64]. In human 293 cells harboring an inserted *attP* site, 15% of the integrations were detected at *attP*, 5% of the rest of the integration events occurred at *psA*, 5–10% were random, whereas the rest of integrations was believed to be distributed over the other ~100 pseudo sites in the human genome [63].
In several studies, reasonably efficient delivery and stable expression of genes relevant in human genetic diseases [65] was achieved in mouse or human cells using ΦC31 recombinase. However, ΦC31 is mutagenic, because it can cause chromosomal aberrations due to recombination between pseudo sites or imperfect recombination reactions [66, 67]. It remains to be tested if insertions of transgenes at pseudo sites in the human genome can cause alterations of host gene expression patterns leading to abnormal cell behavior.

Site-specific transposable elements

Unlike viruses, transposons do not possess envelope genes and hence lack an extracellular phase in their life cycle. This makes their fate closely linked to the fate of the host cell, and may result in integration patterns less mutagenic to the cell. The higher the gene density of a genome, the higher the chance for transposable elements to insert into coding sequences, resulting in potentially fatal consequences to the cell. Significant fractions of genomes with a small proportion of coding regions and extensive intergenic regions can be composed of transposon-derived sequences (e.g., 45% of the human genome), in contrast to organisms having a small genome with high gene density, such as yeast. Ty retrotransposons in *Saccharomyces cerevisiae* are structurally and functionally related to retroviruses. Integration of Ty1, Ty3, and Ty5 retrotransposons is tethered to certain sites in the genome by host proteins (Table 1).

The Ty1 element shows a strong insertion preference for genes transcribed by RNA polymerase III (Pol III). Ninety percent of Ty1 insertions can be found about 1 kb upstream of transfer RNA (tRNA) genes [68]. A second preferred integration area of Ty1 is found upstream of the 5S RNA genes that are also transcribed by Pol III [69]. Targeting of this site by Ty1 elements may thus depend on the same factors as targeting of the tRNA genes. Indeed, components of the Pol III transcription machinery were found to be required for targeting of Ty1 [70]; however, other factors such as chromatin components, physical properties of DNA, or subnuclear localization of the target may as well specify integration sites.

Ty3 integrates one or two base pairs upstream of Pol III transcription start sites. TFIIIB and TFIIIC are important factors for assembly of Pol III complexes at transcription start sites of Pol III-transcribed genes, and are also involved in the recruitment of Ty3 [71]. Though TFIIIB is sufficient to target Ty3, TFIIIC orientates binding of TFIIIB to the TATA box [72], and weakly interacts with Ty3 IN [73]. The Ty5 element interacts with the host protein Sir4p [74], which targets insertions to heterochromatic regions of the genome such as telomers and silent mating locus [75]. Interaction of Ty5 IN with Sir4p is mediated by its targeting domain, a 6-amino-acid motif at the C terminus of Ty5 IN. Mutations within this domain abolish interaction between IN and Sir4p and result in random integration of Ty5 retrotransposons. Concordantly, random integration of Ty5 is observed in cells deficient in Sir4p [74].

Targeting of a specific genomic site may be specified by primary DNA sequence recognized by specific DNA-binding domains (DBDs). In addition, physical properties of the DNA such as kinks due to protein binding, triplex DNA or altered/abnormal DNA structures due to base composition may cause preferential binding of proteins or protein complexes at certain sites. For the bacterial transposon Tn7, both sequence- and structure-specific binding apply. The Tn7 transposon encodes five different proteins: TnsA, B, C, D, and E. Depending on proteins involved in the transposition process, either a particular DNA structure found during conjugation or a specific site in the bacterial genome is targeted [76].

During bacterial conjugation, TnsE seems to recognize DNA structures with recessed 3′-ends during lagging strand DNA synthesis, and directs integration of the transposon to this site. TnsD binds to a specific DNA sequence called *attTn7* in the 3′-end of the bacterial glutamine synthetase (*glmS*) gene in the bacterial genome, followed by insertion of the transposon several base pairs downstream of *glmS* (Table 1). Binding of TnsD creates DNA distortion probably responsible for recruitment of TnsC, which in turn interacts with TnsAB promoting insertion of Tn7 at *attTn7*. Importantly, Tn7 inserts into the human homologue of *glmS* in *E. coli* and test tube reactions [77], but Tn7 transpositional activity in human cells has not been reported.

The eukaryotic microorganism *Dictyostelium discoideum* has a highly compact genome of 34 Mb with 76% coding regions and a surprisingly high transposon load of 10%. Transposons in *D. discoideum* have developed two strategies to avoid genotoxic insertion into coding sequences (Table 1). One of these strategies is nested integrations of transposons forming clusters. For example, the DIRS LTR-retrotransposon family shows no initial target site selectivity, but can be found in few clusters, made up of several copies of themselves [78], located in centromeric and telomeric regions of chromosomes. The other strategy is targeted integration into "safe" regions of the genome free from protein-coding sequences. This strategy is primarily used by non-LTR retrotransposons that insert up- and downstream of tRNA genes [79]. The non-LTR retrotransposons collectively called TRE (tRNA gene-targeting retrotransposable elements) can be divided into two groups: TRE5 elements preferentially integrate about 50 bp upstream of tRNA genes, whereas TRE3 elements favor the integration of 100–150 bp downstream to tRNA genes. An in vivo assay using a reporter gene tagged with a tRNA coding region showed targeted integration of TRE5 in the same manner as in a genomic context, indicating that targeted insertion of

TRE5 is dependent on interactions with Pol III transcription factors [80]. Indeed, the ORF1 protein encoded by the TRE5 element was recently shown to interact with TFIIIB, suggesting a role of this interaction in targeting integration into tRNA genes [81].

In sum, the existence of transposable elements with natural targeting abilities raises promise that recombinase/transposase/IN proteins with target-selective insertion properties can be engineered.

Artificial (imposed) targeting of DANN integration into preselected sequences

None of the vector systems currently used either in preclinical experiments or in clinical trials described above displays DNA sequence preferences specific enough for targeted insertion into a defined location in the human genome. Integration into selected sites in the genome would simultaneously ensure appropriate expression of the transgene (lack of position effects), and prevent hazardous effects to the organism due to insertional mutagenesis of cellular genes (lack of genotoxicity). Targeted gene delivery can rely on distinct molecular strategies. One possibility implies fusion of the recombinase/transposase/IN to a DNA binding domain. Upon binding of the engineered recombinase to a specific target site, integration of the DNA component of the vector system may occur in adjacent regions (Fig. 3a). A more indirect approach uses DNA-binding specificity of interacting proteins. Interaction of proteins bound to specific target sequences can tether either the DNA (Fig. 3b) or the protein (Fig. 3c) component of the vector system to this region of DNA, resulting in integration into nearby regions.

