Review TRENDS in Biotechnology Vol.23 No.8 August 2005
Full text provided by www.sciencedirect.com
SCIENCE DIRECT

Site-directed genome modification: derivatives of DNA-modifying enzymes as targeting tools

Craig J. Coates[1], Joseph M. Kaminski[2], James B. Summers[3], David J. Segal[4], Andrew D. Miller[5] and Andreas F. Kolb[6]

[1]Department of Entomology, Texas A&M University, College Station, TX 77843, USA
[2]Transpovec Corporation, LLC, Nashville, TN 37215, USA
[3]Department of Radiology, University of South Alabama, Mobile, AL 36617, USA
[4]Department of Pharmacology, University of Arizona, Tucson, AZ 85721, USA
[5]Imperial College Genetic Therapies Centre, Department of Chemistry, Imperial College London, London, SW7 2AZ, UK
[6]Hannah Research Institute, Hannah Research Park, Ayr, KA6 5HL, UK

The modification of mammalian genomes is an important goal in gene therapy and animal transgenesis. To generate stable genetic and biochemical changes, the therapeutic genes or transgenes need to be incorporated into the host genome. Ideally, the integration of the foreign gene should occur at sites that ensure their continual expression in the absence of any unwanted side effects on cellular metabolism. In this article, we discuss the opportunities provided by natural DNA-modifying enzymes, such as transposases, recombinases and integrases, to mediate the stable integration of foreign genes into host genomes. In addition, we discuss the approaches that have been taken to improve the efficiency and to modify the site-specificity of these enzymes.

Introduction

Foreign DNA can be transported into eukaryotic cells by physical, chemical or biological methods (e.g. microinjection, liposomes, electroporation, gene-gun or viral vectors). Stable integration of the foreign DNA will only occur in a small proportion of the cells that have taken up the DNA. Integration of the foreign DNA is mediated by cellular DNA repair enzymes and occurs at random sites of the host genome. If the foreign DNA contains sequences that are identical to the host genome, it might be inserted by homologous recombination in a small fraction of the transduced cells.

Retroviral vectors integrate their DNA into the host genome efficiently and murine leukemia virus (MLV)- and lentivirus-based vectors are excellent tools for both animal transgenesis and gene therapy. However, retroviral vectors suffer from serious limitations. In the context of animal transgenesis, the limited packaging capacity of retroviral vectors dictates the use of small transgene expression cassettes, which often lack the regulatory elements required for the tissue-specific and abundant expression of the transgene. The integration at random sites often places the transgene into an environment that is not supportive of its expression. MLV-based vectors are additionally silenced by epigenetic mechanisms. In the context of gene therapy, retroviral vectors also carry the risk of insertional mutagenesis. If a transgene or a therapeutic gene would be inserted into the host genome at a pre-selected site, abundant, continuous expression of the foreign DNA could be accomplished in the absence of genome mutagenesis. Targeting of a defined site in a eukaryotic genome can be achieved by using either DNA-base pairing or sequence-specific DNA–protein interactions. Here we will focus on how natural DNA-modifying enzymes could be modified so that they mediate site-directed integration of foreign DNA. In an accompanying article in this issue of *Trends in Biotechnology* we discuss the design of nucleic acid and protein modules, which can target a foreign gene to a pre-defined site [1].

DNA-modifying enzymes: the tools of the trade

A variety of enzymes can catalyze the insertion of foreign DNA into a host genome. Viral integrases, transposases and site-specific recombinases mediate the integration of virus genomes, transposons or bacteriophages into host genomes. An extensive collection of enzymes with these properties can be derived from a wide variety of sources.

In the context of genome engineering, two properties of the DNA-modifying enzymes are important. First, they need to be sufficiently active in the target organism. Many transposases, for example, have been selected for minimal activity because excess transposon activity would endanger survival of the host and the transposon, [2,3]. Often, hyperactive mutants have to be derived from these enzymes to obtain sufficient levels of DNA mobility [2,4,5]. Second, the site-selectivity of the DNA-modifying enzyme is important. Some of the enzymes discussed subsequently (e.g. Cre or Flp) have a high degree of DNA sequence specificity, whereas others (e.g. retroviral integrases) do not. Transposases have a high degree of site selectivity for the transposon ends, although they appear to have limited

Corresponding author: Kolb, A.F. (a.kolb@hannah.ac.uk).
Available online 1 July 2005

 0167-7799/$ - see front matter © 2005 Elsevier Ltd. All rights reserved. doi:10.1016/j.tibtech.2005.06.009

specificity for the target sequence for *de novo* integration. However, even enzymes that are site-specific in their natural host (i.e. bacteria) might interact with multiple sites in a bigger eukaryotic genome. In addition, nucleosomes and DNA methylation could be obstacles for the activity and accuracy of prokaryotic enzymes in eukaryotic cells.

Recent data suggest that both catalytic activity and site-selectivity of DNA-modifying enzymes can be altered successfully, albeit within limits. In an ideal scenario, a new DNA site specificity could be grafted onto a DNA-modifying enzyme by simply adding a new DNA-binding domain (DBD). However, this is only possible in the rare cases where the catalytic domain and the natural DBD are spatially distinct and interchangeable [4]. Unfortunately, the 3D structure of proteins is often complex and amino acids at the opposite ends of the primary sequence interact to form a crucial protein domain [6].

In subsequent sections, we discuss three groups of enzymes that have attracted major interest as genome engineering tools, namely viral integrases, transposases and site-specific recombinases, and the progress that has been made in defining and re-defining and their target sequences.

Viral DNA integration systems

DNA integration using retroviruses

Retroviruses combine several useful features, including the relative simplicity of their genomes, ease of use and their ability to integrate into the host cell genome, permitting long-term transgene expression in the transduced cells or their progeny. They have, therefore, been used in a large number of gene-therapy protocols [7].

Vectors based on murine leukemia viruses can integrate only in dividing cells and are often silenced by epigenetic mechanisms. This has significantly hampered their use in animal transgenesis [8]. Lentivirus vectors, however, appear to be able to overcome these limitations [9], which has made them attractive candidates for both gene therapy and transgenic applications [10]. Virus entry into cells is limited by the availability of suitable receptor molecules but viral particles can be equipped (pseudotyped) with proteins that enable either a broad or a specific range of cells to be infected [11]. However, the size of foreign DNA that can be packaged successfully in retroviral vectors is limited. The random integration of retroviruses can, under particular circumstances (which are influenced by the integration site, the therapeutic gene in the retroviral vector and the cell type infected), lead to the activation of oncogenes or the inactivation of tumour suppressor genes [12]. Recent reports of leukemia in two gene therapy patients (after retroviral transactivation of the *LMO2* oncogene [12]) have shown that this is more than a theoretical concern.

Retroviral integrases are synthesized as part of the pol protein, which also encompasses the reverse transcriptase and protease domains. An integration complex, crucially relying on the integrase protein but containing several additional viral proteins, is responsible for the integration reaction of human immunodeficiency virus 1 (HIV-1) [13]. Cellular proteins involved in the non-homologous end joining pathway have also been implicated in the completion of the viral integration process [14]. Crystal structures of integrase proteins are available for HIV-1, simian immunodeficiency virus (SIV) and Rous sarcoma virus (RSV). Retroviral integrases possess a catalytic triad of acidic amino acids, the D, D-35-E motif, which coordinates a divalent metal ion (Mg^{2+} or Mn^{2+}). The catalytic domain is in the central part of the protein (Figure 1). The N-terminal fragment consists of three α-helices, which coordinate Zn^{2+} via conserved cysteine and histidine residues. The C-terminus has an all β-strand SH3 fold and binds DNA non-specifically [15] (Figure 1). Mix-and-match mutants of HIV-1 and feline immunodeficiency virus (FIV) integrase have identified an α-helix in the catalytic core region (amino acids 118–121 in the HIV-1 integrase) as being involved in determining target sequence preference [16]. However, the target sequence specificity of retroviral integrases is limited. Hot spots for integration exist but a large number of DNA sequences can serve as integration targets. Integration often occurs in transcriptionally active segments of the genome with MLV integrations favoring regions near a transcriptional start site [17]. Nevertheless, high levels of transcription have been found to be inhibitory to integration of an avian retrovirus [18].

The lack of sequence specificity of integration has been recognized as a limitation of retroviral vectors for many years. Therefore, retroviral integrases, specifically the HIV-1 integrase, were among the first molecules to be engineered to recognize specific sequences by fusing them to the phage λ repressor and the Zif268 transcription factor [19,20]. The fusion proteins were able to direct integration of model substrates bound to purified pre-integration complexes to the vicinity of λ repressor and Zif268 binding sites. However, the generation of functional retroviral integration complexes has proved to be difficult and virions carrying mutated integrase genes were not viable [20,21]. This could be due to several reasons. First, the integrase is derived from a polyprotein encoding two other essential proteins. In the case of HIV-1, the integrase gene also overlaps with the open reading frame (ORF) of another viral protein (vif) and contains an essential splice acceptor site. Drastic sequence alterations in this genome region are refractory to virus viability [22]. However, by fusing the integrase-DBD proteins to the C-terminus of the viral accessory protein vpr and concomitant inactivation of the endogenous integrase by mutating one catalytic amino acid it has been possible to generate virions, which are able to integrate into the genome of human cells, albeit at much reduced efficiency [13]. Fusion proteins linking HIV-1 integrase to the polydactyl zinc-finger protein E2C have been reported to restrict integration of model substrates *in vitro* to a 10-bp region flanking the zinc-finger binding site [23].

