Directed evolution of an HIV-1 LTR specific recombinase for anti-retroviral therapy- a proof of concept study.

DISSERTATION

zur Erlangung des akademischen Grades

Doctor rerum naturalium

(Dr. rer. nat.)

vorgelegt

der Fakultät Mathematik und Naturwissenschaften

der Technischen Universität Dresden

von

INDRANI SARKAR

geboren am 03.12.1974, Kolkata, INDIA.

Gutachter: Dr. Frank Buchholz
Prof. Dr. Francis Stewart
Prof. Dr. Joachim Hauber

Eingereicht am 14.06.2006

Tag der Verteidigung:

Acknowledgements

First and foremost I would like to thank my supervisor, Frank Buchholz for giving me the opportunity to work in his group and for motivating and supporting me through all these years. I am extremely grateful to him for having unlimited patience especially during the long evolution phase of this project and for always giving me helpful advice. He has been very supportive and reassuring in every situation. I am also thankful to him for the great lab retreats he organized. I would like to thank my TAC committee member Francis Stewart for his helpful suggestions. I am grateful to Joachim Hauber and Ilona Hauber (Heinrich-Pette Institute, Hamburg) who collaborated with me on this project and helped me to develop this project and provided me with the HIV-1 LTR based vectors and cell lines. I would like to thank Bianca Habermann for modeling of mutations and also Mayte Pissabarro with whom I had first looked into the recombinase mutations.

A special thanks to all my lab members who made the lab a nice place to work in and also MPI-CBG for the friendly and cooperative atmosphere.

I thank all my friends here who made Dresden warm and fun filled inspite of the cold and dark winters, especially Lawrence Rajendran for always being there through thick and thin, for always reaching out at not so great times and for all the wonderful meals, Aparna Sapra and Vineeth Surendranath for all the great coffee sessions and discussions and get togethers, Jonas Reis and all others with whom I had a great time and felt at home away from home.

Last but not the least I thank my parents for their support and understanding and Angshuman for holding on so strong even at the most difficult times.

I herewith declare that I have produced this paper without the prohibited assistance of third parties and without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such. This paper has not previously been presented in identical or similar form to any other German or foreign examination board.

The thesis work was conducted from *1st June 2002* to *14th June 2006* under the supervision of Dr.Frank Buchholz at *Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG), Dresden, Germany.*

Indrani Sarkar
Dresden, Germany.

1. Summary

The prospect of the work presented in this thesis has been to engineer Cre recombinase to recognize and recombine a sequence from an HIV-1 Long Terminal Repeat (LTR), characterize the recombination proficiency of the evolved recombinase in mammalian cells and explore the potential of the recombinase for a novel antiretroviral strategy.

Retroviruses like HIV-1 integrate into the host chromosome and persist as a provirus flanked by LTRs. Establishment of the provirus is an obligatory step in the retroviral life-cycle and serves to maintain viral sequences in the infected cell. The LTRs of proviral DNA direct the expression of viral genes.

A modified Cre recombinase recognising sequences within each LTR of proviral DNA could mediate excision of the viral coding sequences eliminating the intact provirus and represents a novel direct antiviral strategy targeting the provirus. To realize the potential of a genetic surgery targeting an integrated provirus, it is essential to determine whether such a recombinase can actually be generated that can recombine a sequence within a retroviral LTR. The laboratory evolved Cre recombinase variants with altered site-specificity have proved that it is possible to generate tailored recombinases and hold promise for increasing the applicability of these recombinases for therapeutic purposes.

An *in vitro* evolution strategy called substrate-linked protein evolution strategy (SliDE) was used to evolve a novel Cre recombinase variant recombining an asymmetric sequence termed loxltr belonging to an HIV-1 LTR in *E.coli*.

To obtain the novel Cre variant, an extensive evolution approach was taken, traversing through six pallindromic loxltr subsets, to achieve specificity for the asymmetric loxltr. A novel feature of this strategy was the use of a combinatorial SLiDE by splitting the final loxltr target site into six sub-sets to achieve recombination at the loxltr target. It was possible to generate functional recombinases that recognize the asymmetric loxltr by combining recombinases which separately recognize the loxltr subsets with different mutations all present in the final substrate using DNA shuffling. The synergistic effect of the shuffled libraries of the lower subsets led to the generation of recombinases which recombined the higher loxltr sub-sets and also the final target loxltr. A total of 126 evolution cycles were performed through all the subsets to achieve recombination at the final target loxltr.

Sequence analysis of the evolved recombinase variants showed the sequestering of specific mutations responsible for conferring the change in substrate specificity through the course of evolution. These mutations were modeled onto the wild type Cre crystal structure to get an understanding of the interactions that the mutated amino acids might have with the loxltr target.

To investigate whether the evolved Cre variant, termed Tre, recombines its target in mammalian cells, a lacZ based reporter assay was developed in Hela cells. The recombinase mediated recombination in both transient and stable assays.

To improve the activity of Tre in mammalian cells, the codon usage of the recombinase was adapted to human codon usage preferences by introducing silent base mutations. The codon-optimised recombinase, hTre was tested in Hela cells using the lacZ based reporter assay. In comparison to the unoptimised version, hTre displayed improved recombination potential at the loxltr target both in transient assays and also with the target stably integrated in the genome.

To explore whether recombinases like hTre can actually be applied for an antiretroviral therapy targeting an integrated provirus Tat transactivation assays were performed in an HXB3-LTR based luciferase and chloramphenicol acetyl transferase reporter vectors with the loxltr cloned within the LTRs to investigate whether hTRE can recombine the loxltr site in the context of a real HIV-1 LTR.

Transfection of hTre expression vector in presence of Tat resulted in repression of Tat mediated transactivation of the reporter activity as a result of recombination at the loxltr sites and three-fold reduction of reporter activity was obtained both for the luciferase and CAT vectors.

To analyse the extent of recombination mediated by hTre at the loxltr target integrated in the genome, a PCR based assay was used in Hela cells with the CAT construct stably integrated into the genome. The evolved recombinase mediated recombination to the extent of around 10% as detected by the PCR based assay in the CAT stable cell line. In conclusion this work demonstrates and provides proof of concept that directed evolution can be utilized to develop tailored molecular scissors for therapeutic DNA surgery.

2. Introduction

2.1. Site-specific recombination.

Genetic recombination provides a mechanism for generating genetic diversity beyond that achieved by the independent segregation of chromosomes. Genetic exchange by recombination occurs not only in animals and plants but also in prokaryotes, viruses, plasmids, and even in the DNA of cell organelles such as mitochondria.

Homologous recombination occurs between two homologous DNA segments, and there is relatively little specificity as to the site at which the actual crossover occurs.

Transposition, by contrast, is a process that leaves breaks in DNA that require repair. Transposases cleave endonucleolytically at each transposon end. The 3'OH ends act as the nucleophiles for the target cleavage/joining reaction, the 5' ends of donor DNA flanking the transposon remain free, until degraded or acted upon by cellular repair enzymes. The transposable elements are capable of insertion at several different chromosomal locations and have relatively undefined target site.

Site-specific recombination is distinct from homologous recombination and transposition. Distinguishing features of conservative site-specific recombination (CSSR) are that recombination occurs at specific sites within a short region of sequence identity shared by the participating DNA segments, strand exchange occurs by precise breakage and joining events and recombination is reciprocal.

2.2. The λ- integrase family of recombinases.

Site-specific recombination was first discovered through studies of the formation of lysogens by bacteriophage λ. During the establishment of lysogeny a circular form of bacteriophage λ integrates into the bacterial chromosome by intermolecular recombination between a specific site in the phage *attP* and a specific site *attB* on the bacterial chromosome. Integration requires the phage integrase protein and *E.coli*

encoded protein Integration Host Factor (IHF). Integrase also catalyzes the reverse reaction, excision of the circular λ phage DNA from a bacterial chromosome.