Targeting through fusion to DNA-binding domains

Altering sequence-specificity of most recombinases may prove difficult since they do not have spatially separated catalytic and target DBDs that could be modularly replaced irrespectively of each other. Target specificity can potentially be altered by directed evolution (random mutagenesis techniques followed by activity screening under selective conditions) or by substitution of key amino acids implicated in target recognition. Both approaches yielded mutants of proteins showing more relaxed target-site specificity or even a complete shift in target site preference (reviewed in [82]). Engineering of proteins that specifically bind to desired DNA sequences is expected to pose a major challenge, and may not only lead to altered site specificity, but also to impaired or modified catalytic activity. Fusions of proteins to a specific DBD appear to be a much easier and more direct approach (Table 2).

However, some proteins display sensitivity to fusions with foreign peptides, domains, or proteins, possibly due to altered folding of the resulting chimeric protein. Thus, fusions may result in abolished or limited enzymatic activity. Another factor to consider is that the native DNA-binding capacity of the protein can compete with the foreign DBD of the fusion partner. Requirements for integration of a vector system, such as a TA dinucleotide within an appropriate structural context for the SB transposon, should also be taken into account when selecting a site to be targeted in the genome. Keeping this in mind, fusions between a DBD and a recombinase

protein may overall be a promising approach to targeted gene insertion.

In vitro targeting studies of the IN of avian sarcoma virus (ASV) fused to the DNA-binding domain of the *E. coli* LexA protein showed altered insertion patterns and an insertion hot spot near a tandem LexA operator as compared to unfused IN [83]. HIV IN fusions to the DBD of phage λ repressor protein [84] or to the DBD of the LexA repressor protein [85] were also capable of targeting integrations near their specific binding sites in vitro. These experiments demonstrated the feasibility of using fusions between DBDs and INs to target viral insertions to a certain extent to specific sites.

Transcription factors (TFs) recognize and bind specific DNA sequences followed by recruitment of proteins affecting the transcriptional status of the associated gene. These processes are usually mediated by distinct domains, making it possible to separate these functions. Consequently, the DBD of a TF by itself would preserve its unrestrained DNA-binding capacity (specificity and affinity), serving as a potent source as a fusion partner. TFs are typically classified according to the structure of their DBDs, such as zinc finger (ZF), leucine zipper, helix-turn-helix, helix-loop-helix, and high-mobility group boxes. One naturally occurring ZF is the DBD of transcription factor Gli1 present in vertebrates that recognizes and binds a 9-bp DNA sequence. The bacterial insertion sequence element IS30 was fused to either the cl repressor of phage λ or the Gli1 DBD, and the resulting fusion proteins showed targeted integration into plasmid targets in *E. coli* and zebrafish [86]. This study was the first demonstration that targeted transposition by an engineered transposase could work in vivo.

The DBD of the yeast Gal4 TF contains a ZF domain of the Zn_2Cys_6 type. It recognizes a specific, 17-bp DNA sequence called upstream activating sequence (UAS). Fusions of the Gal4 DBD to the Mos1 (a Tc1/*mariner* transposon from *Drosophila mauritiana*) and PB transposases were tested for their transpositional activities and targeting potentials by applying plasmid-based transposition assays in mosquito embryos [87]. Transposition mediated by the chimeric Mos1 transposase into the UAS-containing target plasmid occurred at a 96% frequency at the same TA located 954 bp away from the targeted UAS sequence. Transposition by the Gal4-PB fusion protein into a plasmid containing the UAS target sequence occurred at a 67% frequency into a TTAA site located 1,103 bp upstream of the UAS.

These results present quite efficient targeting by Mos1- and PB-Gal4 fusions. Binding of the Gal4 DBD to its recognition site presumably brings the fused transposase to close proximity, thereby enhancing the chance of transposon insertions nearby. Chimeric transposases may structurally be limited after UAS binding, allowing transgene integration into only few sites.

An independent study examined the transpositional activities of three different transposase proteins after fusion to Gal4 in cultured human cells [88]. Fusions completely abolished transpositional activity of *Tol2* and SB11 (an early-generation hyperactive mutant of SB), whereas only a slight decrease in activity was observed for Gal4-PB when compared to unfused PB transposase. Targeted transposition by the fusion transposases was not investigated in this study. However, another group reported that only N-terminal fusions to the SB

transposase retained transpositional activity, and that fusion of the Gal4 DBD to HSB5 (a third-generation improved SB transposase) resulted in a drop in transposition efficiency to ~26% of unfused HSB5 [89]. This fusion transposase showed targeted transposon integration in a plasmid-based assay in cultured human cells. Targeted transposition events were enriched about 11-fold in a 443-bp window around a 5-mer UAS site in the target plasmid, as compared with integration patterns mediated by unfused transposase.

Naturally occurring ZFs also include the three-finger transcription factor Zif286 originally identified in the mouse. A chimeric recombinase composed of the DBD of Zif268 and the catalytic domain of the bacterial Tn3 resolvase was successfully assayed for targeting of two inverted Zif268 recognition sites flanking a Tn3 *res* site in *E. coli* [90]. Tn3 belongs to the serine recombinases that have spatially separated catalytic and DNA-binding domains. Functionality of the chimeric protein proves that exchange of the physiological DBD of Tn3 resolvase with a foreign DBD yields a recombinationally competent enzyme. It remains to be investigated whether such a fusion construct is also functional in eukaryotic cells. Zif268 fusions with the HIV IN were also shown to have biased insertion patterns near specific binding sites in vitro [91].

Naturally occurring DBDs have some limitations for use as gene targeting agents. First, some of the DBDs discussed above are derived from proteins that do not have physiological targets in the human genome; thus, specific target sites would need to be introduced into the genome prior to delivery of a transgene. Second, those DBDs that do have physiological binding sites in the human genome recognize short DNA sequences present in multiple copies throughout the human genome, making targeted insertion with these DBDs impractical (for example, a 9-bp recognition sequence of a ZF would be expected to occur >10,000 times in the human genome). Recognition sites of 18 bp would be expected to be unique in the human genome.