DNA integration using adeno-associated virus (AAV)

Adeno-associated virus is a small DNA virus (parvovirus) that is co-replicated in mammalian cells together with helper viruses such as adenovirus, herpes simplex virus or human cytomegalovirus. The viral genome essentially

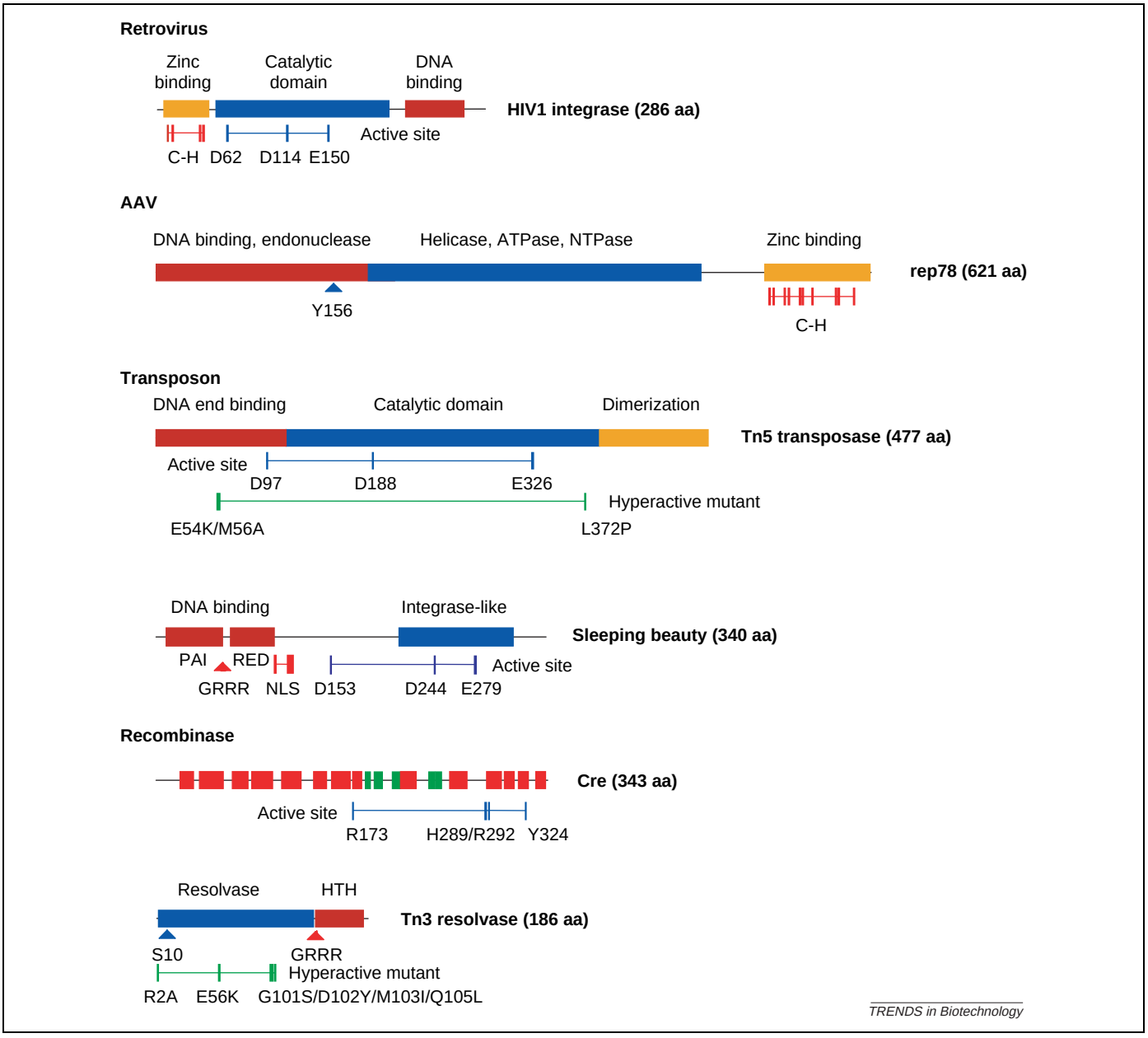

Figure 1. Functional domains of six DNA-modifying enzymes. Catalytic domains are shown as blue boxes and active-site residues are shown as vertical blue lines. Retroviral integrases (e.g. HIV-1 integrase) and transposases [e.g. Tn5 transposase and sleeping beauty transposase (SB10)] share a common D, D-35-E motif at the catalytic center. Catalytic residues in AAV rep78 (Y156, essential for endonuclease activity) and Tn3 resolvase (S10, essential for covalent binding of DNA) are shown with blue arrowheads. Clusters of Zn^{2+}-coordinating cysteine (C) and histidine (H) residues in HIV integrase and rep78 are shown as yellow boxes; the positions of the individual residues are marked as vertical red lines. DNA-binding domains (DBDs) are shown in red [HTH (helix-turn-helix) motif]. Red arrowheads mark the GRRR AT-hook homeodomains in SB10 and Tn3 resolvase. The distribution of α-helices (red boxes) and β-sheets (green boxes) is shown for Cre recombinase. Amino acid changes, which generate hyperactive forms of Tn5 transposase and Tn3 resolvase, are shown as vertical green lines.

consists of only two ORFs (rep, a non-structural protein, and cap, a structural protein) from which (at least) seven different polypeptides are derived by alternative splicing and alternative promoter usage (Figure 2). In the presence of a helper-virus, the rep proteins mediate replication of the AAV genome. Integration, and thus a latent virus infection, occurs in the absence of helper virus. Mechanistically, replication and integration appear to be similar processes [24]. Approximately 70–85% of AAV integrations are found at a defined site (termed AAVS1) of the human chromosome 19q13 in the vicinity of the slow skeletal troponin T gene [24,25]. The integration site resides in an open chromatin structure in human cell lines and enables expression of inserted genes [26]. DNA fragments encompassing between 1.6 kb and 8.2 kb of sequence derived from AAVS1 are sufficient to support site-specific integration of AAV in non-human cell lines and transgenic mice and rats [27].

Integration of AAV is mediated by the rep proteins, which are produced from the same viral gene (Figure 2). The rep proteins perform a variety of functions and possess DNA-binding, endonuclease, ligase, helicase and ATPase activity [28] (Figure 1). To mediate site-specific integration Rep78 interacts with the rep-binding element (RBE) and a stem-loop structure in the inverted terminal repeats (ITR) of AAV, and introduces a nick into one of the

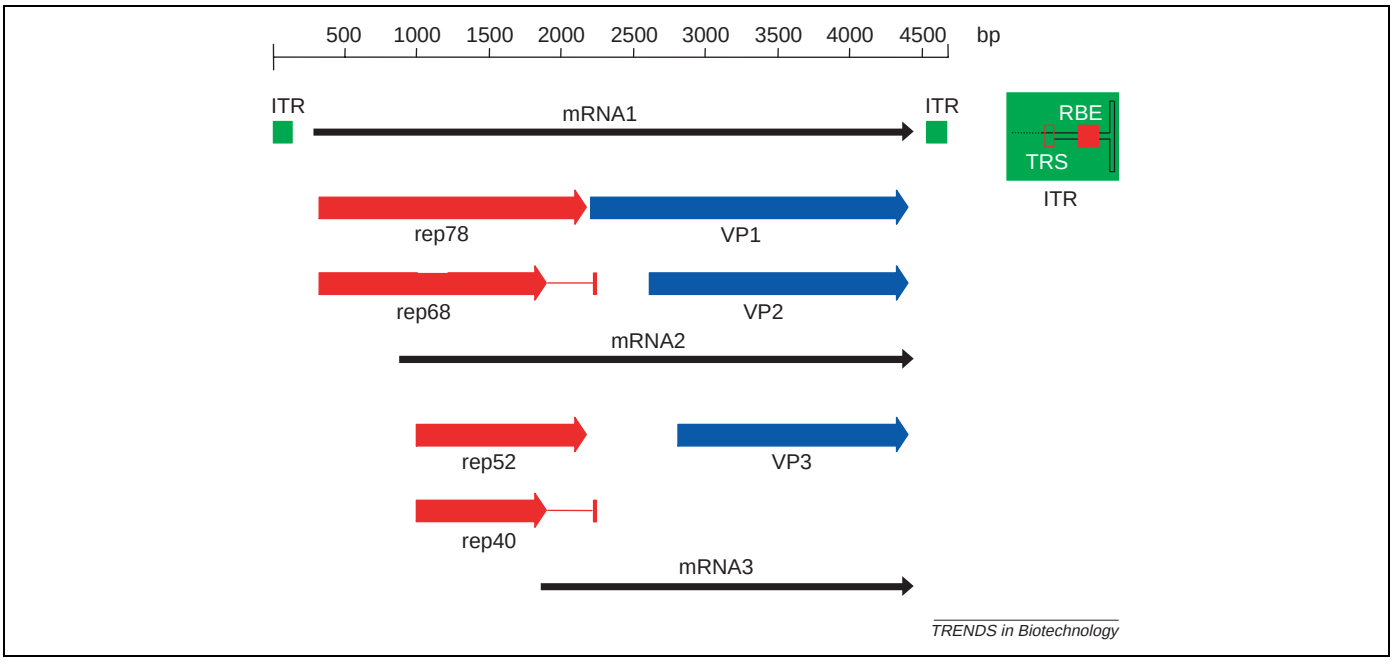

Figure 2. Structure of the AAV2 genome. Open reading frames for the non-structural rep proteins are shown as red arrows. Open reading frames for the structural proteins [VP1–3 (viral protein 1–3)] are shown as blue arrows. Inverted terminal repeats (ITR) are shown as green boxes. The ITR hairpin structure, the rep-binding element (RBE) and the terminal resolution site (TRS) are represented in the expanded version of the ITR.

DNA strands at a specific site termed the terminal resolution site (TRS) [24] (Figure 2). The same sequence elements as in the ITR and the potential for the formation of a stem-loop structure are found at AAVS1 [29]. The crystal structure of the N-terminal 197 amino acids of AAV5-rep78, together with the ITR, suggests that different parts of the domain interact with the RBE, the stem-loop structure and the TRS [30].