Based on sequence similarity, recombinases from bacteria and yeast have been classified into the intergrase family, named after the prototypical phage λ, and resolvase-invertase family, named after the cointegrate-resolving protein encoded by the transposons γδ and Tn3. These two families are not related in protein sequence and employ different mechanisms of recombination. The resolvase family includes most transposon-encoded resolvases and DNA invertases such as Hin and Gin. The λ-integrase family is also referred to as " tyrosine recombinases" (Sherratt and Wigley 1998) which include over 100 members identified on the basis of sequence similarity. The most well studied examples include the integrase protein from bacteriophage λ (Landy 1989), the bacterial XerC and XerD recombinases (Colloms et al. 1990; McCulloch et al. 1994; Colloms et al. 1996), the Cre recombinase from bacteriophageP1 (Abremski and Hoess 1984) and the Flp recombinase from *Saccharomyces cerevisiae* (Sadowski PD, 1995). Site-specific recombinases (SSR) from the lambda integrase family of enzymes catalyze DNA rearrangements that are critical for a variety of important biological functions including the resolution of multimeric plasmids and chromosomes to monomers in order to ensure faithful segregation upon cell division, antigenic phase variation, dissemination of antibiotic- and antiseptic- resistance genes, the amplification of yeast 2μ circle copy number, and the regulation of gene expression (Nunes-Duby, Matsumoto, and Landy 1987; Landy 1993; Esposito and Scocca 1997; Van Duyne 2001).

The defining feature of the λ integrase family is a highly conserved tyrosine nucleophile in combination with a triad of basic amino acids, the arginine-histidine-arginine triad. These residues are absolutely essential for full recombination activity.

λ integrase and its relatives make sequential and ordered pairs of single strand exchanges between two recombinational partners; the first pair of exchanges form a four way Holliday junction and the second pair resolve the junction to complete the recombination. The conserved Tyrosine residue in the recombinase acts as the nucleophile that forms 3'-phosphotyrosine linkages to DNA releasing free 5' hydroxyl groups which then attacks the 3' phosphotyrosine linkages of the partner strands to form a Holliday junction intermediate (Argos et al. 1986; Pargellis et al. 1988; Abremski and Hoess 1992; Kwon et

al. 1997). The mechanistic understanding of the λ integrase family members has greatly improved with the three dimensional structures of varying complexity published for five tyrosine recombinases: Cre (Guo, Gopaul, and van Duyne 1997; Gopaul, Guo, and Van Duyne 1998; Woods et al. 2001; Martin et al. 2002), XerD (Subramanya et al. 1997), HPI integrase (Hickmann et al., 1997), Flp (Chen et al. 2000) and λ-integrase (Wojciak et al. 2002). These structures have revealed a conserved catalytic domain fold.

The Cre recombinase from bacteriophage P1 is one of the best characterized Int family member and have been successfully used for various elegant genome manipulation strategies.

2.3. The Cre-*loxP* system

The Cre recombinase is a 38 kDa protein encoded by the bacteriophage P1. Its role in P1 life cycle are believed to include resolution of dimeric chromosomes formed following DNA replication into monomeric P1 DNAs and also the cyclisation of the linear genome (Sternberg et al. 1986). The DNA sequences where Cre recombinase binding and strand exchange take place are named *loxP* (locus of crossover in P1) (Hoess, Ziese, and Sternberg 1982). The *loxP* site is a 34 bp sequence composed of two pallindromic recombinase binding elements (RBE) or core binding elements surrounding a central 8 bp strand exchange or crossover region (Fig.1).

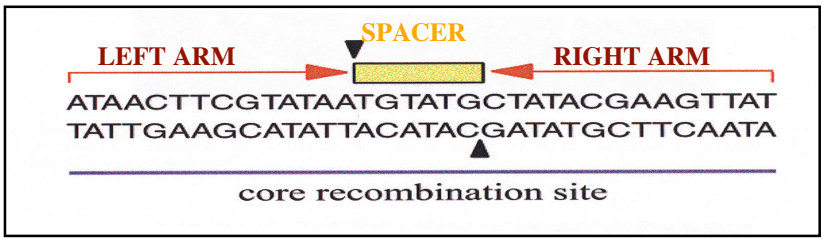


Fig.1. Structure of the *lox*P recognition sequence. The black arrowheads indicate the strand cleavage positions within the spacer.

Two recombinase molecules bind to each core site with a high level of cooperativity (Blakely et al. 1993; Ringrose et al. 1998). The phosphoryl transfer and strand exchange occurs within the central spacer or crossover region. The crossover sequence is

asymmetric and provides directionality to the site. The Cre recombinase does not require any accessory factor to mediate recombination unlike more complex systems like λ-integrase and XerC/D recombinases.

2.4. Mechanism of Cre mediated site-specific recombination.

The Int family of tyrosine recombinases of which Cre is a member and also the type 1B topoisomerases are characterized by the presence of a conserved tyrosine nucleophile, an active site structure and protein fold (Argos et al. 1986; Esposito and Scocca 1997; Nunes-Duby et al. 1998; Sherratt and Wigley 1998; Cheng et al. 2000).

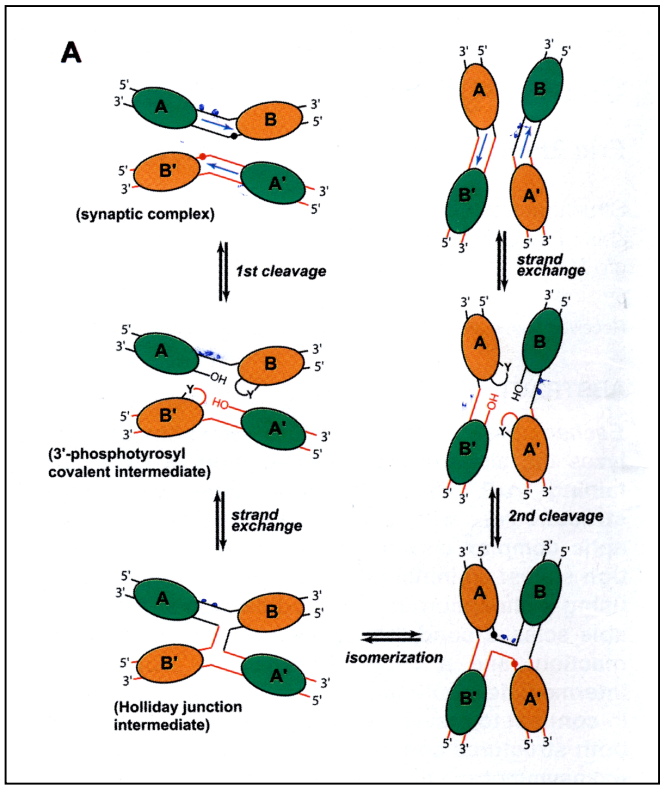


Fig.2. Mechanism of site-specific recombination. Two Cre proteins bind on either side of each of two loxP sites, and then associate together, forming a synapse composed of two DNA molecules and four Cre proteins. A tyrosine (Tyr) residue on opposite Cre proteins cleaves the DNA strand to which the proteins are bound, forming two covalent 3' phosphotyrosine bonds and two free 5'-hydroxyl ends. The free 5'-hydroxyl ends in this Cre-DNA intermediate then react with the opposite phosphotyrosine bonds, yielding a Holliday-type intermediate and regenerating native Cre proteins. Next, a tyrosine residue on the other two

Cre proteins at the synapse cleave the DNA strands to which they are bound, again forming covalent 3' phosphotyrosine bonds and free 5'-hydroxyl ends to give a second Cre-DNA intermediate. Finally, the free 5'hydroxyl ends of this intermediate attack the opposite phosphotyrosine bonds, yielding the recombinant DNA products still bound to regenerated native Cre proteins. [Adapted from F. Guo, D. Gopaul, and G. Van Duyne, 1997, *Nature* **389**:40.]

As shown in Fig.2 the basic mechanism includes a first set of strand cleavage and ligation producing a Holliday junction intermediate and this is then resolved by a second set of strand cleavage and rejoining to yield recombinant products. The 8 bp asymmetric spacer divides the loxP into two 13 bp domains, the left arm and right arm (Fig.1). The DNA cleavages occur at the scissile sites between the first and second bases of the spacer, that is between A and T of the top strand or between G and C of the bottom strand. The interior six nucleotides are exchanged during the recombination reaction.