Artificial ZFs, especially the C_2H_2 type, offer a potential solution. Their modular character in structure and function is the key advantage in engineering of proteins that are able to recognize theoretically any sequence in the human genome [92]. Each individual zinc finger binds 3–4 bp DNA, thus a set of 64 domains would cover recognition of any desired DNA sequence. ZF nucleases (ZFNs) consisting of the FokI cleavage domain fused to a ZF represent an attractive technology for targeted gene repair by homologous recombination. Two ZFNs need to heterodimerize in order to cleave DNA at the target site. So far, the use of ZFNs has exhibited cytotoxic effects on cells probably resulting from off-target DNA cleavage. Recent work, however, shows reduction of cytotoxic effects after redesign of the dimerization interface of the nucleases [93]. In combination with integrase-defective lentiviral vectors as a delivery tool, high levels of gene repair and gene addition into a variety of human cells were recently accomplished [94].

Fusions of engineered ZFs to recombinase proteins could enable selective insertion of a transgene into a desired region of the genome. The synthetic E2C ZF protein is a six-finger ZF recognizing an 18-bp target site in the 5′-untranslated region of the human *erbB-2* gene. E2C fusions to transcriptional activator and repressor domains have been used to regulate endogenous *erbB-2* gene

expression [95]. Fusions of E2C to HIV IN were shown to target retroviral integration near the 18-bp E2C binding site in cell-free reactions [96]. The E2C/IN fusion protein was then tested for targeting of the E2C locus in cultured human cells using a quantitative real-time PCR assay showing an approximately tenfold increase of insertions near the E2C binding site in the genome as compared to unfused IN. However, virions containing the fusion proteins exhibited poor infectivity ranging from 1 to 24% compared to viruses containing wild-type IN [97].

A recent publication reported the generation of fusion proteins of E2C and the HSB5 hyperactive SB transposase [89]. As seen before [98], fusion proteins showed reduced transpositional activity as compared to unfused transposase, but about 20% transposition activity could be rescued by applying a glycine/serine linker between the ZF and transposase domains and by using a human codon-optimized E2C gene. This optimized fusion protein showed targeted transposon integration in a plasmid-based assay in cultured human cells. Targeted transposition events were enriched about eightfold in a 443-bp window around a 5-mer repeat of the E2C binding site in the target plasmid, as compared with integration patterns mediated by unfused HSB5. However, cell-based assays failed to detect targeting of the E2C binding site in a genomic context.

Similarly, the artificial three-finger protein Jazz, binding to a 9-bp sequence in the promoter region of the human utrophin gene [99], was fused to the SB transposase. The fusion protein retained about 15% transpositional activity when compared to wild-type transposase, but targeted transposition events on the genome level could not be identified [100]. One possibility to explain failure of targeting in a genomic context could be physical constraints on the transposase upon site-specific binding in that the tranposase is unable to interact with a TA dinucleotide to integrate the transposon. This may especially hold true for GC-rich DNA sequences at the *erbB-2* promoter region.

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

Targeting through interaction with DNA-binding proteins

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

domain to the DBD of phage λ repressor protein [101]. In an analogous fashion, Sir4p (which, as described above, mediates targeted insertion of the yeast retrotransposon into heterochromatin in yeast) fused to the *E. coli* LexA DBD was shown to result in integration hot spots for Ty5 near LexA operators [102]. Domain swaps in recombinase proteins by changing protein–protein interaction domains could also lead to modified integration patterns. Indeed, replacing the targeting domain of Ty5 IN, which interacts with Sir4p, with a heterologous domain interacting with a protein fused to LexA, also leads to insertions near the LexA operators [102] (Table 3).

Different approaches to targeting were taken in work involving the SB transposon system [100]. The principle of these approaches was to bring either component of the SB transposon system (transposon DNA or transposase protein) in close proximity to a specific site in a human cell environment (Table 3). Components of a first approach were a LexA operator site incorporated into an SB transposon, a fusion protein consisting of LexA and a SAF-box, and unmodified SB transposase. The SAF-box is a domain first identified in the human scaffold attachment factor (SAF-A) that specifically binds to scaffold/matrix attachment regions (S/MARs) [103]. S/MAR elements are bound to the nuclear matrix, thereby structuring chromosomal DNA by forming chromatin loops. Transgenes flanked by S/MARs have shown expression independent from their site of integration. Therefore, a possible way to minimize silencing effects on transgene expression could be the insertion of a transgene into S/MARs. For targeted transposition into S/MARs to occur, the LexA-SAF-box fusion protein was expected to bind the LexA operator-containing transposon. This protein–DNA complex would then be tethered to S/MAR regions of chromosomes through SAF-box binding, whereas transposition into linked sites would occur upon recruitment of SB transposase (Fig. 3b). An increase in transposon insertions within a 1-kb range of genomic S/MAR sequences was observed as compared to controls with fusion proteins lacking the SAF-box. In this study, targeting by a protein with highly specific DNA-binding properties, the tetracycline repressor (TetR), was also sought. A transgenic HeLa cell line incorporating a tetracycline response element (TRE)-driven EGFP gene as a targeted locus was created. In this experiment, a targeting fusion protein consisting of TetR and LexA was applied. Integrations upstream of the EGFP gene were determined, yielding insertions into two TA sites within the EGFP promoter region 44 and 48 bp downstream of the TRE region. No insertions into this region were detected with transposons lacking the LexA operator sequence, suggesting that interaction between the targeting protein and the transposon DNA is indeed required for targeted transposition events.

As shown for HIV IN and LEDGF/p75, protein–protein interactions can tether integration complexes to certain regions of the genome, suggesting that such a mechanism can be adapted for targeted transposon insertion as well (Fig. 3c and Table 3). Accordingly, a previously identified protein–protein interaction domain of the SB transposase was built into an experimental setup aiming at targeted transposition in human cells. This domain spans the N-terminal helix-turn-helix domain (termed N57 for containing 57 amino acids) of the SB transposase [104]. Importantly, coexpression of N57 together with full-length transposase had no dominant negative effect on transposition [104]. Targeted transposition into the chromosomal TRE-EGFP region using a TetR-N57 fusion was monitored in human cells [100]. On average, >10% of cells undergoing transposition were found to contain targeted events within the TRE-EGFP locus. Insertions obtained by this strategy occurred at multiple sites within a 2.5-kb window and featured some insertion hot spots.

An overall advantage of applying technologies based on protein–DNA and/or protein–protein interactions for the manipulation of target site selection of transposases is that the transposase does not need to be modified, thereby eliminating the decrease in transpositional activity associated with direct fusions.