Linking the N-terminal domain of the homologous goose parvovirus (GPV) rep1 protein to the remainder of the AAV rep78 protein resulted in a chimeric protein, which preferentially bound to the GPV ITR rather than the AAV ITR, indicating that a sequence-specific DBD resides in this N-terminal part of the rep proteins. In addition, it could be shown that the catalytic centre mediating endonuclease activity is also located in the N-terminal region of the rep78 protein [28]. A tyrosine residue (Tyr156) was found to be essential for endonuclease and ligase activity of rep78 [31] (Figure 2). Rep78 also interacts with several cellular proteins, including high-mobility group chromosomal protein 1 (HMGP-1), protein kinase A, transcription factors Sp1 and PC4, and Topors, a topoisomerase binding protein ([32] and references therein). These interactions can modulate site-specific integration, AAV transcription and host cell gene transcription.

AAV vectors for gene therapy usually lack the rep genes to extend the limited packaging capacity of the virus. Therefore, the packaged DNA remains predominantly episomal in the infected cells [33]. A low level of AAV integration occurs in the absence of rep78 at a large number of sites [34] and is enhanced significantly by DNA double-strand breaks [35]. Therefore, directing double-stand breaks to defined regions of the genome could be one route to accomplish site-directed transgene integration at sites other that AAVS1 [36]. Plasmid based systems using the AAV rep-proteins and the ITR of AAV have been used successfully to support site-specific integration [37]. In addition, chimeric virus systems combining the advantages of adenovirus and AAV have been established [27,38]. These approaches have been made possible by expressing the rep protein conditionally, thereby overcoming its inhibitory effect on cell viability.

In summary, AAV has the useful ability to target a transcriptionally permissible site in the human genome. This targeting potential can be used for the transfer of foreign DNA in hybrid vectors or plasmid-based systems. The specificity of the natural rep protein is already useful for gene therapy approaches; therefore, the alteration of rep site-specificity has not been explored in great detail but the N-terminal domain of rep78 might be a useful starting point to try to redefine target specificity.

Transposon-based DNA integration systems

Transposons are segments of mobile DNA that can be found in a variety of organisms. Although active transposons are found in many prokaryotic systems and insects, no functional natural transposons exist in vertebrates. The *Drosophila P* element transposon has been used for many years as a genome engineering tool [39]. Bacterial transposons, especially conjugative transposons, are major contributors to the spread of antibiotic resistance genes. Non-vertebrate transposons function in vertebrate cells, albeit at low efficiency [40,41]. The sleeping beauty transposon was established from non-functional transposon copies found in salmonid fish and is significantly more active in mammalian cells than prokaryotic or insect transposons [41]. Several laboratories have demonstrated the ability of the enzyme to mediate stable integration of foreign DNA into mouse or human cells [42,43].

A minimal transposon consists of a transposase gene and the terminal repeat sequences, which are recognized by the transposase. However, other genes (e.g. therapeutic

Table 1. Alterations of DNA-binding properties of DNA-modifying enzymes

Enzyme	Target sequence: wild-type target site and mutated target site[a,b]	Number of amino acid changes[c]	Method	Refs
Cre (343 aa)	loxP ATAACTTCGTATA GCATACAT TATACGAAGTTAT loxH ATATATACGTATA TAGACATA TATACGTATATAT (m5/26)	9–15	Substrate-linked *in vitro* evolution by error-prone PCR	[83]
Cre (343 aa)	loxP ATAACTTCGTATA GCATACAT TATACGAAGTTAT loxM7 ATAACTCTATATA GCATACAT TATATAGAGTTAT (m6/26)	5	Randomization of amino acid residues in direct contact with DNA	[82]
Flp (423 aa)	FRT GAAGTTCCTATAC TTTCTAGA GAATAGGAACTTC mFRT11 GAAGTTTCTATAG TTTCTAGA CAATAGAAACTTC (m2/26)	2 or 3	Error-prone PCR and DNA shuffling	[84]
Flp (423 aa)	FRT GAAGTTCCTATAC TTTCTAGA GAATAGGAACTTC mFRT71 GAAGTTTCTATAG TTTCTAGA CAATAGAAACTTC (m2/26)	1	Error-prone PCR and DNA shuffling	[84]
ϕC31 (613 aa)	attP CCCCAACTGGGGTAACCT TTG AGTTCTCTCAGTTGGGGG ψatt TAAGTACTTGGGTTTCCC TTG GTGTCCCCATGGAGATTT (m22/39)	~10	DNA shuffling	[90]
Tn3 resolvase (186 aa)	resI CGT TCGA AATATTATAAATT ATCAG ACA ZR GCGTGGGCG ACGA AATATTATAAATT TGCAT CGCCCACGC (m10/28) Zif-268 Zif-268	Exchange of 37 amino acids of HTH for 94 amino acids of zinc finger	Generation of chimeric protein	[4]
Tn5 (477 aa)	outside end CTGACTCTTATACACAAGT inside end CTGTCTCTTGATCAGATCT (m7/19)	4–7	DNA shuffling	[45]

[a]The upper strand of the target sequences and the corresponding mutants are shown. Nucleotides that are identical in both sequences are marked with a '|'. The number of mutated nucleotides with respect to the wild-type sequence is indicated in brackets [e.g. (m5/26) denotes that 5 out of 26 nucleotides are mutated].
[b]Abbreviation: ZR, Z-resolvase binding site.
[c]The number of amino acid changes required to generate a recombinase with an altered site-specificity is shown.

genes) can also be included in the transposon. Transposases possess a catalytic triad of two aspartate and one glutamate residue (D, D-35-E), which, as in retroviral integrases, complex a divalent ion. Transposases can act via different reaction mechanisms: either a free hairpin DNA is excised from the original site and is then transported as a synaptic complex to a new integration site or, alternatively, one DNA strand is cleaved and attached to a new target site before the second strand is cleaved [44].

Crystal structures have been solved for several transposases and transposase fragments including Tn5 transposase, MuA and Ty3 transposase [44]. These structures have demonstrated that the catalytic core region is indeed related to a similar fold in retroviral integrases [6,44]. The Tn5 crystal structure also demonstrates that the three domains, which can be deduced from the primary sequence (N-terminal, catalytic and C-terminal as shown in Figure 1), do not have discreet functions. All of them participate in DNA-binding and the formation of the synaptic complex [6]. Tn5 transposase mutants that were generated by DNA shuffling and selected for their ability to interact with an altered transposon end sequence (Table 1) carried mutations in all three 'domains' of the protein (Figure 3) [45]. Nevertheless, biochemical evidence in several systems demonstrates that the helix-turn-helix motives at the N-terminus of the transposase proteins are crucial for interaction with the transposon ends. Sequence comparisons suggest a similarity to DNA-binding motives found in other proteins, especially Pax transcription factors [46]. DNA encompassing the Tn5 outside ends partially protects the N-terminus of the transposase protein against proteolytic cleavage [47]. The N-terminal domain of the Tc3 transposase could be crystallized together with a 21-bp fragment of the transposon ITR [44]. An alanine scanning mutagenesis approach of the N-terminal (amino acids 1–123) DNA-binding region of the sleeping beauty transposase (SB10) identified four amino acids that crucially influence interaction with the transposon ends [48]. Ten other mutations were found in the same region that significantly increased transposition activity. Fragments of the N-terminal domain of SB10 specifically bind to the transposon ends [49].

The DNA-binding specificity of SB10 was modulated by generating chimeras with the DBD of Tc1 transposase, mixing segments of the paired-like DNA binding domains [PAI–RED (Figure 1)]. The PAI domain was found to determine DNA-binding specificity of the chimeric proteins. However, the chimeras were unable to cross-mobilize the transposons specified by the PAI domain [49].

Different transposons have adopted different strategies to maintain a relationship with their host that enables long-term coexistence [50]. In some cases, transposases have been selected for inactivity. Activity of the Tn5 transposon and the *Drosophila P* element is controlled by inhibitor molecules, which are essentially inactive deletion mutants of the functional transposase [2,51]. In