The main mechanistic features of the Int site-specific recombination mechanism have been defined by crystal structures of Cre-lox complexes (Guo, Gopaul, and van Duyne 1997; Gopaul, Guo, and Van Duyne 1998; Guo, Gopaul, and Van Duyne 1999; Woods et al. 2001; Martin et al. 2002; Ennifar et al. 2003). The crystal structures of the synaptic complexes formed between a symmetric loxP site and two Cre mutants that are defective in strand cleavage have revealed that the DNA bent sharply at end of the crossover or spacer region and that the bend direction appeared to determine which of the stands in the DNA duplex will be cleaved first (Guo, Gopaul, and Van Duyne 1999). There remains a controversy regarding the order of strand exchange in Cre mediated recombination. Analysis of in vivo recombination products suggested that cleavage was initiated on the lower strand (Hoess and Abremski 1985), and this view was restated later (Lee and Saito 1998); (Lee and Sadowski 2001). On the other hand, strand composition analysis of the Cre/lox Holliday junction intermediate revealed that Cre initiates recombination by cleaving the upper strand at the scissile A residue at the left arm, and preferential association of the cleaving or active Cre subunit to left arm dictates first strand exchange and ensures productive orientation of loxP sites in the recombination synapses (Martin et al. 2002). Recently two crystal structures of the full tetrameric Cre-loxP synaptic complex has been reported (Ennifar et al. 2003). The pre-cleavage synaptic complex with a phosphorothioate modification at the scissile bond also revealed a kink at the TG/AC

residues in left side of the spacer next to the scissile ApT residues and they also showed that the cleavage competent Cre subunit bound to the scissile ApT residues in the top strand. Their studies suggested that the Cre induced kink activates the neighbouring scissile phosphate for cleavage, a result contradictory to that of the Van Duyne lab which proposed that bending at one-half of the spacer stimulates cleavage at the opposite half (Guo, Gopaul, and Van Duyne 1999). The issue of spacer identity has been addressed in an extensive mutational study (Lee and Saito 1998) and it was found that identity of the innermost six bases was important for recombination though their findings suggested that Cre cleaves the bottom strand first. Very recently experiments from the Van Duyne lab (Ghosh et al. 2005) have addressed the issue of order of strand exchange in the context of synaptic complex formed at the beginning of recombination pathway and have shown through biochemical analyses that Cre preferentially cleaves the bottom strands of loxP under conditions favouring synapsis, however, in absence of synapsis, top strand cleavage also occurs but at a lower rate. The above studies reveal that there is a structural basis for the observation that DNA strands are exchanged in a defined order although the intricacies of Cre mediated recombination still remains a puzzle.

2.5. The spacer orientation determines the outcome of recombination.

The asymmetric spacer or core sequence confers directionality to the target site, consequently the relative orientation of the target sites with respect to one another determines the outcome of recombination. The possible recombination outcomes are described in Fig.3. Cre will excise a circular molecule from two directly repeated target sites, invert the DNA between two inverted sites and exchange sequences distal to target sites present on two linear molecules, such as a pair of non-homologous chromosomes.

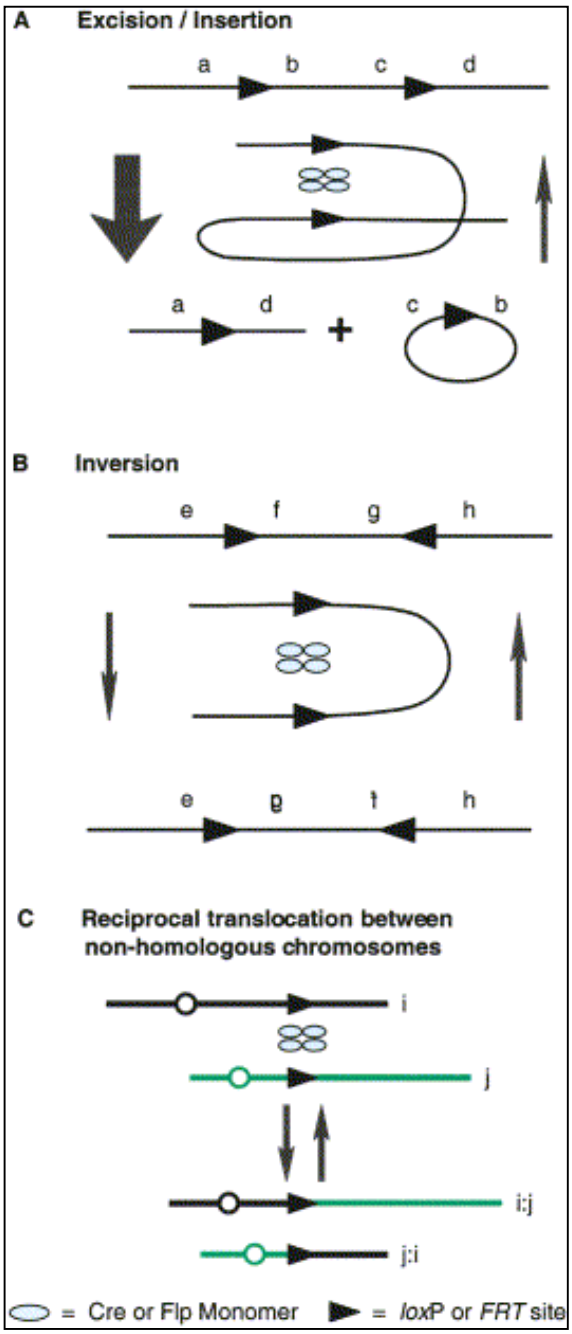


Fig 3. The spacer orientation determines the outcome of recombination.

2.6. Application of site-specific recombinases.

The Cre/loxP system is widely used for genomic manipulations as Cre efficiently recombines loxP sites without the need for any accessory cofactors and have been used effectively to create gene deletions, inversions, insertions and exchanges in exogenous systems like mammalian cell culture and mice. The first indication that this phage enzyme works in eukaryotic cells emerged from in vitro cell culture experiments (Sauer and Henderson 1988; Sauer and Henderson 1990). These experiments were followed by the demonstration that Cre works well in mice and also when expressed from a transgene (Lakso et al. 1992; Orban, Chui, and Marth 1992). The excision reaction being irreversible has been widely exploited and is the basis for conditional gene inactivation and molecular fate mapping that has allowed gene function and cell deployment to be studied in both embryos and adult stages.

1.Tissue-specific knockouts.

Traditional gene targeting strategy has two serious pitfalls. The first being that elimination of the gene activity throughout the entire animal can result in early embryonic lethality which precludes analysis of gene function at later stages and secondly the maintenance of the positive selection cassette within the gene targeted can disrupt neighbouring gene expression due to presence of regulatory elements in the selection cassettes. These problems have been circumvented by employing site-specific recombination to create knock out (KO) mice. In such studies, standard gene-targeting techniques are used to produce a mouse in which an essential region of the gene of interest is flanked by directly repeated loxP sites or floxed, so that tissue specific cre expression results in inactivation of this allele. In most cases the loxP sites are placed in introns, but some have also been placed in 5' or 3' flanking regions as well. Before recombination the conditional allele should have wild- type activity. By crossing a mouse line with a conditional allele to an effector mouse line expressing cre in a tissue-specific manner, progeny are produced in which the conditional allele is inactivated only in those tissues expressing cre. The first report of conditional gene modification examined the effect of a null mutation in the essential polβ gene in T-cells (Gu et al. 1994). Since the

first published results of the tissue-specific KO, the use of this strategy has steadily increased and many diferent tissue-specific cre-expressing mice exist till date.

The presence of positive selection marker like neo can interfere with expression of the floxed allele, compromising the wild type-activity in absence of recombination. This problem had been solved by using three loxP sites such that both the gene of interest and neo are floxed. A partial Cre mediated recombination event removes the neo gene, by transiently introducing a cre expression cassette into ES cells (Gu et al. 1994). This can also be achieved *in vivo* by injecting zygotes with *cre* expression cassette or by breeding mice carrying the floxed allele to mice that mosaically express cre in the germ line (Holzenberger et al. 2000a). Inclusion of the neo gene might generate a hypomorphic allele which in combination with the null and conditional allele can be very informative (Meyers, Lewandoski, and Martin 1998; Nagy et al. 1998). Apart from the tri-lox strategy of removal of selection cassettes dual recombinase strategy has been used to create an allelic series for Fgf8, where Fgf8 was floxed and neo cassette was flrted to allow for Flp-mediated recombination (Meyers, Lewandoski, and Martin 1998).

2.Use of Cre in cell lineage analysis.