Conclusion

As discussed in this review, there are several factors affecting site-selectivity of integrating vector systems. These include accessibility of specific chromosomal sites by chromatin components, primary sequence, and physical structure of the DNA at the targeted region, endogenous expression of proteins that may compete for binding, and the specificity as well as capacity of chimeric proteins in DNA-binding as well as in catalytic functions. Both naturally targeted recombinase systems (such as ΦC31) and targeting systems engineered from promiscuously integrating vectors (such as *Sleeping Beauty*) show off-target effects in the context of the human genome. For the former, the capacity of the recombinase to act at endogenous pseudo sites can lead to genomic rearrangements. For the latter, despite the fact that targeted integrations can be generated, non-targeted insertions can still occur at high frequencies because the natural DNA-binding capacities of the transposase competes with that of the foreign DBD used for targeting. Keeping such off-target effects at a minimum remains a major challenge. Although several hurdles are yet to be overcome before technologies of targeted gene insertion can be considered for applications, recent evidence suggests that target-selected transgene insertion into desired regions in the human genome is a realistic goal.

Corresponding Author

Zoltán Ivics, zivics@mdc-berlin.de

References

1. Follenzi A, Santambrogio L, Annoni A (2007) Immune responses to lentiviral vectors. Curr Gene Ther 7:306–315.

2. Wu X, Li Y, Crise B, Burgess SM (2003) Transcription start regions in the human genome are favored targets for MLV integration. Science 300:1749–1751.

3. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F (2002) HIV-1 integration in the human genome favors active genes and local hotspots. Cell 110:521–529.

4. Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC, Ecker JR, Bushman FD (2004) Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol 2:E234.

5. Baum C, von Kalle C, Staal FJ, Li Z, Fehse B, Schmidt M, Weerkamp F, Karlsson S, Wagemaker G, Williams DA (2004) Chance or necessity? Insertional mutagenesis in gene therapy and its consequences. Mol Ther 9:5–13.

6. Hacein-Bey-Abina S, Le Deist F, Carlier F, Bouneaud C, Hue C, De Villartay JP, Thrasher AJ, Wulffraat N, Sorensen R, Dupuis-Girod S et al (2002) Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy. N Engl J Med 346:1185–1193.

7. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Lebouich P, Lim A, Osborne CS, 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.

8. Baum C (2007) What are the consequences of the fourth case? Mol Ther 15:1401–1402.

9. Thrasher AJ, Gaspar HB (2007) Severe adverse event in clinical trial of gene therapy for X-SCID. ASGT press release.

10. Aiuti A, Slavin S, Aker M, Ficara F, Deola S, Mortellaro A, Morecki S, Andolfi G, Tabucchi A, Carlucci F et al (2002) Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. Science 296:2410–2413.

11. Volpers C, Kochanek S (2004) Adenoviral vectors for gene transfer and therapy. J Gene Med 6(Suppl 1):S164–S171.

12. Raper SE, Chirmule N, Lee FS, Wivel NA, Bagg A, Gao GP, Wilson JM, Batshaw ML (2003) Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. Mol Genet Metab 80:148–158.

13. Christ M, Lusky M, Stoeckel F, Dreyer D, Dieterle A, Michou AI, Pavirani A, Mehtali M (1997) Gene therapy with recombinant adenovirus vectors: evaluation of the host immune response. Immunol Lett 57:19–25.

14. Dai Y, Schwarz EM, Gu D, Zhang WW, Sarvetnick N, Verma IM (1995) Cellular and humoral immune responses to adenoviral vectors containing factor IX gene: tolerization of factor IX and vector antigens allows for long-term expression. Proc Natl Acad Sci USA 92:1401–1405.

15. Muruve DA, Cotter MJ, Zaiss AK, White LR, Liu Q, Chan T, Clark SA, Ross PJ, Meulenbroek RA, Maelandsmo GM et al (2004) Helper-dependent adenovirus vectors elicit intact innate but attenuated adaptive host immune responses in vivo. J Virol 78:5966–5972.

16. Ehrhardt A, Xu H, Kay MA (2003) Episomal persistence of recombinant adenoviral vector genomes during the cell cycle in vivo. J Virol 77:7689–7695.

17. McCarty DM, Young SM Jr, Samulski RJ (2004) Integration of adeno-associated virus (AAV) and recombinant AAV vectors. Annu Rev Genet 38:819–845.

18. Nakai H, Montini E, Fuess S, Storm TA, Grompe M, Kay MA (2003) AAV serotype 2 vectors preferentially integrate into active genes in mice. Nat Genet 34:297–302.

19. Li S, Ma Z (2001) Nonviral gene therapy. Curr Gene Ther 1:201–126.

20. Goncz KK, Prokopishyn NL, Chow BL, Davis BR, Gruenert DC (2002) Application of SFHR to gene therapy of monogenic disorders. Gene Ther 9:691–694.

21. Russell DW, Hirata RK (1998) Human gene targeting by viral vectors. Nat Genet 18:325–330.

22. Wu H, Ceccarelli DF, Frappier L (2000) The DNA segregation mechanism of Epstein–Barr virus nuclear antigen 1. EMBO Rep 1:140–144.

23. Piechaczek C, Fetzer C, Baiker A, Bode J, Lipps HJ (1999) A vector based on the SV40 origin of replication and chromosomal S/MARs replicates episomally in CHO cells. Nucleic Acids Res 27:426–428.

24. Kaul R, Murakami M, Choudhuri T, Robertson ES (2007) Epstein–Barr virus latent nuclear antigens can induce metastasis in a nude mouse model. J Virol 81:10352–10361.

25. Hammerschmidt W, Sugden B (2004) Epstein–Barr virus sustains Burkitt's lymphomas and Hodgkin's disease. Trends Mol Med 10:331–336.

26. Hadlaczky G (2001) Satellite DNA-based artificial chromosomes for use in gene therapy. Curr Opin Mol Ther 3:125–132.

27. Ivics Z, Hackett PB, Plasterk RH, Izsvak Z (1997) Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell 91:501–510.

28. Miskey C, Izsvak Z, Plasterk RH, Ivics Z (2003) The Frog Prince: a reconstructed transposon from Rana pipiens with high transpositional activity in vertebrate cells. Nucleic Acids Res 31:6873–6881.

29. Ivics Z, Izsvak Z (2006) Transposons for gene therapy! Curr Gene Ther 6:593–607.

30. Yant SR, Ehrhardt A, Mikkelsen JG, Meuse L, Pham T, Kay MA (2002) Transposition from a gutless adeno-transposon vector stabilizes transgene expression in vivo. Nat Biotechnol 20:999–1005.