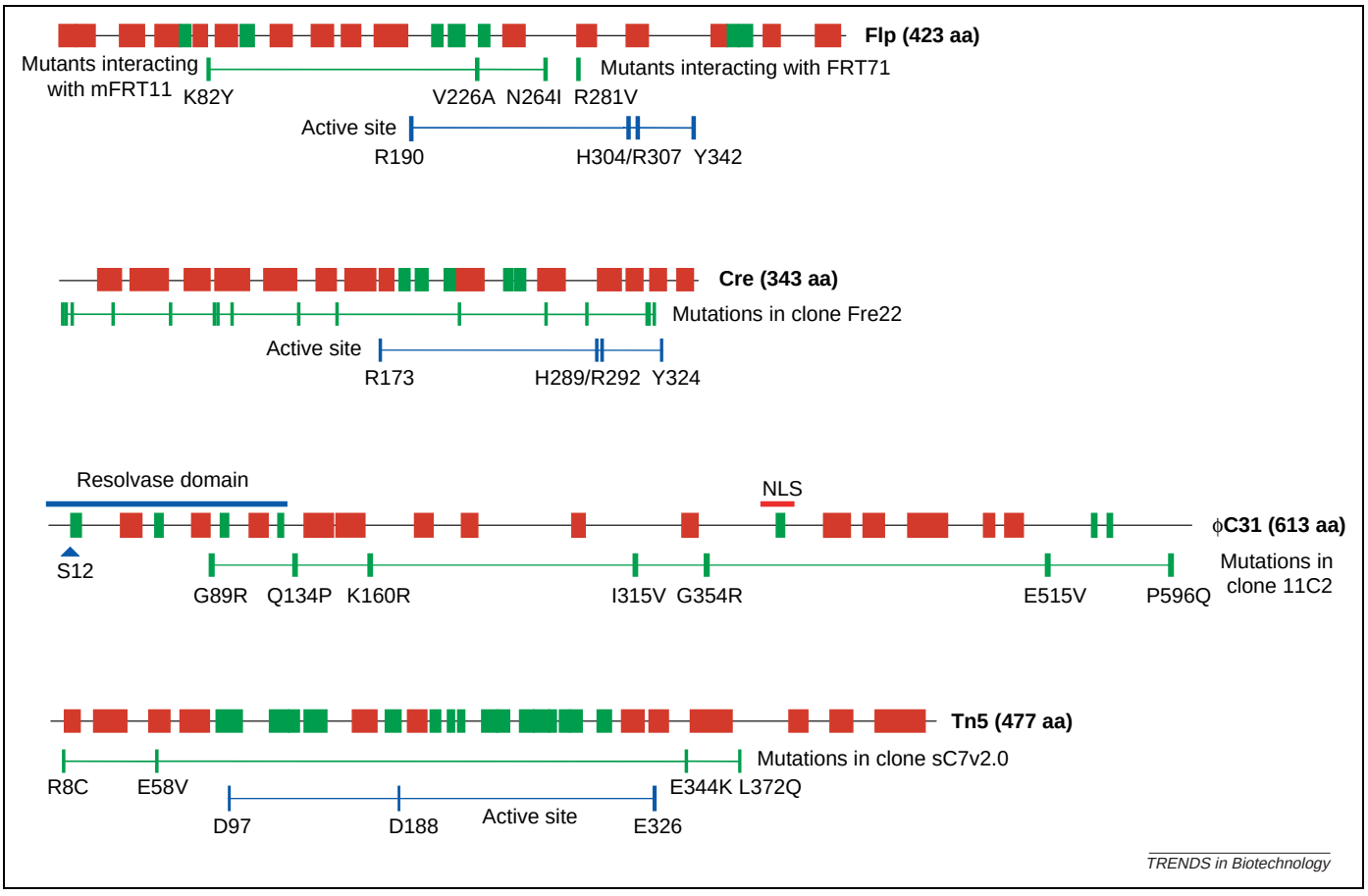

Figure 3. Distribution of mutations that affect target site selection. Flp, Cre, ϕC31 integrase and Tn5 transposase are shown as a sequence of α-helices (red boxes) and β-sheets (green boxes) derived from their crystal structures (Cre, Flp and Tn5 transposase) or structure prediction using the predict protein program [110]. Active-site residues are shown as vertical blue lines (Flp, Cre and Tn5) or blue arrowheads (ϕC31). Mutations that alter target site specificity are shown as vertical green lines. Mutations of Flp, Cre, fC31 and Tn5 transposase are derived from Refs [74], [73], [90] and [45], respectively.

bacteria and yeast, some transposons have been selected to mediate integration into defined sites in the host genome that are not refractory to host survival [3,52].

Most transposases, although having a tight sequence specificity with regard to their cognate transposon ends [49], are relatively flexible with regard to their *de novo* integration site in that they either recognize a loose or short consensus sequence. For example, the *mariner* and *Tc1* superfamily of transposons recognize TA target sites and duplicate them upon insertion [53]. *PiggyBac* (a transposase derived from the moth *Trichoplusia ni*) recognizes and duplicates TTAA target sites [54], whereas the *Hermes* element (derived from the house fly *Musca domestica*) recognizes and duplicates an 8bp consensus sequence [55]. Recent studies with SB10 have demonstrated that target site-selection is dependent on DNA structure rather than sequence *per se* [56–58]. Furthermore, sleeping beauty-mediated transposition is strongly enhanced by CpG methylation [59]. SB10, like retroviral integrases and AAV rep78, has been shown to interact with a host cell factor. HMGB-1 enhances SB transposition significantly and appears to assist the process of synaptic complex formation [60]. Interaction with host cell factors might also influence target site-selection. The yeast retro-transposon Ty5 interacts with the host protein SIR4C to target heterochromatin sites of the host genome (e.g. telomeres and silent mating loci). This strategy ensures that transposition avoids genome areas of high transcriptional activity. By replacing the so-called targeting domain of Ty5 integrase, which mediates the interaction with SIR4C, with peptides that redirect binding to other cellular proteins, it was possible to modulate site-selectivity [52].

Some transposons (e.g. Tn7) have a marked target site selectivity in their natural hosts. The Tn7 transposase complex contains four proteins TnsA, B, C and D. TnsD recognizes a singular sequence in the *Escherichia coli* genome at the 3′ end of the glucosamine synthetase ORF and mediates assembly of the transposition complex, which then catalyzes transposition into a site downstream of the *glmS* gene [3]. TnsD binding to DNA results in a distortion of the DNA target site, which then acts as a landing site for the TnsABC complex [61]. DNA bending promotes transposition and recombination in several systems, presumably by facilitating the formation of active recombinase–DNA complexes [60]. Tn7 transposition is also modulated by interaction with two host proteins, ribosomal protein L29 and acyl carrier protein [62]. Tn7 is also able to use the highly conserved human homologue of glmS, gfpt-1 (70% sequence identity to the *E. coli* gene) as a target site in human cells [63]. It is noteworthy that the DBD specifying the transposition target is located on a different polypeptide than the transposase catalytic domain (present on the TnsABC complex, with the

catalytic D, D-35-E motif on TnsB). TnsD might be more resilient with regard to changes in its site-selectivity because mutations in the DBD do not influence the catalytic domain located on TnsB directly. Similar situations might exist in variety of other transposon systems.

Many transposases do not have a pronounced site-selectivity, therefore, it was reasoned that they could be linked to DBDs of proteins, which can provide this specificity and guide the transposase to a defined genomic site [64]. Such an approach has been shown to be viable using the *E. coli* IS30 transposase. This protein has a strong sequence preference for a singular site in the *E. coli* genome and most integrations are found at this hot spot [65]. The DBDs derived from the cI repressor of phage λ (which directs integration primarily into the vicinity of the λ operator in *E. coli*) or the DBD of the mammalian transcription factor Gli1 were simply attached to the C-terminus of the full length transposase and analyzed in *E. coli* and zebrafish. Interestingly, attachment of the cI DBD enhanced transposition activity. The DBD redirected a significant subset of integration events to sites around the λ operator sequence but did not abolish the affinity of the transposase for its natural target in the *E. coli* genome. Juxtaposition of the transposase and the transcription factor Gli1 DBD was able to direct integrations of a marker gene in the context of the IS30 transposon to the vicinity of a Gli1 binding site placed in the zebrafish genome. Surprisingly, most integration events were not due to transposition but illegitimate recombination events [40]. This suggests that the DBD correctly targets the desired site and that the catalytic activity of the transposase modifies the host DNA such that DNA repair enzymes are recruited.

Recombinase-based DNA integration systems

Site-specific recombinases are enzymes that catalyze DNA strand exchange between DNA segments that possess only a limited degree of sequence homology [66,67]. They bind to recognition sequences that are between 30 and 200 nucleotides in length, cleave the DNA backbone, exchange the two DNA double helices involved and religate the DNA. Characteristically, a covalent bond between the protein and the DNA target is formed and resolved as part of the reaction mechanism. In some site-specific recombination systems, a single polypeptide is sufficient to perform all of these reactions, whereas other recombinases require a varying number of accessory proteins to fulfil these tasks. Site-specific recombinases can be clustered into two protein families with distinct biochemical properties, namely tyrosine recombinases (in which the DNA is covalently attached to a tyrosine residue) and serine recombinases (where covalent attachment occurs at a serine residue). The most popular enzymes used for genome modification approaches are Cre (a tyrosine recombinase derived from *E. coli* bacteriophage P1) and ϕC31 integrase (a serine recombinase derived from the Streptomyces phage ϕC31).

Several other bacteriophage derived site-specific recombinases (including Flp, lambda integrase, bacteriophage HK022 recombinase, bacteriophage R4 integrase and phage TP901–1 integrase) have been used successfully to mediate stable gene insertions into mammalian genomes [5,68–71]. This suggests that these enzymes can be the basis for the development of gene integration systems, which carry DNA-binding sites interacting with precise genomic locations.

Targeted integration using Cre

Tyrosine recombinases carry a catalytic centre consisting of a R, H, R triad and a tyrosine residue that acts as a nucleophil (Figure 1). Crystal structures are available for several members of the family including lambda integrase, Cre, Flp and *E. coli* XerD/C [72–75].

The bacteriophage P1 protein Cre is the most frequently used recombinase in biomedical research. Cre-mediated recombination does not require any co-factors and is effective in prokaryotic and eukaryotic cells. Cre mediates excision of DNA segments that are flanked by its target site, loxP, during the life-cycle of the bacteriophage [76]. LoxP sites consist of two 13bp inverted repeats, which flank an 8bp unidirectional spacer segment. Cre cleaves the two DNA segments involved in the strand exchange within this spacer segment and generates a protruding end of six nucleotides. The protruding ends of the two DNA strands anneal by conventional base-pairing and are ligated by Cre. For efficient recombination to occur, two conditions have to be met. First, the recombinase needs to have a significant affinity to the 13bp inverted repeat segments and, second, the sequences of the two 6bp protruding ends need to be identical in both DNA strands [77]. Cre can catalyze the deletion of DNA segments, which are flanked by direct repeats of lox sites or inversions of DNA segments flanked by inverted repeats of lox sites. Both reactions are reversible and Cre can, therefore, also catalyze DNA integration. However, as the sequences of the lox sites are not altered during the reaction, every integration product remains a target for a deletion reaction. The local site concentration is higher in the integrated state than in the excised state, therefore, deletion of DNA is the thermodynamically favored reaction. However, stable integration of genes into a host genome can be achieved by a process termed recombinase-mediated cassette exchange [68,78], as discussed in a recent review in this journal [79]. We found that under certain experimental conditions, transgene integration in murine cells carrying suitable target sites is almost exclusively due to Cre-mediated cassette exchange (and not to random integration). At the same time, integration is limited to the RMCE target in the genome (D.A. Sorrell and A.F. Kolb, unpublished). This demonstrates that Cre can be both efficient and highly site-selective.