Reporter mice in which a histological marker is activated by Cre- mediated recombination can be extremely useful in analyzing cell lineages. Examples of such reporter mice are Cre-activable β-galactosidase (Mao, Fujiwara, and Orkin 1999; Soriano 1999)or the enhanced GFP (Mao et al. 2001) or EYFP and ECFP varieties (Srinivas et al. 2001) in the ROSA26 locus allowing constitutive transgene expression. Two well known Cre reporter lines are the Z/AP (Lobe et al. 1999) and Z/EG (Novak et al. 2000) in which Cre-mediated recombination causes a switch from l*acZ* expression to alkaline phosphatase or EGFP expression, respectively. Such reporter lines are useful for molecular fate mapping studies as they represent a link between gene expression and the fate of specific cell lineages. Fate mapping of mid-hindbrain border (Zinyk et al. 1998), neural crest cell derivatives (Joseph et al. 2004), pancreatic islets (Herrera 2000) and memory T cells (Jacob and Baltimore 1999) was achieved using this technique.

3.Cre-mediated gene activation and tissue-specific gene repair.

Transgene activation by Cre-mediated recombination is achieved by removal of an intervening 'Stop' sequence (Lakso et al. 1992), or by restoring an ORF that is interrupted by floxed sequences. In such cases the remaining loxP sequence after recombination must be tolerated as a part of the open reading frame.

Cre-mediated recombination can be used to achieve tissue- specific gene repair in a design similar to that of a tissue-specific KO experiment. Human skin disorders have been modeled in mice using this approach (Cao et al. 2001; Arin and Roop 2004). In these studies mouse keratin genes containing point mutations as in human disease were silenced by inclusion of a floxed *neo* gene. Topical application of an inducer of Cre activity resulted in activation of the mutant allele and resulted in phenotypes similar to that in disease syndromes.

4.Chromosomal rearrangements in mice to model human disease.

Chromosomal abnormalities are a principal cause of fetal loss and developmental disorders in humans and chromosomal translocations are involved in the genesis of human cancers. Chromosomal rearrangements in mice can thus be used to model these disorders and genetically dissect their causes. The Cre/loxP recombination system has been used to create mouse models to recapitulate human chromosomal rearrangements (Buchholz et al. 2000; Collins et al. 2000). Mouse balancer chromosomes have been developed using Cre/loxP technology by tagging chromosomal inversions with recessive lethal mutations and coat colour markers (Zheng et al. 2000). Deletion chromosomes can be engineered to provide segmental haploidy in the diploid mouse genome. Recessive mutations induced in these deletions can be detected by crossing mutatnt mice to mice hemizygous for different regions of the genome. It was possible to generate megabase genomic rearrangements in mice by using chromosomal engineering technology (Ramirez-Solis, Liu, and Bradley 1995).

Defined chromosomal deletions, duplications and inversions provide powerful tools for mouse functional genomics. In vivo chromosomal engineering can be used to produce genetic mosaics, defined as individuals heterozygous for a particular mutation, but

containing patches of homozygous mutant cells, genetic mosaics are generated when mitotic recombination occurs at G2 and is followed by X-segregation and clones of homozygous mutant daughter cells are produced from heterozygous mothers. Genetic mosaics allow assessment of effect of gene function on clonal cell lineages. Cre/loxP technology was used to reproducibly induce mitotic recombination in mouse ES cells for multiple allelic lox sites after transient *cre* expression (Liu, Jenkins, and Copeland 2002).

5. Application of Cre/loxP technology to recombineering.

Cre/loxP technology has also been combined with recombinogenic engineering or recombineering techniques. Highly efficient phage based *E.coli* homologous recombination systems have been developed which allows genomic DNA in BACs to be modified and sub-cloned without using restriction enzymes and ligases (Zhang et al. 1998; Zhang et al. 2000). This strategy facilitates the introduction of target sites of recombinases to be introduced into the mouse genome. Recombineering technique has been employed to produce gene-targeting constructs to create conditional alleles in which recombinase target sites are positioned at significant distances from each other (Liu, Jenkins, and Copeland 2003; Testa et al. 2003).

2.7. Temporal control of recombinase expression.

The first inducible Cre mouse line used the Mx1 promoter, which can be activated by injecting mice with interferon-α or β or with double-stranded RNA (poly-inosinic-polycytidilic acid) (Kuhn et al. 1995). Such mice have been used to regulate gene expression in the liver and immune system where recombination efficiency is high.

To add temporal control to site-specific recombinase activity, ligand regulated forms of Cre and also Flp have been developed. A mutated estrogen receptor (ER) ligand binding domain (LBD) has been fused to C terminus of Cre (Feil et al. 1996; Kellendonk et al. 1996) or Flp (Logie and Stewart 1995).

Currently there are three different ERs available, all of which are insensitive to endogenous β-estradiol, but are responsive to the synthetic estrogen antagonist 4-hydroxy tamoxifen. The ligand regulated switch for turning on recombinase activity has two major

applications, first, broad expression of recombinase:ER fusions in most tissues to enable temporal control of recombinase activity (Seibler et al. 2003), and secondly, tissue-restricted expression of these fusions are achieved using specific enhancers (Logie and Stewart 1995).

The C-terminus of Cre has also been fused to a mutated progesterone receptor (PR) LBD. The fusion is responsive to the synthetic steroid RU486, but not to endogenous progesterone (Kellendonk et al. 1996), however this system suffers from leakiness and has been improved to confer lower background activity (Wunderlich et al. 2001).

A second mechanism to achieve temporal control of Cre or Flp expression involves the tetracycline responsive system (Gossen and Bujard 1992). Inducing Cre expression with Dox in the rtTA system results in a tighter regulation in comparison to tTA sysyem (Utomo, Nikitin, and Lee 1999; Holzenberger et al. 2000b).

2.8. loxP variants.

Three loxP mutants, lox511, lox514 and lox 512 with single-base substitution in the spacer region were investigated to understand how the asymmetric spacer region confers directionality to the recombination reaction and the importance of spacer homology for directionality. One of these mutants, lox511, was found to be incompatible in recombination with wild-type loxP site and recombined only with like site. Incompatibility is referred to the ability of a *lox* site to discriminate against recombination with another *lox* site not sharing spacer region homology.

This study showed that most bases in the loxP core or spacer can deviate from the wild-type as long as the cores match between recombining lox sites (Hoess, Wierzbicki, and Abremski 1986). Minor changes in the pallindromic half –sites of loxP have also been tolerated by Cre in several cases (Abremski and Hoess 1985; Abremski et al. 1988; Sauer 1992; Albert et al. 1995; Sauer 1996). Further investigation of the functionality of cre on mutant lox sites was carried out using 24 single-base substitution spacer mutants and 30 double-base substitution spacer mutants using in vitro recombination assays. The study demonstrated the exclusivity of two loxP mutants, lox2272 and lox5171 with two-base substitutions, showing efficient recombination among themselves but not cross recombination with loxP site (Lee and Saito 1998).

There are two classes of mutations in the loxP variant sites that can alter the specificity and directionality of recombination reaction. The first class are the spacer variants. As discussed above spacer region homology is important for efficient recombination between directly repeated lox sites in cis. Crystallographic studies of Cre recombinase bound to lox site suggest that base pairing is required between exchanging strands in the recombination complex for efficient recombination (Guo, Gopaul, and van Duyne 1997). The spacer region not only confers directionality but also specificity to recombination reaction. The second class of lox mutations are within the 13 bp palindromes or inverted repeats, two sites *lox71* and *lox66* are examples of such inverted repeat variants. Mutations within the pallindromic arms weaken the affinity of recombinase for its binding site. Recombination between lox sites with mutations in complementary pallindromic arms generates a wild type loxP site and a site with doubly mutant arms. This double mutant no longer serves as a recombination substrate and undergoes no further recombination. This kind of recombination reaction between pallindromic lox mutants is almost exclusively unidirectional. The lox sites containing arm mutations have been used to stably integrate plasmid DNA to both plant and mammalian chromosomes (Albert et al. 1995; Araki, Araki, and Yamamura 2002).

2.9. Applications of loxP variants in genome engineering.

Recombinase-mediated cassette exchange.

Recombinase-mediated cassette exchange (RMCE) is a method to achieve site-specific chromosomal integration at a defined genomic locus. The findings that Cre can tolerate variations in its target sites led to the development of RMCE.