31. Bowers WJ, Mastrangelo MA, Howard DF, Southerland HA, Maguire-Zeiss KA, Federoff HJ (2006) Neuronal precursor-restricted transduction via in utero CNS gene delivery of a novel bipartite HSV amplicon/transposase hybrid vector. Mol Ther 13:580–588.

32. Mates L, Izsvak Z, Ivics Z (2007) Technology transfer from worms and flies to vertebrates: transposition-based genome manipulations and their future perspectives. Genome Biol 8(Suppl 1):S1.

33. Vigdal TJ, Kaufman CD, Izsvák Z, Voytas DF, Ivics Z (2002) Common physical properties of DNA affecting target site selection of Sleeping Beauty and other Tc1/mariner transposable elements. J Mol Biol 323:441–452.

34. Liu G, Geurts AM, Yae K, Srinivasan AR, Fahrenkrug SC, Largaespada DA, Takeda J, Horie K, Olson WK, Hackett PB (2005) Target-site preferences of Sleeping Beauty transposons. J Mol Biol 346:161–173.

35. Yant SR, Wu X, Huang Y, Garrison B, Burgess SM, Kay MA (2005) High-resolution genome-wide mapping of transposon integration in mammals. Mol Cell Biol 25:2085–2094.

36. Walisko O, Schorn A, Rolfs F, Devaraj A, Miskey C, Izsvak Z, Ivics Z (2008) Transcriptional activities of the Sleeping Beauty transposon and shielding its genetic cargo with insulators. Mol Ther 16:359–369.

37. Dupuy AJ, Jenkins NA, Copeland NG (2006) Sleeping beauty: a novel cancer gene discovery tool. Hum Mol Genet 15(Spec No 1):R75–R79.

38. Carlson CM, Frandsen JL, Kirchhof N, McIvor RS, Largaespada DA (2005) Somatic integration of an oncogene-harboring Sleeping Beauty transposon models liver tumor development in the mouse. Proc Natl Acad Sci USA 102:17059–17064.

39. Wilson MH, Coates CJ, George AL Jr (2007) PiggyBac transposon-mediated gene transfer in human cells. Mol Ther 15:139–145.

40. Ding S, Wu X, Li G, Han M, Zhuang Y, Xu T (2005) Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. Cell 122:473–483.

41. Koga A, Suzuki M, Inagaki H, Bessho Y, Hori H (1996) Transposable element in fish. Nature 383:330.

42. Kawakami K (2007) Tol2: a versatile gene transfer vector in vertebrates. Genome Biol 8(Suppl 1):S7.

43. Balciunas D, Wangensteen KJ, Wilber A, Bell J, Geurts A, Sivasubbu S, Wang X, Hackett PB, Largaespada DA, McIvor RS et al (2006) Harnessing a high cargo-capacity transposon for genetic applications in vertebrates. PLoS Genet 2:e169.

44. Lewinski MK, Yamashita M, Emerman M, Ciuffi A, Marshall H, Crawford G, Collins F, Shinn P, Leipzig J, Hannenhalli S et al (2006) Retroviral DNA integration: viral and cellular determinants of target-site selection. PLoS Pathog 2:e60.

45. Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, Shinn P, Ecker JR, Bushman F (2005) A role for LEDGF/p75 in targeting HIV DNA integration. Nat Med 11:1287–1289.

46. Ge H, Si Y, Roeder RG (1998) Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation. EMBO J 17:6723–6729.

47. Llano M, Vanegas M, Fregoso O, Saenz D, Chung S, Peretz M, Poeschla EM (2004) LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. J Virol 78:9524–9537.

48. Barr SD, Leipzig J, Shinn P, Ecker JR, Bushman FD (2005) Integration targeting by avian sarcoma-leukosis virus and human immunodeficiency virus in the chicken genome. J Virol 79:12035–12044.

49. Linden RM, Winocour E, Berns KI (1996) The recombination signals for adeno-associated virus site-specific integration. Proc Natl Acad Sci USA 93:7966–7972.

50. Weitzman MD, Kyostio SR, Kotin RM, Owens RA (1994) Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. Proc Natl Acad Sci USA 91:5808–5812.

51. Urcelay E, Ward P, Wiener SM, Safer B, Kotin RM (1995) Asymmetric replication in vitro from a human sequence element is dependent on adeno-associated virus Rep protein. J Virol 69:2038–2046.

52. Cortes ML, Oehmig A, Saydam O, Sanford JD, Perry KF, Fraefel C, Breakefield XO (2008) Targeted integration of functional human ATM cDNA into genome mediated by HSV/AAV hybrid amplicon vector. Mol Ther 16:81–88.

53. Recchia A, Perani L, Sartori D, Olgiati C, Mavilio F (2004) Site-specific integration of functional transgenes into the human genome by adeno/AAV hybrid vectors. Mol Ther 10:660–670.

54. Young SM Jr, Samulski RJ (2001) Adeno-associated virus (AAV) site-specific recombination does not require a Rep-dependent origin of replication within the AAV terminal repeat. Proc Natl Acad Sci USA 98:13525–13530.

55. Yu Y, Bradley A (2001) Engineering chromosomal rearrangements in mice. Nat Rev Genet 2:780–790.

56. Sauer B, Henderson N (1990) Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. New Biol 2:441–449.

57. Thyagarajan B, Guimaraes MJ, Groth AC, Calos MP (2000) Mammalian genomes contain active recombinase recognition sites. Gene 244:47–54.

58. Sarkar I, Hauber I, Hauber J, Buchholz F (2007) HIV-1 proviral DNA excision using an evolved recombinase. Science 316:1912–1915.

59. Buchholz F, Angrand PO, Stewart AF (1998) Improved properties of FLP recombinase evolved by cycling mutagenesis. Nat Biotechnol 16:657–662.

60. Loonstra A, Vooijs M, Beverloo HB, Allak BA, van Drunen E, Kanaar R, Berns A, Jonkers J (2001) Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. Proc Natl Acad Sci USA 98:9209–9214.

61. Thorpe HM, Smith MC (1998) In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. Proc Natl Acad Sci USA 95:5505–5510.

62. Chalberg TW, Portlock JL, Olivares EC, Thyagarajan B, Kirby PJ, Hillman RT, Hoelters J, Calos MP (2006) Integration specificity of phage phiC31 integrase in the human genome. J Mol Biol 357:28–48.

63. Thyagarajan B, Olivares EC, Hollis RP, Ginsburg DS, Calos MP (2001) Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. Mol Cell Biol 21:3926–3934.