Genomic sites that represent potential targets for Cre activity have been identified in the human and the mouse genome [80]. However, no Cre-mediated integration of therapeutic genes into these sites has been demonstrated to date. In addition, despite the frequent use of Cre in whole animal systems (e.g. mouse) few side effects, which could be ascribed to Cre-mediated genome alterations, have been documented [81]. This argues against a strong affinity of Cre towards genomic pseudo-sites.

The analysis of Cre by X-ray crystallography has shown that large portions of the proteins participate in

DNA interactions and that individual sub-domains are difficult to define. Nevertheless, the site-specificity of Cre can be modulated to achieve integration of therapeutic genes. This has been accomplished in two independent sets of experiments (Table 1). Santoro and Schulz randomized two pairs of five amino acid residues in Cre that were known to make contact with its DNA target and tested the resulting library of mutants for their ability to recombine several mutated lox sites [82]. Buchholz and Stewart used error-prone PCR to introduce mutations in the entire Cre protein [83] In this approach, a lox-like sequence derived from human chromosome 22 (loxH), which is not readily recombined by native Cre, was used as a recombination target for the mutant enzymes. Both approaches yielded Cre derivatives with altered site-specificity (Table 1). The amino acid alterations, which enabled the mutant Cre protein to recombine the loxH site were distributed throughout the protein (shown for mutant Fre22 in Figure 3).

The yeast recombinase Flp is closely related to Cre, is effective in prokaryotic and eukaryotic cells and also recognizes a minimal 34bp target sequence (Table 1). However, Flp is less active than Cre due to its lower stability at physiological temperature. The target site-specificity of Flp could be modified successfully [84]. Although the alterations of the target site (mFRT11 and mFRT71) (Table 1) only affect two nucleotides, these changes are at crucial positions and naïve Flp recombinase has a recombination activity below 0.1% at these sites. By using an error-prone PCR approach, protein mutants were derived that were able to recombine the mutated FRT sites. A single amino acid change at lysine 82 (either to methionine, threonine or tyrosine) resulted in an Flp protein that recognized both FRT and mFRT11. An additional mutation (V226A) diminished activity at FRT, while maintaining activity at mFRT11. A single amino acid change (R281V) was sufficient to alter the specificity of Flp from FRT to mFRT71. Combinations of the two sets of mutations were able to recombine an mFRT11–71 site carrying both sequence alterations, albeit at reduced efficiency. These experiments demonstrate that changes in site-specificity in tyrosine recombinases can be introduced with few amino acid changes. Interestingly, the V226A mutation, which shifted the specificity of Flp to a new target site, is not directly involved in DNA interactions [74,84]. This indicates that the influence of single amino acids on site-specificity is not always predictable, even in the presence of structural information on the DNA–protein interaction. These results also suggest that random mutagenesis and functional screening could be the most powerful approach to isolating recombinase mutants with altered site-specificities.

Targeted integration using ϕC31 recombinase

Recently, a site-specific recombinase has been purified from the Streptomyces bacteriophage ϕC31 [85]. The ϕC31 recombinase is a member of the resolvase family and mediates phage integration. In this process the bacteriophage attP site recombines with the corresponding attB site in the bacterial genome. The crossover generates two sites, attL and attR, which are no longer a target for recombinase action, in the absence of accessory proteins. The reaction also takes place in mammalian cells and can therefore be used to mediate site-specific integration of therapeutic genes [86]. Natural sites in the mouse and human genome have been used successfully as recombination targets [87]. In fact, even if a wild-type attP site is present in the genome of murine or human cells, only 50% of integrations of a plasmid equipped with a corresponding attB site will take place at this site, whereas all other integration events will occur at so called pseudo att sites. About 100 different sites in the human genome were able to act as potential integration sites in human cell lines [87]. However, only two sites in the mouse genome were found to be viable integration sites in mouse liver *in vivo* [88], whereas around 50 sites could be targeted in murine cell lines [87]. This suggests that chromatin structure poses constraints on the recombinase-mediated integration of therapeutic genes *in vivo*.

Stable integration of a plasmid carrying a therapeutic factor IX gene and an attB site in the liver of mice after high-pressure tail vein injection was increased by 10-fold when a ϕC31 recombinase gene was co-injected [88]. This indicates that the recombinase efficiently mediates plasmid integration. Similarly, several transgenes encoding extracellular-matrix proteins were introduced into human keratinocytes to correct hereditary diseases [89]. Integration at six different genomic pseudo-att sites was observed in these experiments with a bias towards a single site on chromosome 8. This demonstrates that the ϕC31 recombinase can be used to stably integrate therapeutic genes at predictable sites *in vivo*. The ϕC31 recombinase also appears to be amenable to alterations of its site-specificity. By using a DNA shuffling approach, the site-specificity of the ϕC31 integrase has been modified successfully [90]. The resulting protein recognized a pseudo-att site in the human genome three-times better than the native fC31 integrase (Table 1). The mutations altering the specificity of the integrase in the DNA shuffling experiment were found throughout the protein (Figure 3) arguing against the DBD residing in a defined section of the primary sequence.

Targeted integration using Z-resolvases

The site-specificity of tyrosine-recombinases has been difficult to modify by direct protein engineering because the catalytic domain and the DNA recognition domain are closely interwoven [67]. Therefore, changes in specificity are often accompanied by a loss in activity [91]. Serine-recombinases might be more amenable to engineering approaches because X-ray crystallographic data suggest that the DBD is spatially distinct from the catalytic domain [92]. A hyperactive derivative of Tn3 resolvase has been modified by exchange of the natural DBD for a zinc-finger domain of the human zinc-finger transcription factor Zif268 [4]. The DNA site-specificity of the resulting chimeric protein, termed Z-resolvase, had been switched to that of Zif268. Zinc-finger proteins can be modified by *in vitro* protein evolution to recognize any DNA sequence [93]; therefore, this approach could theoretically enable development of chimeric recombinases that can integrate therapeutic genes into precise genomic locations. However, at present it is unclear how far the catalytic domain of the Tn3 resolvase contributes to the sequence specificity.

In published experiments, the central sequence of the resI target site of Tn3 resolvase remained unchanged (Table 1). Therefore, future experiments will have to demonstrate whether this sequence is essential for resolvase activity.

Applications

Several applications require the site-specific integration of genes into eukaryotic genomes. Here we have listed three illustrative examples that are current issues in our laboratories.

Expression of proteins in milk

Large amounts of pharmaceutically relevant proteins can be expressed in the milk of ruminants [94]. Even complicated multi-component proteins are correctly folded and carry the expected glycosylation and phosphorylation patterns. However, the generation of transgenic ruminants is expensive. A major contributor to this fact is that microinjection technology leads to the random integration of the transgene construct. The integration site often exerts a dominant and silencing effect on transgene expression. Therefore, a large number of transgenic lines need to be screened to identify one that expresses the transgene at sufficient levels. However, if the transgene encoding the relevant protein would be inserted into a highly expressed milk protein gene [95], it would be equipped with all the regulatory elements that warrant abundant expression in the lactating mammary gland. In addition, the transgene would be inserted as a single copy, thereby avoiding epigenetic silencing mechanisms that are often induced by multi-copy transgene arrays. Third, if the insertion site is chosen carefully, the transgene could replace an endogenous milk protein gene – not all casein proteins are required for milk secretion [96]. Therefore, the transcriptional and translational capacity, which is taken up by a dispensable casein gene, could be used to support expression of a transgene [97]. This approach would also make the modified milk less complex in its protein content and alleviate purification of the pharmaceutically relevant protein from milk. Currently, the precise integration of genes into ruminant genomes requires a homologous recombination step in fetal cells. Subsequently, the targeted cells can be used as the basis for nuclear transfer into suitable oocytes to generate transgenic ruminants. Although homologous recombination into genes that are expressed in fetal fibroblasts is possible [98], genes that are not expressed (like casein genes) are extremely difficult to target. We could not detect a single targeting event at the ovine β-casein gene in 720 foetal fibroblast clones (J. McWhir and A.F. Kolb, unpublished). The use of DNA-modifying enzymes with defined sequence specificities could provide an alternative to the homologous recombination approach in this scenario. Similar considerations also apply to the development of cell line-based protein expression systems.

Generation of pathogen-resistant mosquito strains

Transposons have been used successfully as transgene vectors in a variety of species including mouse, zebrafish, *Caenorhabditis elegans*, silk worms, *Drosophila* and mosquitoes [40,99,100]. In some species (e.g. *Drosophila*) the use of transposable elements is an established laboratory technique, whereas in other species the transposon-mediated production of transgenic lines is still difficult and not widespread [101]. Transposons are valuable transgene vectors to investigate, for example, the control of mosquito gene expression [102] or to generate mosquito lines, which are refractory to pathogen infection and/or transmission [103]. Active transposons could be used to spread genes in populations for the purposes of population replacement or modification, including the spread of a refractory gene to prevent pathogen transmission [104]. The use of DBD-transposase fusions might increase transposition rates and would also result in the integration of the transgene into known and selected genomic sites. This could mitigate some of the concerns associated with the use of 'unrestrained' autonomous transposons integrating at random locations throughout a host genome. Furthermore, the use of a site-specific transposon might reduce concerns over the potential for horizontal transmission, given that the chosen target sequence might not exist in non-host genomes. Even if specific target sites do exist in non-target organisms, the results of horizontal transmission could then be predicted more accurately based on the now known likely sites of integration. This would help guide risk management strategies for these new technologies. If the refractory mechanism involved the silencing of an endogenous gene, the use of an artificial DBD-transposase element could also be useful because the autonomous element would continue to insert at the specific gene on any un-modified chromosomes, ensuring complete gene knockout.