RMCE can be defined as a site-specific recombination reaction that entails the exchange of one DNA cassette for another. It is a two-step procedure. Heterospecific or incompatible lox sites are targeted to the genomic locus of interest by homologous recombination. Action of recombinase inserts a replacement sequence into this pre-tagged site by double reciprocal crossover. Such a method permits re-iterative modification of a particular genomic locus. This insertion reaction exploits the fact that recombination is efficiently mediated between pairs of homotypic, but not heterotypic target sites. Normally a selectable marker flanked by heterospecific lox sites is inserted into a

chromosomal site, then a DNA sequence of choice is inserted into the tagged site by replacing the selectable marker in presence of the recombinase. Mechanistically this process involves two recombination events, an integration step followed by an excision step, leaving heterotypic target sites flanking the exchanged cassette. In RMCE reactions the target sites may be directly or inversely oriented. The advantage of using inverted sites is that cassette excision is prevented should there be promiscuous recombination between pair of heterotypic target sites. RMCE has been adapted to create isogenic cell lines (Bouhassira, Westerman, and Leboulch 1997; Feng et al. 1999; Kolb 2001). Though RMCE proved to be very efficient in bacteria and cell culture, the system suffers from promiscuous recombination between heterospecific target sites leading to deletion of floxed cassette.

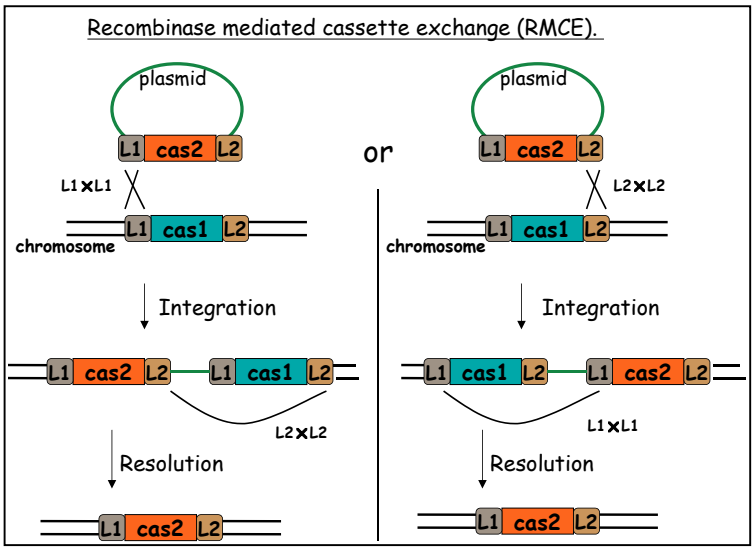


Fig.4. Schematic representation of RMCE. L1 and L2 are a pair of hetero-specific target sites flanking a chromosomally integrated cassestte cas1 to be exchanged with the incoming cassette cas2 flanked by the same pair of sites. Recombination through either L1 or L2 followed by recombinant resolution results in the exchange of chromosomally integrated cas1 by incoming cas2.

The re-excision can be prevented by using inverted target sites, but this requires additional screening to determine the orientation of the exchanged gene. This procedure enabled studies on the orientation dependence of expression parameters at a given

genomic site. Three sites investigated in this way displayed either stable or silencing position effects and demonstrated that a transgene at a given genomic locus can be differentially imprinted depending on its orientation (Feng et al. 2001).

FLEx switch.

Another very recent use of mutant lox511 and wild type loxP site was in a novel strategy called FLEx switch (Schnutgen et al. 2003). The strategy uses the ability of Cre to both invert or excise DNA depending on the orientation of the target sites. This Cre-dependent genetic switch works in a manner such that the expression of a given gene is turned off while the expression of another gene is concomitantly turned on. Stable DNA inversion can be achieved using this strategy, it also involves two recombination events like RMCE, the first is inversion followed by an excision step. The final product can serve as a substrate for subsequent rounds of RMCE.

To broaden the applicability of Cre recombinase, directed evolution strategies have been used to change the substrate specificity of Cre.

2.10. Directed molecular evolution.

Directed molecular evolution is the laboratory approach to Darwinian evolution and is a powerful tool for exploring and altering enzyme functions. The process is an algorithm of mutation, recombination and selection and has been widely applied industrially to redesign a myriad of protein properties including thermal and solvent stability, enzyme selectivity, solubility and protease susceptibility. Another approach to enzyme engineering is to make specific modifications which requires a detailed and often unattainable understanding of sequence and function and this can be bypassed by directed evolution.

One of the most important aspects of a directed evolution experiment is the construction of the diverse molecular libraries of enzyme variants. The methods of choice for creation of molecular diversity are error-prone PCR, oligonucleotide-directed randomization and *in vitro* recombination.

Random mutatgenesis of the whole gene is the most straightforward approach for library construction and requires no structural or mechanistic information while uncovering potentially beneficial mutations. Error-prone PCR followed by sequential rounds of screening and selection of improved variants has been used to realize wide number of engineering goals, like creation of enantioselective enzymes like cyclohexane monooxygenases (Reetz et al. 2004). Error-prone PCR was also used to enhance the selectivity of cytosine deaminase designed for killing tumor cells (Mahan et al. 2004), for increased activity of cytochromes 450 which are important for the metabolism of drugs and other xenobiotics (Kumar et al. 2005).

The most dramatic results in the field of directed evolution were achieved by using *in vitro* recombination methods to enrich diversity by exploring much larger sequence space. **DNA shuffling** pioneered by Stemmer (Stemmer 1994) is a method for *in vitro* homologous recombination of pools of selected mutant genes by random fragmentation and PCR reassembly (Fig.5).

Staggered extension process or StEP is another shuffling format which uses PCR based recombination with very short annealing or extension cycles (Zhao et al. 1998). Full length chimeric genes are created when fragments denatured from parent templates switch to a new template based on high homology. Due to template switching the growing polynucleotides contain sequence information from different parental genes and the resulting progeny has multiple crossovers from parent templates.

Recently several modifications of DNA shuffling have been reported. Exon shuffling (Kolkman and Stemmer 2001), family DNA shuffling (Crameri et al. 1998), sequence-homology-independent protein recombination or SHIPREC (Sieber, Martinez, and Arnold 2001). Other homology dependent recombination methods have been developed like degenerate homoduplex recombination for engineering a variety of mammalian epidermal growth factors (Coco et al. 2002), synthetic shuffling (Ness et al. 2002) and assembly of designed oligonucleotides (Zha, Eipper, and Reetz 2003). In these methods overlapping oligonucleotides encoding all the degeneracy in two or more parental genes are used in a PCR reassembly reaction. The result is a library where all the shuffled variants are equally likely regardless of linkage.

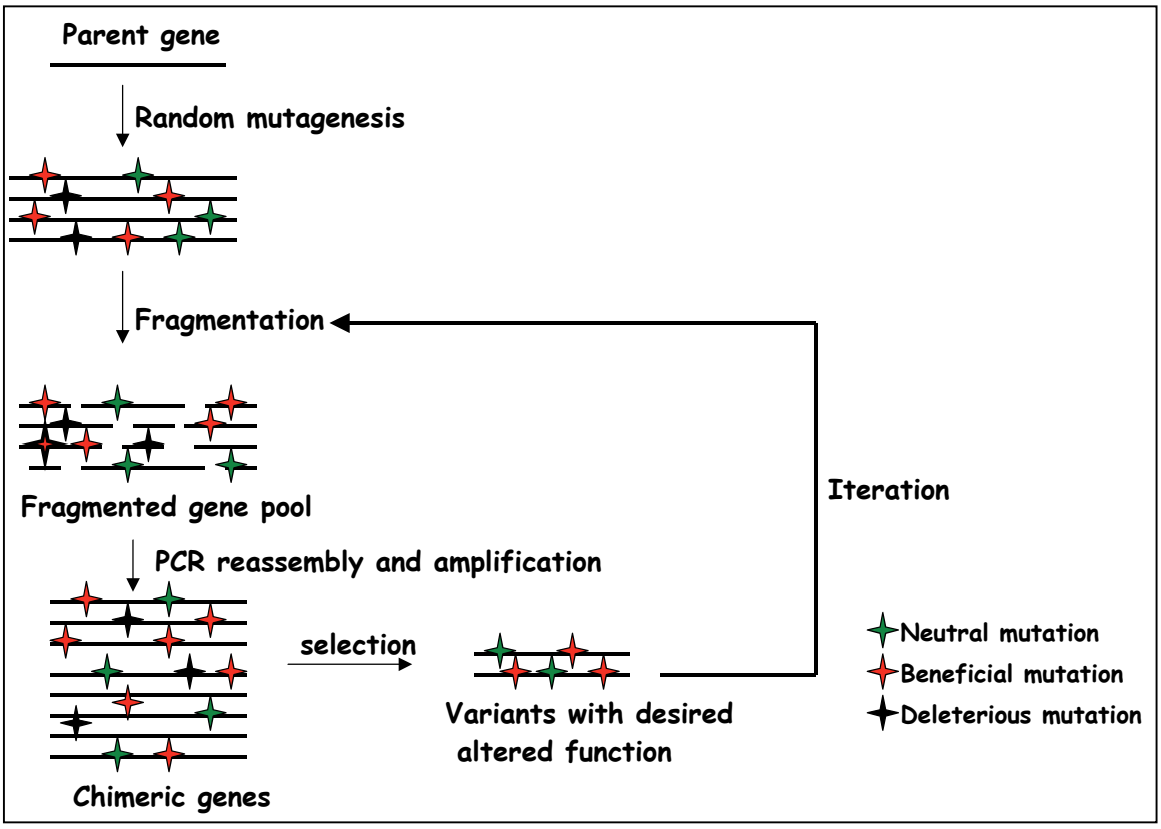


Fig.5. Schematic representation of DNA shuffling strategy.