64. Ginsburg DS, Calos MP (2005) Site-specific integration with phiC31 integrase for prolonged expression of therapeutic genes. Adv Genet 54:179–187.

65. Glover DJ, Lipps HJ, Jans DA (2005) Towards safe, non-viral therapeutic gene expression in humans. Nat Rev Genet 6:299–310.

66. Liu J, Jeppesen I, Nielsen K, Jensen TG (2006) Phi c31 integrase induces chromosomal aberrations in primary human fibroblasts. Gene Ther 13:1188–1190.

67. Ehrhardt A, Engler JA, Xu H, Cherry AM, Kay MA (2006) Molecular analysis of chromosomal rearrangements in mammalian cells after phiC31-mediated integration. Hum Gene Ther 17:1077–1094.

68. Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. Genome Res 8:464–478.

69. Bryk M, Banerjee M, Murphy M, Knudsen KE, Garfinkel DJ, Curcio MJ (1997) Transcriptional silencing of Ty1 elements in the RDN1 locus of yeast. Genes Dev 11:255–269.

70. Bachman N, Gelbart ME, Tsukiyama T, 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.

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

72. Yieh L, Hatzis H, Kassavetis G, Sandmeyer SB (2002) Mutational analysis of the transcription factor IIIB-DNA target of Ty3 retroelement integration. J Biol Chem 277:25920–25928.

73. Aye M, Dildine SL, Claypool JA, Jourdain S, Sandmeyer SB (2001) A truncation mutant of the 95-kilodalton subunit of transcription factor IIIC reveals asymmetry in Ty3 integration. Mol Cell Biol 21:7839–7851.

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

75. Zou S, Ke N, Kim JM, Voytas DF (1996) The Saccharomyces retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci. Genes Dev 10:634–645.

76. Peters JE, Craig NL (2001) Tn7: smarter than we thought. Nat Rev Mol Cell Biol 2:806–814.

77. Kuduvalli PN, Mitra R, Craig NL (2005) Site-specific Tn7 transposition into the human genome. Nucleic Acids Res 33:857–863.

78. Loomis WF, Welker D, Hughes J, Maghakian D, Kuspa A (1995) Integrated maps of the chromosomes in Dictyostelium discoideum. Genetics 141:147–157.

79. Winckler T, Dingermann T, Glockner G (2002) Dictyostelium mobile elements: strategies to amplify in a compact genome. Cell Mol Life Sci 59:2097–2111.

80. Winckler T, Szafranski K, Glockner G (2005) Transfer RNA gene-targeted integration: an adaptation of retrotransposable elements to survive in the compact Dictyostelium discoideum genome. Cytogenet Genome Res 110:288–298.

81. Chung T, Siol O, Dingermann T, Winckler T (2007) Protein interactions involved in tRNA gene-specific integration of Dictyostelium non-long terminal repeat retrotransposon TRE5-A. Mol Cell Biol 27:8492–8501.

82. Collins CH, Yokobayashi Y, Umeno D, Arnold FH (2003) Engineering proteins that bind, move, make and break DNA. Curr Opin Biotechnol 14:665.

83. Katz RA, Merkel G, Skalka AM (1996) Targeting of retroviral integrase by fusion to a heterologous DNA binding domain: in vitro activities and incorporation of a fusion protein into viral particles. Virology 217:178–190.

84. 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.

85. Goulaouic H, Chow SA (1996) Directed integration of viral DNA mediated by fusion proteins consisting of human immunodeficiency virus type 1 integrase and Escherichia coli LexA protein. J Virol 70:37–46.

86. Szabo M, Muller F, Kiss J, Balduf C, Strahle U, Olasz F (2003) Transposition and targeting of the prokaryotic mobile element IS30 in zebrafish. FEBS Lett 550:46–50.

87. Maragathavally KJ, Kaminski JM, Coates CJ (2006) Chimeric Mos1 and piggyBac transposases result in site-directed integration. FASEB J 20:1880–1882.

88. Wu SC, Meir YJ, Coates CJ, Handler AM, Pelczar P, Moisyadi S, Kaminski JM (2006) piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. Proc Natl Acad Sci USA 103:15008–15013.

89. Yant SR, Huang Y, Akache B, Kay MA (2007) Site-directed transposon integration in human cells. Nucleic Acids Res 35:e50.

90. Akopian A, He J, Boocock MR, Stark WM (2003) Chimeric recombinases with designed DNA sequence recognition. Proc Natl Acad Sci USA 100:8688–8691.

91. Bushman FD, Miller MD (1997) Tethering human immunodeficiency virus type 1 preintegration complexes to target DNA promotes integration at nearby sites. J Virol 71:458–464.

92. Mandell JG, Barbas CF 3rd (2006) Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. Nucleic Acids Res 34:W516–W523.

93. Szczepek M, Brondani V, Buchel J, Serrano L, Segal DJ, Cathomen T (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. Nat Biotechnol 25:786–793.

94. Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, Urnov FD, Galli C, Gregory PD et al (2007) Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat Biotechnol 25:1298–1306.

95. Beerli RR, Segal DJ, Dreier B, Barbas CF 3rd (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proc Natl Acad Sci USA 95:14628–14633.

96. Tan W, Zhu K, Segal DJ, Barbas CF 3rd, Chow SA (2004) Fusion proteins consisting of human immunodeficiency virus type 1 integrase and the designed polydactyl zinc finger protein E2C direct integration of viral DNA into specific sites. J Virol 78:1301–1313.

97. Tan W, Dong Z, Wilkinson TA, Barbas CF 3rd, Chow SA (2006) Human immunodeficiency virus type 1 incorporated with fusion proteins consisting of integrase and the designed polydactyl zinc finger protein E2C can bias integration of viral DNA into a predetermined chromosomal region in human cells. J Virol 80:1939–1948.

98. Wilson MH, Kaminski JM, George AL Jr (2005) Functional zinc finger/sleeping beauty transposase chimeras exhibit attenuated overproduction inhibition. FEBS Lett 579:6205–6209.

99. Corbi N, Libri V, Fanciulli M, Tinsley JM, Davies KE, Passananti C (2000) The artificial zinc finger coding gene 'Jazz' binds the utrophin promoter and activates transcription. Gene Ther 7:1076–1083.

100. Ivics Z, Katzer A, Stuwe EE, Fiedler D, Knespel S, Izsvak Z (2007) Targeted Sleeping Beauty transposition in human cells. Mol Ther 15:1137–1144.