Non-viral gene therapy

The most obvious application of targeted genome modification would be in the area of gene therapy. Current vectors, including retroviruses, AAV and adenoviruses, do not usually lead to a complete and life-long cure of a genetic deficiency. Rather, the procedure needs to be repeated because: (i) cells that express the therapeutic gene are lost from the body; (ii) the viral DNA only has a limited life-time in an episomal state; or (iii) the therapeutic gene is silenced by epigenetic mechanisms [105]. In the case of viral vectors, re-application often causes a pronounced immune response, which can in itself be a risk factor [106], but might also diminish the efficacy of the gene therapy treatment. The generation of a non-viral gene delivery and integration system would overcome many of these limitations because the most immunogenic proteins associated with viral vectors would no longer be required. Gene therapy vectors based on DNA-modifying enzymes could use liposomes as a delivery tool, thereby minimizing immunogenicity. Site-directed integration into a defined site, which ensures continuous expression of the therapeutic gene, could provide an optimized supply of the therapeutic protein. In addition, integration of the functional therapeutic gene into the locus of the mutated gene could provide the foreign DNA with all the required regulatory signals but would also enable the therapy of genes, which exert a dominant negative phenotype. In addition, vectors based on DNA-modifying enzymes could be produced at a

Table 2. Advantages and disadvantages of different gene integration systems

Integration system	Advantages	Disadvantages	Applications	Use
Retrovirus	Efficient and stable integration	Integration not site-specific, only in dividing cells, often silenced by epigenetic mechanisms	Research, gene therapy (in clinical trials)	Frequent, in many laboratories
Lentivirus	Efficient and stable integration, also in dividing cells, can escape epigenetic gene silencing	Integration not site-specific, HIV-based vectors associated with disease	Research, gene therapy (in clinical trials)[a], animal transgenesis	Frequent, in many laboratories
AAV	Site-specific integration into transcriptionally active chromatin, not associated with any disease	Most AAV based vectors are only episomal due to lack of rep protein	Research, gene therapy (in clinical trials)[b]	Frequent, in many laboratories
Sleeping beauty	Stable integration	Integration not site-specific, not as efficient as retroviral vectors	Research, gene therapy (clinical trials planned)	In several laboratories
Cre	High site-selectivity	Reversible reaction, requires major modifications to interact efficiently with endogenous sites of the human genome	Research	Frequent, in many laboratories
ϕC31	Unidirectional recombinase, stable integration into sites present in the human genome	Promiscuous activity, not as efficient as retroviral vectors	Research	In several laboratories

[a]Data from [109].
[b]See http://www.targen.com/, http://www.parkinson.org/ and http://www.avigen.com/.

large scale from few purified components making gene therapy both more cost effective and safer [41,64].

Conclusions

Stable transgene insertions into host genomes can be catalyzed by several available viral, transposon and recombinase systems (Table 2). AAV and especially retroviruses provide a high frequency of integration and could, therefore, be useful for gene therapy and the generation of transgenic animals. However, the lack of site-specificity of integration poses several problems. First, random integration of therapeutic genes can lead to insertion mutagenesis, which can result in cell transformation. Second, epigenetic mechanisms often silence foreign DNA, which has been inserted at random sites in the genome. Third, viral vectors often have a limited packaging capacity. Finally, non-specifically integrated therapeutic genes are unable to correct dominant negative genetic defects. Huge

Table 3. Integration frequency in three gene transfer systems

Integration system	Cell system	Integration frequency	Refs
Lentivirus	Human primary CD4+ lymphocytes (purified from blood)	99% of cells stably transformed	[109]
Sleeping beauty	HeLa cells (human)	10% of transfected cells stably transformed in the presence of transposase; 0.1% random integration	[43]
Sleeping beauty	Primary human keratinocytes	17.5% of non-selected cells express a b-gal reporter gene in the presence of transposase; 0.3% of non-selected cells express a b-gal reporter gene in the absence of transposase	[107]
ϕC31	HEK293 cells (human) carrying an attP target site	0.1% of transfected cells stably transformed in the presence of recombinase; 0.01% random integration	[87]
ϕC31	HEK293 cells (human) no attP site present, integration occurs at ψ att sites	0.07% of transfected cells stably transformed in the presence of recombinase; 0.008% random integration	[87]
ϕC31	3T3 cells (mouse) carrying an attP target site	4.5% of transfected 3T3 mouse cells stably transformed in the presence of recombinase; 0.35% random integration	[87]
ϕC31	3T3 cells (mouse) no attP site present, integration occurs at ψ att sites	1.2% of transfected 3T3 mouse cells stably transformed in the presence of recombinase; 0.25% random integration	[87]
ϕC31	Primary human keratinocytes	15% of non-selected cells express a b-gal reporter gene in the presence of recombinase; 0.5% of non-selected cells express a b-gal reporter gene in the absence of recombinase	[89]

progress has been made over past five years in defining the interactions of DNA-modifying enzymes with their DNA targets at a biochemical and structural level. Integration site preferences for several enzymes have been documented at a genome wide level. Vectors based on site-specific recombinases and transposases have shown considerable efficacy of gene transfer *in vitro*. Several promising strategies have been developed for the alteration of site-selectivity of DNA-modifying enzymes (Figure 4).

Nevertheless, two crucial questions remain. First, how well will these novel vectors work in a realistic gene therapy scenario? Second, how safe will they be? Enzymes like ϕC31 recombinase or SB10 clearly improve the rate of stable gene integration above the level of random integration and have also been tested successfully in animal experiments [89,107]. Combinations of viral vectors and transposases have been demonstrated to link high transduction efficiency with DNA integration capability [108]. Integration efficiencies of up to 15% of non-selected cells were obtained with both the sleeping beauty and the ϕC31 system. These are significant integration rates although no non-viral DNA integration system has yet reached the insertion efficiency of retroviral vectors [109] (Table 3). However, both sleeping beauty and ϕC31 recombinase are promiscuous in their target site selection and still carry a risk of eliciting insertional mutagenesis. Therefore, at present their advantage over the retrovirus system lies in reduced immunogenicity and the potential of large scale production of the DNA transfer system.

Changes in site-selectivity have been introduced in various enzymes via several methods (Figure 4) including site-directed mutagenesis, random mutagenesis or exchange or the addition of a heterologous DBD. An ideal enzyme system for the targeted modification of eukaryotic genomes would combine high site-selectivity (similar to Cre), high efficiency (similar to retroviral vectors) and substantial packaging capacity.

The site-specificity of retroviral integrases can be altered but incorporating the modified enzymes into viral vectors has remained difficult. Cre and Flp were successfully mutated to recognize target sites that deviate from their wild-type recognition sequences in crucial positions (Table 1). However, reactions catalyzed by Cre or Flp are reversible and to use these enzymes for targeted gene transfer into mammalian genomes, two adjacent sites would need to be targeted to enable a cassette exchange approach. This strategy would require the development of two complementary mutants of the tyrosine recombinases. *In vitro* mutagenesis approaches were also successful in generating ϕC31 recombinase and Tn5 transposase mutants with altered site-specificity. Surprisingly, site-selectivity can, in some cases, be re-directed by the simple addition of sequence specific DBDs. The experiments conducted with IS30 transposase suggest that site-specificity of the resulting fusion proteins might correlate with binding affinity of the DBD. Zinc-finger proteins could provide the most promising DBDs for such an approach because they combine high site-selectivity with excellent binding affinity [1].

Past experience in the area of gene therapy shows that any benefit needs to be balanced carefully against

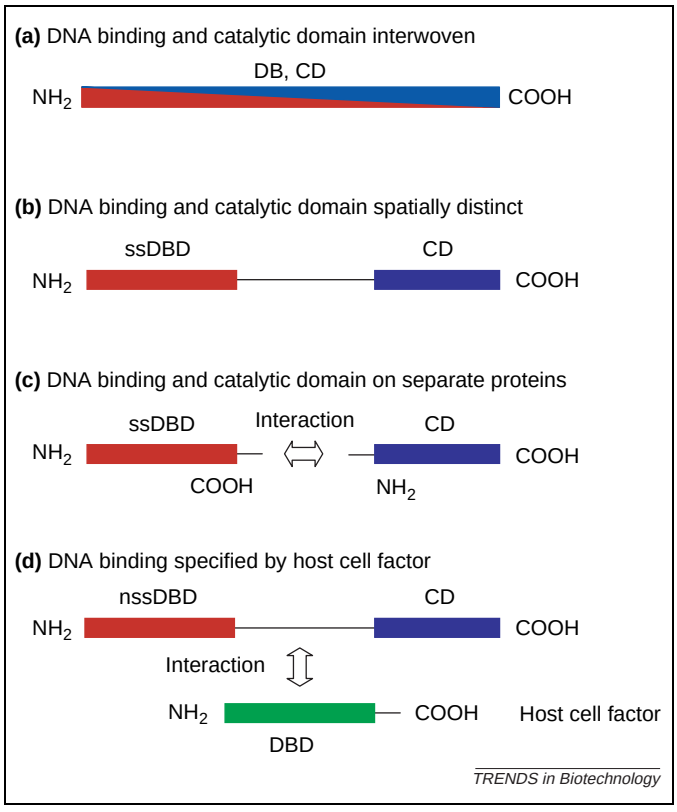

Figure 4. Altering the site specificity of DNA-modifying enzymes. DNA-binding domains (DBDs) are shown in red and catalytic domains (CDs) are shown in blue. **(a)** If the DBD is closely interwoven with the catalytic domain (e.g. Cre or Tn5 transposase) site specificity can be modulated by site-directed mutagenesis of amino acids that make contact with the DNA, or by random mutagenesis approaches (e.g. *in vitro* evolution and DNA shuffling) with subsequent selection of mutants with a desired DNA sequence specificity. **(b)** If a sequence-specific DBD (ssDBD) and the catalytic domain (CD) are spatially distinct, site specificity can be modulated by exchanging the DBD for a DNA-binding module with a different specificity. **(c)** If the ssDBD and the catalytic domain are located on two different proteins, the domains that mediate the interaction between the two proteins need to be maintained but could be linked to alternative DNA-binding modules or catalytic domains to alter site specificity. **(d)** In some cases, host proteins with which the DNA-modifying enzyme interacts specify the site-preference for integration. In this case, the interaction domain could be modified such that alternative host proteins with different site-specificities act as binding partners. Abbreviation: nssDBD, non-sequence-specific DNA-binding domain.

potential risks. The development of gene therapy vectors that target a defined site in the host genome will clearly improve the safety of genetic medicines. However, DNA-modifying enzymes have the potential to cleave host DNA or to mobilize endogenous host sequences (like transposons). This aspect, which is beyond the scope of this article, will have to be addressed in future experiments. The available data suggest that it is possible to derive DNA-modifying enzymes with novel site-specificities by using sophisticated mutagenesis and selection systems. Often, only a few amino acid changes are required to modify sequence specificity. This clearly demonstrates that the development of gene therapy vectors and transgenesis systems with improved efficiency and an enhanced safety profile is a technical possibility.