Homology independent structure guided methods have also been developed. Sequence-independent site-directed chimeragenesis (Hiraga and Arnold 2003) is a method where semi-rationally designed peptide building blocks are shuffled. The method relies on a structure based computational predictor SCHEMA which defines peptide fragments from parental proteins which can be interchanged with minimum structural interference. Recently SCHEMA guided recombination of three cytochromes P450 was used to create chimeras of properly folded novel P450 proteins (Otey et al. 2006).

Apart from library diversity, a lot of focus is also been given to in vitro selection of novel enzyme variants. *In vitro* compartmentalization (IVC) using a water-in-oil emulsion has emerged as a promising alternative for selecting catalysts (Tawfik and Griffiths 1998; Doi and Yanagawa 1999; Griffiths and Tawfik 2003).

Therapeutic proteins like IFN-αs (Chang et al. 1999b), high affinity antibodies (Boder, Midelfort, and Wittrup 2000; Hanes et al. 2000), protease inhibitors like plasminogen activator inhibitor-1 (Stoop et al. 2001) have also been engineered using directed evolution and shuffling techniques. Besides engineering of gene therapy vehicles has also been achieved (Soong et al. 2000) through DNA family shuffling of envelope genes of MLVs for developing new tropism for CHO cells.

2.11. Directed evolution of Cre variants.

Applications of Cre are currently restricted by the requirement of appropriately positioned loxP sites in the targeted DNA regions. The ability to target for new target sites different from a canonical loxP would significantly increase the use of Cre. A number of groups have developed protein evolution strategies to manipulate the enzymatic properties of Cre to create new recombinase variants (Buchholz and Stewart 2001; Rufer and Sauer 2002; Santoro and Schultz 2002). Such studies were aimed at changing the target specificity of Cre to broaden the applicability of the recombinase to novel targets and also to get a better understanding of the mechanism of site-specific recombination and represent a step towards the custom evolution of recombinases that recognize unique sites within the mammalian genome.

A Flp recombinase variant with enhanced thermostability, Flpe, was generated from a randomly mutagenised library of Flp coding sequence using successive rounds of screening and DNA shuffling (Buchholz, Angrand, and Stewart 1998). Flpe exhibits about four fold higher recombinase activity in cell culture compared to wild type and is also used to mediate efficient recombination in mice (Farley et al. 2000; Rodriguez et al. 2000; Awatramani et al. 2003). Flp variants with an altered target site (mFRT) specificity has also been generated using a dual reporter screen in E.coli (Voziyanov, Stewart, and Jayaram 2002). Besides Flp and Cre, another site specific recombinase, the phage ϕC31 integrase was subjected to directed evolution using two cycles of shuffling and screening in E.coli to obtain evolved integrases with better sequence specificity and integration frequency at a pseudo *attP* sequence on human chromosome 8 (Sclimenti, Thyagarajan, and Calos 2001).

Such directed evolution strategies can be further used to expand the recombination repertoire of Cre and related recombinases for important practical applications in genome manipulations.

2.12. Aim of my work.

Site-specific recombinases like Cre have proved to be highly efficient tools for genomic manipulations. The Cre recombinase favours excision over integration as the product of integration becomes a substrate for re-excision if recombinase activity is still around. In fact most of the genomic manipulation strategies developed exploit the virtually irreversible excision recombination mediated by Cre. The efficiency of Cre mediated site-specific recombination has been used in a wide variety of exogenous systems like mammalian cell culture, mice, flies and zebrafish.

Retroviruses like HIV-1 integrate into the host chromosome and persist as a provirus flanked by viral long terminal repeats or LTRs. Establishment of the provirus is an obligatory step in the retroviral life-cycle and serves to maintain viral sequences in the infected cell. The LTRs of proviral DNA direct the expression of viral genes. A molecular strategy to excise the proviral DNA from infected cells could be envisioned by applying the principle of site-specific recombination mediated through sequences in the flanking LTR repeats of the pro-viral genome. The Cre/lox system has been indicated as a possible antiviral strategy against retroviruses (Flowers et al. 1997; Lee and Park 1998). A modified Cre recombinase recognising sequences within each LTR of proviral DNA could mediate excision of the viral coding sequences eliminating the intact provirus and represents a novel direct antiviral strategy targeting the provirus. To realize the potential of a genetic surgery targeting an integrated provirus, it is essential to determine whether such a recombinase can actually be generated that can recombine a sequence within a retroviral LTR. The laboratory evolved Cre recombinase variants with altered site-specificity have proved that it is possible to generate tailored recombinases and hold promise for increasing the applicability of these recombinases. Directed evolution has developed into a mature technology and can be used to engineer solutions for the

development of human therapeutics and generate new protein variants that exhibit traits never observed before. To maximize the utility of evolved recombinases it is essential to demonstrate the feasibility of directing evolution towards more variant and asymmetric target sites.

The aim of my work has been to use directed evolution to generate a custom designed novel Cre recombinase recognizing an asymmetric sequence within an HIV-l LTR, investigate the recombination efficiency of this novel recombinase in bacteria and mammalian cells and provide proof of principle that the applicability of such evolved enzymes could be further extended for therapeutic anti-retroviral DNA surgery.

3. Abbreviations.

Amp Ampicillin

Ara L-Arabinose

CAT Chloramphenicol acetyl transferase

Cm Chloramphenicol

Cre Bacteriophage P1 encoded site-specific recombinase

DIG Digoxigenin

DMEM Dulbecco's modified Eagles medium

dNTP Deoxyribonucleotide triphosphates

dUTP Deoxyuridine triphosphate

FCS Fetal calf serum

IRES Internal ribosome entry site

Kb Kilobase pairs of DNA

Loxltr HIV-1 LTR derived target sequence

LoxP Cre recombinase recognition target sequence

LTR Long terminal repeat

Luc Luciferase

μg/ml microgram/mililitres

μl microlitres

OD Optical density

PAC Puromycin acetyl transferase

PBS Phosphate buffered saline

Pen-strep Penicillin-streptomycin for use in mammalian cell culture medium

SDS Sodium dodecyl sulphate

X-gal Histochemical reagent for detection of Beta galactosidase activity, (5-Bromo-4-Chloro-3-Indolyl-β-D-Galactopyranoside).

4. Materials and Methods

4.1. Enzymes, chemicals, reagents and kits.

All restriction enzymes, T4 DNA ligase ligase and Calf intestinal alkaline phosphatase were purchased from New England Biolabs (NEB), USA.

Taq DNA polymerase, Deoxyribonucleotides (dNTPs), 10X PCR buffer and $MgCl_2$ was purchased from Bioline.

Agarose was purchased from Invitrogen Life Technologies.

All cell culture reagents, that is DMEM, FCS, Pen-Strep and Trypsin were purchased from Gibco. Puromycin was purchased from Invivogen and Geniticin was purchased from Gibco.

Qiagen's Effectene transfection reagent was used for all transfections.

DNA isolation was performed with Qiagen's QIAprep Spin Mini-prep Kit and Qiagen Plasmid Maxi Kit.

Genomic DNA from Hela cells was isolated with QIAamp DNA Blood minikit.

Gel extraction and PCR purification was performed with QIAEX11 gel extraction Kit (Qiagen) and QIAquick PCR purification kit (Qiagen).

X-Gal was purchased from DiagnosticsChemicals Limited, Charlottetown, PEI.

L-arabinose was purchases from Sigma-Aldrich.

LB liquid bacterial medium and LB-agar plates were provided by the institute kitchen.

4.2. Synthetic oligonucleotides

All the oligonucleotides in this work were ordered from Biosprings, Germany.