101. Ciuffi A, Diamond TL, Hwang Y, Marshall HM, Bushman FD (2006) Modulating target site selection during human immunodeficiency Virus DNA integration in vitro with an engineered tethering factor. Hum Gene Ther 17:960–967.

102. Zhu Y, Dai J, Fuerst PG, Voytas DF (2003) Controlling integration specificity of a yeast retrotransposon. Proc Natl Acad Sci USA 100:5891–5895.

103. Kipp M, Gohring F, Ostendorp T, van Drunen CM, van Driel R, Przybylski M, Fackelmayer FO (2000) SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA. Mol Cell Biol 20:7480–7489.

104. Izsvák Z, Khare D, Behlke J, Heinemann U, Plasterk RH, Ivics Z (2002) Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in Sleeping Beauty transposition. J Biol Chem 277:34581–34588.
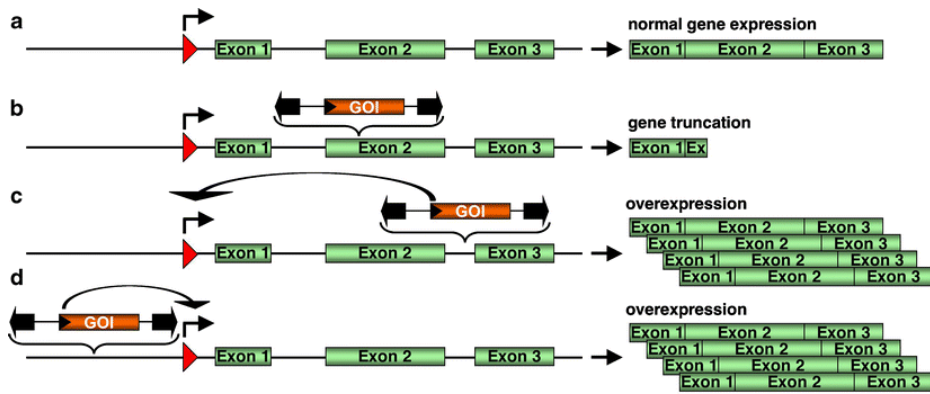
Voigt K et al.

Fig.1: Possible mutagenic consequences of transgene integration in or close to a transcription unit. **a** The figure depicts a hypothetical transcription unit with a promoter (*red arrow*) and three exons. Normal gene expression results in physiological levels of the correctly spliced protein. **b** A gene of interest (GOI) carried by an integrating vector inserts into an exon, thereby resulting in a truncated gene product. The black arrows flanking the GOI represent retroviral long terminal repeats or transposable element terminal inverted repeats. **c** Transgene insertion occurs in an intron. An enhancer linked to the GOI upregulates transcription of the endogenous gene, resulting in overexpression and/or ectopic expression. **d** Transgene insertion occurs upstream of the targeted gene. An enhancer linked to the GOI upregulates transcription of the endogenous gene, resulting in overexpression and/or ectopic expression
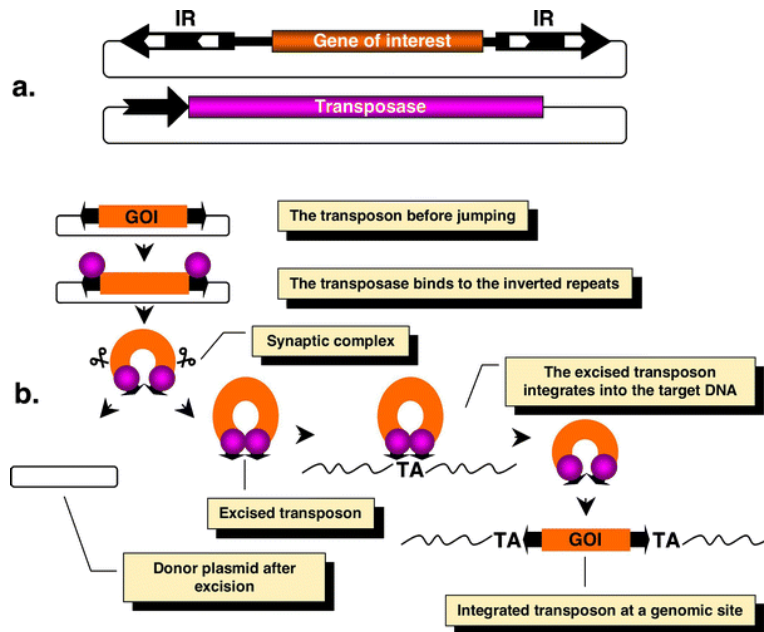
Fig.2: Transposon-based gene vectors. **a** Components and structure of a two-component gene transfer system based on *Sleeping Beauty*. A gene of interest (*orange box*) to be mobilized is cloned between the terminal inverted repeats (IR, *black arrows*) that contain binding sites for the transposase (*white arrows*). The transposase gene (purple box) is physically separated from the IRs, and is expressed in cells from a suitable promoter (black arrow). **b** Mechanism of cut-and-paste transposition. The transposable element carrying a gene of interest (GOI, *orange box*) is maintained and delivered as part of a plasmid vector. The transposase (*purple sphere*) binds to its sites within the transposon inverted repeats (*black arrows*). Excision takes place in a synaptic complex. Excision separates the transposon from the donor DNA. The excised element integrates into a TA site in the target chromosomal DNA (*wavy lines*) that will be duplicated and will be flanking the newly integrated transposon
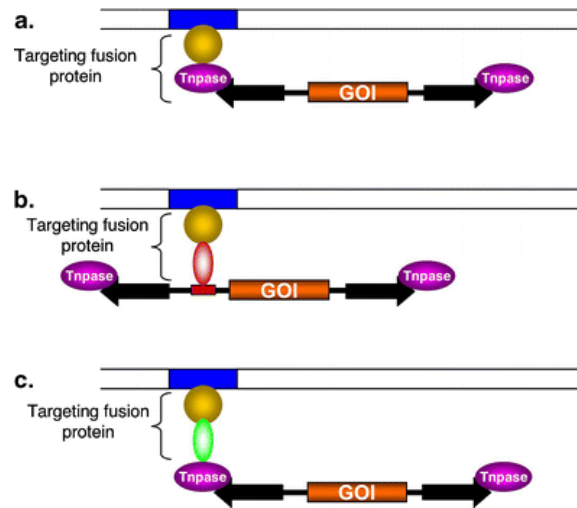
Fig.3: Experimental strategies for target-selected transgene integration by transposable element gene vectors. The common components of the targeting systems include a transposable element that contains the IRs (*arrowheads*) and a gene of interest (GOI) equipped with a suitable promoter. The transposase (*purple sphere*) binds to the IRs and catalyzes transposition. A DNA-binding protein domain (*yellow circle*) recognizes a specific sequence (*blue 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 (*red oval*) that binds to a specific DNA sequence within the transposable element (*red 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 (*green oval*) that interacts with the transposase. In this strategy, neither the transposase nor the transposon is modified