References

1 Kolb, A.F. *et al*. (2005) Site-directed genome modification: nucleic acid and protein modules for targeted integration and gene correction. *Trends Biotechnol*. doi: 10.1016/j.tibtech.2005.06.005

2 Reznikoff, W.S. (2003) Tn5 as a model for understanding DNA transposition. *Mol. Microbiol.* 47, 1199–1206
3 Craig, N.L. (1997) Target site selection in transposition. *Annu. Rev. Biochem.* 66, 437–474
4 Akopian, A. *et al*. (2003) Chimeric recombinases with designed DNA sequence recognition. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8688–8691
5 Christ, N. and Droge, P. (2002) Genetic manipulation of mouse embryonic stem cells by mutant lambda integrase. *Genesis* 32, 203–208
6 Davies, D.R. *et al*. (2000) Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* 289, 77–85
7 McTaggart, S. and Al-Rubeai, M. (2002) Retroviral vectors for human gene delivery. *Biotechnol. Adv.* 20, 1–31
8 Chen, W.Y. *et al*. (1997) Reactivation of silenced, virally transduced genes by inhibitors of histone deacetylase. *Proc. Natl. Acad. Sci. U. S. A.* 94, 5798–5803
9 Kafri, T. (2004) Gene delivery by lentivirus vectors an overview. *Methods Mol. Biol.* 246, 367–390
10 Hofmann, A. *et al*. (2003) Efficient transgenesis in farm animals by lentiviral vectors. *EMBO Rep.* 4, 1054–1060
11 Sandrin, V. *et al*. (2003) Targeting retroviral and lentiviral vectors. *Curr. Top. Microbiol. Immunol.* 281, 137–178
12 Sadelain, M. (2004) Insertional oncogenesis in gene therapy: how much of a risk? *Gene Ther.* 11, 569–573
13 Holmes-Son, M.L. and Chow, S.A. (2002) Correct integration mediated by integrase-LexA fusion proteins incorporated into HIV-1. *Mol. Ther.* 5, 360–370
14 Daniel, R. *et al*. (2004) Evidence that stable retroviral transduction and cell survival following DNA integration depend on components of the nonhomologous end joining repair pathway. *J. Virol.* 78, 8573–8581
15 Wang, J.Y. *et al*. (2001) Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J.* 20, 7333–7343
16 Appa, R.S. *et al*. (2001) Role of the nonspecific DNA-binding region and alpha helices within the core domain of retroviral integrase in selecting target DNA sites for integration. *J. Biol. Chem.* 276, 45848–45855
17 Wu, X. and Burgess, S.M. (2004) Integration target site selection for retroviruses and transposable elements. *Cell. Mol. Life Sci.* 61, 2588–2596
18 Maxfield, L.F. *et al*. (2005) Relationship between retroviral DNA-integration-site selection and host cell transcription. *Proc. Natl. Acad. Sci. U. S. A.* 102, 1436–1441
19 Bushman, F. (1994) Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. *Proc. Natl. Acad. Sci. U. S. A.* 91, 9233–9237
20 Bushman, F. and Miller, M. (1997) Tethering human immunodeficiency virus type 1 preintegration complexes to target DNA promotes integration at nearby sites. *J. Virol.* 71, 458–464
21 Katz, R.A. *et al*. (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
22 Holmes-Son, M.L. and Chow, S.A. (2000) Integrase-LexA fusion proteins incorporated into human immunodeficiency virus type 1 that contains a catalytically inactive integrase gene are functional to mediate integration. *J. Virol.* 74, 11548–11556
23 Tan, W. *et al*. (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
24 Linden, R.M. *et al*. (1996) The recombination signals for adeno-associated virus site-specific integration. *Proc. Natl. Acad. Sci. U. S. A.* 93, 7966–7972
25 Dutheil, N. *et al*. (2000) Adeno-associated virus site-specifically integrates into a muscle-specific DNA region. *Proc. Natl. Acad. Sci. U. S. A.* 97, 4862–4866
26 Lamartina, S. *et al*. (2000) Characteristics of the adeno-associated virus preintegration site in human chromosome 19: open chromatin conformation and transcription-competent environment. *J. Virol.* 74, 7671–7677
27 Bakowska, J.C. *et al*. (2003) Targeted transgene integration into transgenic mouse fibroblasts carrying the full-length human AAVS1 locus mediated by HSV/AAV rep(+) hybrid amplicon vector. *Gene Ther.* 10, 1691–1702
28 Yoon, M. *et al*. (2001) Amino-terminal domain exchange redirects origin-specific interactions of adeno-associated virus Rep78 *in vitro*. *J. Virol.* 75, 3230–3239
29 Jang, M.Y. *et al*. (2005) Stable secondary structure near the nicking site for adeno-associated virus type 2 Rep proteins on human chromosome 19. *J. Virol.* 79, 3544–3556
30 Hickman, A.B. *et al*. (2004) The nuclease domain of adeno-associated virus rep coordinates replication initiation using two distinct DNA recognition interfaces. *Mol. Cell* 13, 403–414
31 Smith, R.H. and Kotin, R.M. (2000) An adeno-associated virus (AAV) initiator protein, Rep78, catalyzes the cleavage and ligation of single-stranded AAV ori DNA. *J. Virol.* 74, 3122–3129
32 Weger, S. *et al*. (2002) Topors, a p53 and topoisomerase I binding protein, interacts with the adeno-associated virus (AAV-2) Rep78/68 proteins and enhances AAV-2 gene expression. *J. Gen. Virol.* 83, 511–516
33 Tenenbaum, L. *et al*. (2003) Evaluation of risks related to the use of adeno-associated virus-based vectors. *Curr. Gene Ther.* 3, 545–565
34 Nakai, H. *et al*. (2005) Large-scale molecular characterization of adeno-associated virus vector integration in mouse liver. *J. Virol.* 79, 3606–3614
35 Miller, D.G. *et al*. (2004) Adeno-associated virus vectors integrate at chromosome breakage sites. *Nat. Genet.* 36, 767–773
36 Porteus, M.H. *et al*. (2003) Efficient gene targeting mediated by adeno-associated virus and DNA double-strand breaks. *Mol. Cell. Biol.* 23, 3558–3565
37 Kogure, K. *et al*. (2001) Targeted integration of foreign DNA into a defined locus on chromosome 19 in K562 cells using AAV-derived components. *Int. J. Hematol.* 73, 469–475
38 Ueno, T. *et al*. (2000) Site-specific integration of a transgene mediated by a hybrid adenovirus/adeno-associated virus vector using the Cre/loxP-expression-switching system. *Biochem. Biophys. Res. Commun.* 273, 473–478
39 Ryder, E. and Russell, S. (2003) Transposable elements as tools for genomics and genetics in Drosophila. *Brief Funct. Genomic Proteomic* 2, 57–71
40 Szabo, M. *et al*. (2003) Transposition and targeting of the prokaryotic mobile element IS30 in zebrafish. *FEBS Lett.* 550, 46–50
41 Izsvak, Z. and Ivics, Z. (2004) Sleeping beauty transposition: biology and applications for molecular therapy. *Mol. Ther.* 9, 147–156
42 Luo, G. *et al*. (1998) Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 95, 10769–10773
43 Yant, S.R. *et al*. (2000) Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. *Nat. Genet.* 25, 35–41
44 Rice, P.A. and Baker, T.A. (2001) Comparative architecture of transposase and integrase complexes. *Nat. Struct. Biol.* 8, 302–307
45 Naumann, T.A. and Reznikoff, W.S. (2002) Tn5 transposase with an altered specificity for transposon ends. *J. Bacteriol.* 184, 233–240
46 Rezsohazy, R. *et al*. (1993) The IS4 family of insertion sequences: evidence for a conserved transposase motif. *Mol. Microbiol.* 9, 1283–1295
47 Braam, L.A. and Reznikoff, W.S. (1998) Functional characterization of the Tn5 transposase by limited proteolysis. *J. Biol. Chem.* 273, 10908–10913
48 Yant, S.R. *et al*. (2004) Mutational analysis of the N-terminal DNA-binding domain of sleeping beauty transposase: critical residues for DNA binding and hyperactivity in mammalian cells. *Mol. Cell. Biol.* 24, 9239–9247
49 Izsvak, Z. *et al*. (2002) Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in sleeping beauty transposition. *J. Biol. Chem.* 277, 34581–34588
50 Stellwagen, A.E. and Craig, N.L. (1998) Mobile DNA elements: controlling transposition with ATP-dependent molecular switches. *Trends Biochem. Sci.* 23, 486–490
51 Lee, C.C. *et al*. (1998) DNA binding by the KP repressor protein inhibits *P* element transposase activity *in vitro*. *EMBO J.* 17, 4166–4174
52 Zhu, Y. *et al*. (2003) Controlling integration specificity of a yeast retrotransposon. *Proc. Natl. Acad. Sci. U. S. A.* 100, 5891–5895
53 Plasterk, R.H. *et al*. (1999) Resident aliens: the Tc1/mariner superfamily of transposable elements. *Trends Genet.* 15, 326–332
54 Handler, A.M. (2002) Use of the piggyBac transposon for germ-line transformation of insects. *Insect Biochem. Mol. Biol.* 32, 1211–1220