4.3. Bacterial strain

For all purposes the strain E.coli DH5α (F'phi80d*lac*Z delta(*lac*ZYA-*arg*F)U169 *deo*R *rec*A1 *end*A1 *hsd*R17 (rk-, m k+) *pho*A *sup*E44lambda-*thi*-1 *gyr*A96 *rel*A1/F' *pro*AB+ *lac*IqZdeltaM15 Tn10(*tet*r) was used.

4.4. Plasmids

1). **Evolution vector pEVO**.

This plasmid is based on the pBAD33 vector (Guzman et al. 1995) and contain two directly repeated recombinase target sites (loxP or loxltr and loxltr subsets). The plasmid allow arabinose induclible expression of the recombinases which can be efficiently repressed with glucose in E.coli. The recombinases are expressed by the araC promoter. The plasmid offers convenient cloning of the recombinases using *BsrG*1 and *Xba*1 sites flanking the recombinase coding region, and also cloning of directly repeated lox sites by using flanking Bgl11 restriction sites. The map of the vector is shown in Fig.6a.

2). **PSVpaX and PSVloxltr reporter vectors.**

The pSVpaX and pSVloxltr are multihost reporter vectors. They are recombinase excision reporter plasmids for use in both mammalian cells and *E.coli*. Monitoring of lacZ activity allows easy detection of recombination. In bacteria, recombination at the target lox sites results in removal of a promoter within the puromycin acetyl transferase (pac) gene that results in prevention of lacZ expression and renders white colonies on X-Gal plates. In mammalian cells, recombination turns on lacZ expression by removal of the 1.1 kb pac region.
The pSVpaX and pSVloxltr are identical except for the directly repeated loxP or loxltr sites flanking the pac region, respectively. The map is shown in Fig. 6b.

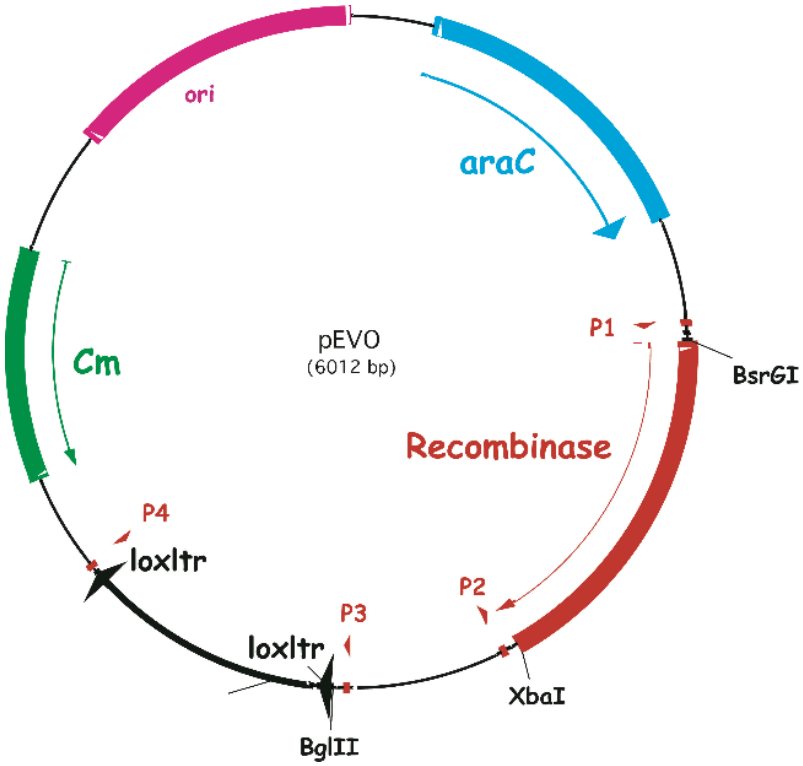
3). **Mammalian recombinase expression vectors.**

This group of expression plasmids were available commercially from Clontech. The expression cassette contains the human cytomegalovirus major immediate early promoter/enhancer, a synthetic intron known to enhance the stability of the transcript, an ECMV IRES followed by an antibiotic resistance marker. The IRES facilitated the expression of the selection marker (Fig.6c). The unoptimised Tre recombinase was cloned into the hygromycin version of the vector by using *BsrG*1/*Nhe*1 sites in the MCS, and the optimized hTre was cloned into neomycin version of the vector by using *Nhe*1 and *Bam*H1 sites in the MCS of the vector (Fig.6d).

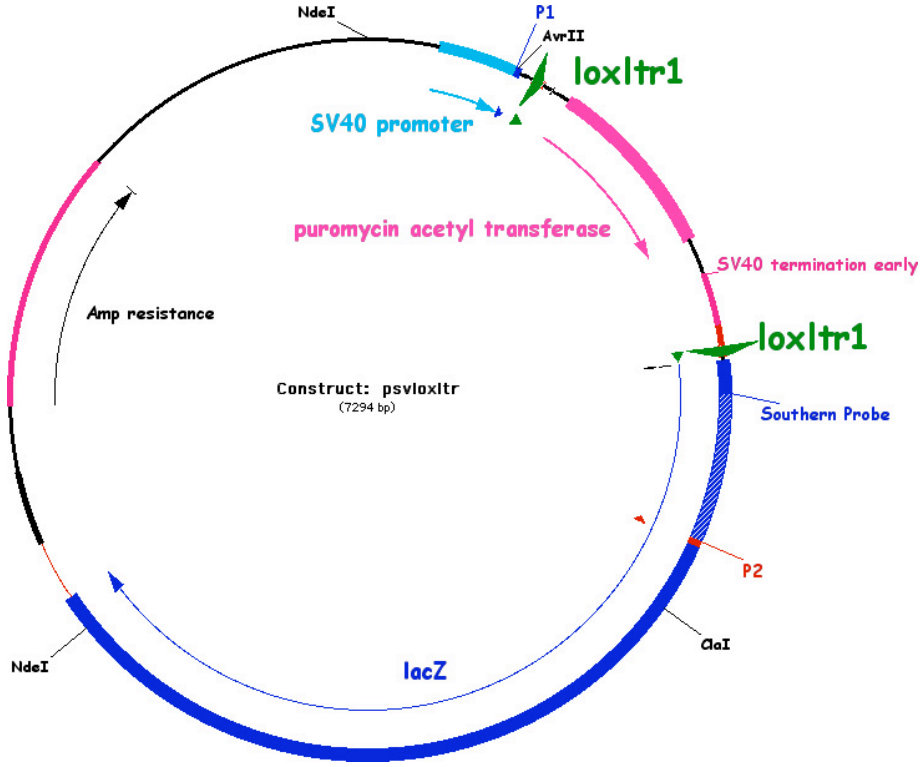
4). **HIV-1 HXB3 based reporter vectors (from J. Hauber lab, HPI, Hamburg).**

These vectors contain the HIV-1 HXB3 long terminal repeat, which drives the expression of a downstream reporter gene. The 5' long terminal repeat has the TAR binding site where Tat protein binds and transactivates reporter genes. The loxltr sequence was cloned within the long terminal repeats of the vector. The HIV_T2_LUC and HIV_T1_LUC contain the firefly luciferase as the reporter gene, and the HIV_T2_CAT has the chloramphenicol acetyl transferase as the reporter gene (maps shown in Fig.26 and Fig.28). The HIV_T2_CAT stable Hela cell lines were also provided by the Hauber Lab.