Table1: Integrating genetic elements showing targeted insertion in their natural hosts

Recombinase	Integration site	Efficiency of targeting	Natural host/origin	Cofactors	Reference
Cre recombinase	*loxP*; pseudo *loxP* sites exist in mammalian genomes	0,12% for human *lox* h7q21 (site specific/total integrations) (~100% excision for wt *loxP* in *E. coli*)	*Escherichia coli* (Bacteriophage P1)	No	[57]
Flp recombinase	FRT, pseudo sites in mammalian cells unknown	unknown	*Saccharomyces cerevisiae*	No	[59]
ΦC31 integrase	*attP*; pseudo *attP* sites; at least 11 in humans	5% for human psA site; 15% at inserted *attP* site	*Streptomyces lividans*	No	[63]
Tn7 transposase	Original target site: *attTn7* in the *E. coli glmS* gene; *gfpt-1* and *gfpt-2* are human orthologs of *glmS*	Frequency of targeting *attTn7* in *E. coli* 100%; Frequency of targeting *gfpt-1* (31%) is comparable to targeting *glmS* (32%) *in vitro*; for *gfpt-2* (23%)	Bacterial	TnsD transposase subunit binding to attTn7 in *glmS*	[76,77]
Ty1	within 750 bp of tRNA or other RNA pol III transcribed genes	90% of insertions	*Saccharomyces cerevisiae*	Interaction of integrase with TFIII components	[68,70]
Ty3	within 750 bp of tRNA or other RNA pol III transcribed genes	95% of insertions	*Saccharomyces cerevisiae*	Interaction of integrase with TFIIIB and C	[68]
Ty5	silent heterochromatin	21% of insertions adjacent or within transcriptional silencers flanking HML and HMR or subtelomeric repeat in chromosome III	*Saccharomyces cerevisiae*	Interaction of integrase with Sir4	[74,75]
TRE3	100–150 bp downstream of tRNA genes	100%	*Dictyostelium discoideum*	Not known	[80]
TRE5	50 bp upstream of tRNA genes	100%	*Dictyostelium discoideum*	Interaction of ORF1 with TFIIIB	[80,81]
DIRS	cluster formation on extremities of chromosomes	87% into transposons	*Dictyostelium discoideum*	Not known	[78]
AAV Rep	AAVS1	70–90%	*Homo sapiens*	Not known	[49,50]

Table2: Targeting of gene delivery systems by direct fusion to DANN-binding domains

Protein	Efficiency of targeting	Activity of chimeric enzyme	Experimental system	Reference
ASV IN/LexA	Significantly enriched integration adjacent to binding site	Full processing activity; similar joining activity	In vitro	[83]
HIV IN/				
λ-repressor	Enriched integration near binding sites	Fusion retained known activities of HIV IN	In vitro	[84]
LexA	Enriched integration near target site	No appreciable change in activities compared to wt	In vitro	[85]
Zif268	Different integration patterns when compared to wt with integration hot spots near targeted sites	Infectivity abolished, but restored when mixing in wt IN (up to 93% of wt only when mixed 1:1)	In vitro	[91]
E2C (ZF)	>60% of insertions within 10 bp of binding site (5% for wt) for GC-rich strand (for G-rich strand 14% bzw 32%; wt 5%)	no appreciable change in processing and joining activity	In vitro	[96]
	~ tenfold higher preference for integration near the E2C site as compared to wt IN	up to ~20% activity of viruses compared to wt	cell culture (HeLa, chromosomal target)	[97]
IS30/				
cI repressor	More than tenfold	Similar activity to wt ($2{,}5\times10^{-2}$ vs 2×10^{-2})	_E. coli_ (plasmid target)	[86]
Gli1	6 insertions near binding site (5 illegitimate, one legitimate); none found on target lacking the binding site	Low excision and integration efficiency	Zebrafish (plasmid target)	[86]
Tn3 resolvase/ Zif268	100% of plasmid molecules recovered were resolved (indicating efficient binding to binding site and recombinational activity)	44–98% (of resolved substrate molecules in recovered plasmid DNA)	_E. coli_	[90]
PB/Gal4	67% of insertions into location ~1 kb upstream of binding site	> tenfold increased transpositional activity (compared to transposition into plasmid lacking target site)	mosquito embryos (plasmid target)	[87]
Mos1/Gal4	96% of insertions into location ~1 kb upstream of binding site	> tenfold increased transpositional activity (compared to transposition into plasmid lacking target site)	mosquito embryos (plasmid target)	[87]
SB/				
Gal4	11-fold increase in a window around target site	26%	Cell culture (HeLa, plasmid target)	[89]
E2C (ZF)	Eightfold increase in a window around target site	20%	Cell culture (HeLa, plasmid target)	[89]
Jazz (ZF)	No targeted transposition	15%	Cell culture (HeLa, chromosomal target)	[100]

Table3: Targeting of gene delivery systems by fusing a DANN-binding domain to a protein domain that interacts with the recombinase

Protein	Interaction partner	Efficiency of targeting	Efficiency of transposition	Experimental system	References
HIV IN	LEDGF/p75 fused to λ repressor	Significantly enriched integration near the λ operator	Wild-type activity	In vitro	[101]
Ty5 IN	Sir4p fused to LexA	11%	100%	Yeast (plasmid target)	[102]
TD replaced with a Rad9p motif	LexA-FHA1	11%	100%	Yeast (plasmid target)	[102]
TD replaced with a 12 aa NpwBP motif	LexA-Npw38	14%	100%	Yeast (plasmid target)	[102]
SB	N57-TetR	10%	100%	Cell culture (HeLa transgenic for TRE-EGFP, chromosomal target)	[100]
Transposon	**Interaction partner**	**Efficiency of targeting**	**Efficiency of transposition**	**Experimental system**	**References**
LexA-SB transposon	LexA-TetR	Two insertions (out of 400 in total) 44 bp and 48 bp downstream of TRE	100%	Cell culture (HeLa transgenic for TRE-EGFP, chromosomal target)	[100]
	LexA-SAF Box	>Fourfold increase in the frequency of targeting a MAR	100%	Cell culture (HeLa, chromosomal target)	[100]