55 Guimond, N. *et al*. (2003) Patterns of Hermes transposition in *Drosophila melanogaster*. *Mol. Genet. Genomics* 268, 779–790
56 Liu, G. *et al*. (2005) Target-site preferences of Sleeping Beauty transposons. *J. Mol. Biol.* 346, 161–173
57 Yant, S.R. *et al*. (2005) High-resolution genome-wide mapping of transposon integration in mammals. *Mol. Cell. Biol.* 25, 2085–2094
58 Vigdal, T.J. *et al*. (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
59 Yusa, K. *et al*. (2004) Enhancement of Sleeping Beauty transposition by CpG methylation: possible role of heterochromatin formation. *Mol. Cell. Biol.* 24, 4004–4018
60 Zayed, H. *et al*. (2003) The DNA-bending protein HMGB1 is a cellular cofactor of Sleeping Beauty transposition. *Nucleic Acids Res.* 31, 2313–2322
61 Kuduvalli, P.N. *et al*. (2001) Target DNA structure plays a critical role in Tn7 transposition. *EMBO J.* 20, 924–932
62 Sharpe, P.L. and Craig, N.L. (1998) Host proteins can stimulate Tn7 transposition: a novel role for the ribosomal protein L29 and the acyl carrier protein. *EMBO J.* 17, 5822–5831
63 Kuduvalli, P.N. *et al*. (2005) Site-specific Tn7 transposition into the human genome. *Nucleic Acids Res.* 33, 857–863
64 Kaminski, J.M. *et al*. (2002) Design of a nonviral vector for site-selective, efficient integration into the human genome. *FASEB J.* 16, 1242–1247
65 Nagy, Z. *et al*. (2004) Analysis of the N-terminal DNA binding domain of the IS30 transposase. *Mol. Microbiol.* 54, 478–488
66 Groth, A.C. and Calos, M.P. (2004) Phage integrases: biology and applications. *J. Mol. Biol.* 335, 667–678
67 Kolb, A.F. (2002) Genome engineering using site-specific recombinases. *Cloning Stem Cells* 4, 65–80
68 Schlake, T. and Bode, J. (1994) Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. *Biochemistry* 33, 12746–12751
69 Kolot, M. *et al*. (1999) Site-specific recombination in mammalian cells expressing the Int recombinase of bacteriophage HK022. *Mol. Biol. Rep.* 26, 207–213
70 Olivares, E.C. *et al*. (2001) Phage R4 integrase mediates site-specific integration in human cells. *Gene* 278, 167–176
71 Stoll, S.M. *et al*. (2002) Phage TP901-1 site-specific integrase functions in human cells. *J. Bacteriol.* 184, 3657–3663
72 Kwon, H.J. *et al*. (1997) Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science* 276, 126–131
73 Guo, F. *et al*. (1997) Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* 389, 40–46
74 Chen, Y. *et al*. (2000) Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. *Mol. Cell* 6, 885–897
75 Subramanya, H.S. *et al*. (1997) Crystal structure of the site-specific recombinase, XerD. *EMBO J.* 16, 5178–5187
76 Abremski, K. *et al*. (1983) Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. *Cell* 32, 1301–1311
77 Hoess, R.H. and Abremski, K. (1984) Interaction of the bacteriophage P1 recombinase Cre with the recombining site loxP. *Proc. Natl. Acad. Sci. U. S. A.* 81, 1026–1029
78 Waterhouse, P. *et al*. (1993) Combinatorial infection and *in vivo* recombination: a strategy for making large phage antibody repertoires. *Nucleic Acids Res.* 21, 2265–2266
79 Srivastava, V. and Ow, D.W. (2004) Marker-free site-specific gene integration in plants. *Trends Biotechnol.* 22, 627–629
80 Thyagarajan, B. *et al*. (2000) Mammalian. genomes contain active recombinase recognition sites. *Gene* 244, 47–54
81 Schmidt, E.E. *et al*. (2000) Illegitimate Cre-dependent chromosome rearrangements in transgenic mouse spermatids. *Proc. Natl. Acad. Sci. U. S. A.* 97, 13702–13707
82 Santoro, S.W. and Schultz, P.G. (2002) Directed evolution of the site specificity of Cre recombinase. *Proc. Natl. Acad. Sci. U. S. A.* 99, 4185–4190
83 Buchholz, F. and Stewart, A.F. (2001) Alteration of Cre recombinase site specificity by substrate-linked protein evolution. *Nat. Biotechnol.* 19, 1047–1052
84 Voziyanov, Y. *et al*. (2003) Stepwise manipulation of DNA specificity in Flp recombinase: progressively adapting Flp to individual and combinatorial mutations in its target site. *J. Mol. Biol.* 326, 65–76
85 Thorpe, H.M. and Smith, M.C. (1998) *In vitro* site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc. Natl. Acad. Sci. U. S. A.* 95, 5505–5510
86 Groth, A.C. *et al*. (2000) A phage integrase directs efficient site-specific integration in human cells. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5995–6000
87 Thyagarajan, B. *et al*. (2001) Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol. Cell. Biol.* 21, 3926–3934
88 Olivares, E.C. *et al*. (2002) Site-specific genomic integration produces therapeutic Factor IX levels in mice. *Nat. Biotechnol.* 20, 1124–1128
89 Ortiz-Urda, S. *et al*. (2002) Stable nonviral genetic correction of inherited human skin disease. *Nat. Med.* 8, 1166–1170
90 Sclimenti, C.R. *et al*. (2001) Directed evolution of a recombinase for improved genomic integration at a native human sequence. *Nucleic Acids Res.* 29, 5044–5051
91 Hartung, M. and Kisters-Woike, B. (1998) Cre mutants with altered DNA binding properties. *J. Biol. Chem.* 273, 22884–22891
92 Yang, W. and Steitz, T.A. (1995) Crystal structure of the site-specific recombinase gamma delta resolvase complexed with a 34 bp cleavage site. *Cell* 82, 193–207
93 Greisman, H.A. and Pabo, C.O. (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* 275, 657–661
94 Wilmut, I. *et al*. (1991) Production of pharmaceutical proteins in milk. *Experientia* 47, 905–912
95 Kolb, A.F. (2002) Engineering immunity in the mammary gland. *J. Mammary Gland Biol. Neoplasia* 7, 123–134
96 Kumar, S. *et al*. (1994) Milk composition and lactation of beta-casein-deficient mice. *Proc. Natl. Acad. Sci. U. S. A.* 91, 6138–6142
97 McClenaghan, M. *et al*. (1995) Secretory proteins compete for production in the mammary gland of transgenic mice. *Biochem. J.* 310, 637–641
98 McCreath, K.J. *et al*. (2000) Production of gene-targeted sheep by nuclear transfer from cultured somatic cells. *Nature* 405, 1066–1069
99 Rowan, K.H. *et al*. (2004) Tn5 as an insect gene vector. *Insect Biochem. Mol. Biol.* 34, 695–705
100 Dupuy, A.J. *et al*. (2002) Mammalian germ-line transgenesis by transposition. *Proc. Natl. Acad. Sci. U. S. A.* 99, 4495–4499
101 Riehle, M.A. *et al*. (2003) Towards genetic manipulation of wild mosquito populations to combat malaria: advances and challenges. *J. Exp. Biol.* 206, 3809–3816
102 Coates, C.J. *et al*. (1999) Promoter-directed expression of recombinant fire-fly luciferase in the salivary glands of Hermes-transformed Aedes aegypti. *Gene* 226, 317–325
103 Kokoza, V. *et al*. (2000) Engineering blood meal-activated systemic immunity in the yellow fever mosquito, Aedes aegypti. *Proc. Natl. Acad. Sci. U. S. A.* 97, 9144–9149
104 James, A.A. (2003) Blocking malaria parasite invasion of mosquito salivary glands. *J. Exp. Biol.* 206, 3817–3821
105 Swindle, C.S. and Klug, C.A. (2002) Mechanisms that regulate silencing of gene expression from retroviral vectors. *J. Hematother. Stem Cell Res.* 11, 449–456
106 Kochanek, S. *et al*. (1996) A new adenoviral vector: replacement of all viral coding sequences with 28 kb of DNA independently expressing both full-length dystrophin and beta-galactosidase. *Proc. Natl. Acad. Sci. U. S. A.* 93, 5731–5736
107 Ortiz-Urda, S. *et al*. (2003) Sustainable correction of junctional epidermolysis bullosa via transposon-mediated nonviral gene transfer. *Gene Ther.* 10, 1099–1104
108 Yant, S.R. *et al*. (2002) Transposition from a gutless adeno-transposon vector stabilizes transgene expression *in vivo*. *Nat. Biotechnol.* 20, 999–1005
109 Lu, X. *et al*. (2004) Safe two-plasmid production for the first clinical lentivirus vector that achieves >99% transduction in primary cells using a one-step protocol. *J. Gene Med.* 6, 963–973
110 Rost, B. *et al*. (1997) Protein fold recognition by prediction-based threading. *J. Mol. Biol.* 270, 471–480