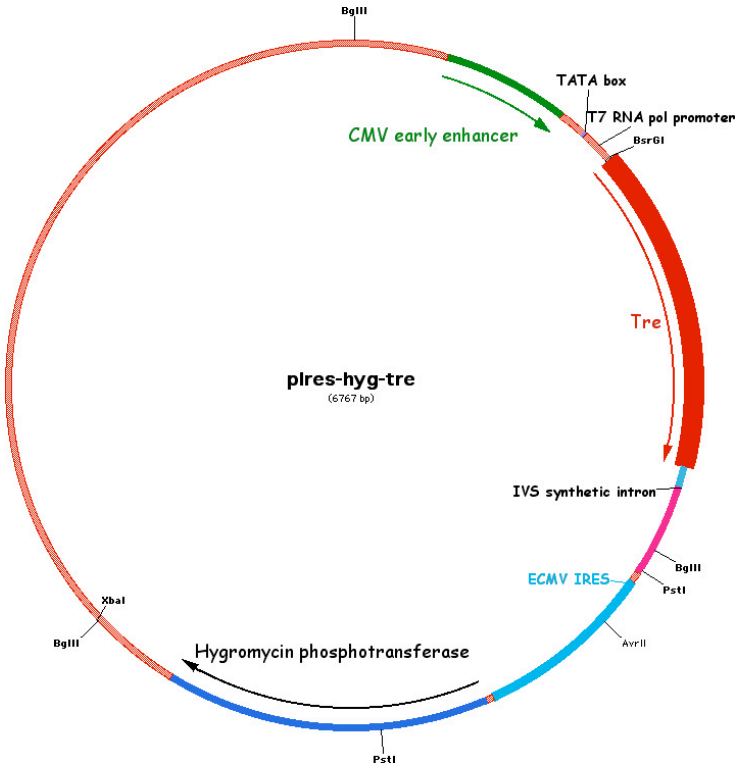
(a).


ori
araC
Cm
pEVO
(6012 bp)
P1
BsrGI
Recombinase
P4
loxItr
P2
P3
loxItr
XbaI
BglII

(b).


NdeI
P1
AvrII
loxItr1
SV40 promoter
puromycin acetyl transferase
SV40 termination early
Amp resistance
loxItr1
Construct: psvloxItr
(7294 bp)
Southern Probe
P2
ClaI
NdeI
lacZ

(c).


BglII
TATA box
T7 RNA pol promoter
BsrGI
CMV early enhancer
Tre
pIres-hyg-tre
(6767 bp)
IVS synthetic intron
BglII
PstI
ECMV IRES
AvrII
XbaI
BglII
Hygromycin phosphotransferase
PstI

(d).

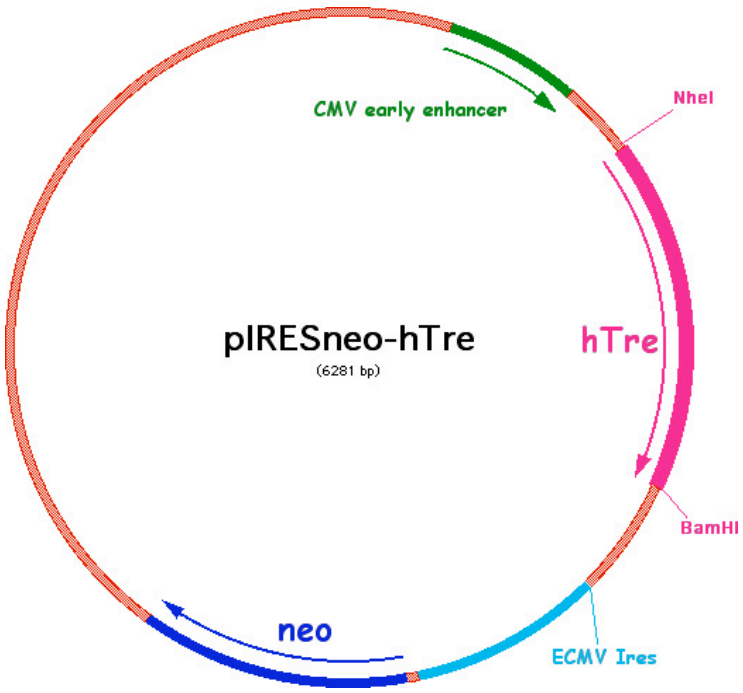

CMV early enhancer
NheI
pIRESneo-hTre
(6281 bp)
hTre
BamHI
neo
ECMV Ires

Fig. 6 Vectors used in this project.

(a). Evolution vector pEVO.

(b) Reporter vector used for recombination assay in Hela cells. The psvpaX vector is the same except for presence of loxP sites in place of loxltr. P1 and P2 represent the primer pairs used to assay for recombination in Hela cells.

(c) Recombinase expression vectors used in Hela cells.

4.5. Construction of pEVO-loxltr vectors and psvloxltr vectors.

A 690 base pairs fragment flanked by loxltr sites was constructed by using oligonucleotides:

(5'ATGAGATCT**ACAACATCCTATTACACCCTATATGCCAACATGG**AAGCTTGCATGCCTGCAGATCGAG-3')

(5'TTGAGATCT**CCATGTTGGCATATAGGGTGTAATAGGATGTTGT**TCGAACTGTACCGGTTGTTAGTGA-3') using the pEVO-loxP vector as a template.

The design of the oligos were designed that the resulting fragment would be flanked also by *Bgl*11 sites. The PCR fragment was digested with *Bgl*11 and CIP and cloned into similarly digested pEVO vector. Clones were sequenced using the primer (5'-CAATAACCCTGATAAATG-3'). All the loxltr subset vectors were also constructed in a similar fashion.

The psvloxltr reporter vector was constructed in a two-step method using psvpaH as a template. The first loxltr site was cloned in place of the loxH site by using oligos (5'-ACAACCGGTACAGTTCGAACAACATCCTATTACACCCTATATGCCAACATGGAAGCTTGCATGCCTGCAGGTCGGCCGCCACGACCGG-3') and

(5'-TTCGACGCTCTCCGGCGT-3'). *BstB*1/*Sac*11 sites were used for cloning and correct clone verified by sequencing. The second loxltr site was cloned by using primers (5'-GTCTGGATCC**ACAACATCCTATTACACCCTATATGCCAACATGG**ACCGGGCCACCATGGTCGCGAGT-3') and (5'-TCATCGATAATTTCACC-3') and using *BamH*1 and *Cla*1 sites for cloning.

4.6. Preparation of recombined versions of psvloxltr and HIV_T2_CAT vectors

The pEVO-loxltr-Tre25 recombinase was digested with *Sac*1 and *Xba*1 and was cloned into the similarly digested pBAD33-Cre vector. The resulting pBAD-Tre25 (Cm resistant) vector was used to prepare chemically competent cells with recombinase expression induced with 100μg/ml L-arabinose. These cells are transformed with the reporter vectors (psvloxltr or HIV_T2_CAT, both Amp resistant) and recombinase induced with the same arabinose concentration and grown for 2 hours at 37°C. Then a part of the cells are plated onto Ampicillin (100μg/ml) containing LB-agar plates and the rest of the cells are cultured overnight in LB containing Chloramphenicol (25μg/ml) and Ampicillin (100μg/ml) and 100μg/ml of L-arabinose. Plasmid DNA was isolated from the overnight culture and transformed into *E.coli* DH5α competent cells. The culture was plated onto ampicillin LB agar plates. Random colonies were picked up and restreaked on Cm Plated to segregate the recombinase plasmid from the recombined reporter plasmids.

4.7. DNA shuffling and cycling.

DNA shuffling was carried out to recombine the obtained Cre mutants in vitro. The basic protocol was according to Stemmer (Stemmer 1994). DNA fragments containg the Cre gene or mutant libraries were randomly fragmented by sonication using 12% amplitude for 5 minutes with 30s pulse on and 5s pulse off. After every third cycle the libraries were shuffled, for instance the third library was shuffled by mixing 3 parts of library 3, 2 parts of library 2 and 1 part of library 1. The sonicated fragments (100-500 bp in size) were purified by the PCR purification kit from Qiagen. To reassemble the fragments 10μl of the sonicated DNA was added to a 50μl PCR mix containing 1X PCR buffer, 250μM of each dNTP, 3mM of $MgCl_2$ and 1.5units of BioTaq DNA polymerase and amplified

according to the following program: 35 cycles of 94°C, 60s; 55°C, 60s; 72°C, 90s. Three microlitres of this primerless PCR reaction product was used as a template for the next PCR with primers P1 (5'-TCTACTGTTTCTCCATA-3') and P2 (5'- GCGGATGAGAGAAGATT-3') in a 50µl PCR mix with the same cycling conditions. The shuffled library was digested with *BsrG*1 and *Xba*1 and recloned into a similarly digested pEVO vector used for directed evolution.

Plasmid DNA isolated after each evolution cycle was digested with Nde1 and amplified with the same PCR conditions as stated above using primers P1 and P4 (5'-TGTCGCCCTTATTCCCT-3'), digested with *BsrG*1 and *Xba*1 and subcloned into similarly digested pEVO. The resulting recombinase libraries were transformed into *E.coli* DH5α chemically competent cells and grown in LB liquid culture containing chloramphenicol at a concentration of 25µg/ml and L-arabinose to induce recombinase expression.

4.8. Cell culture

Hela cells were cultured at 37°C under 5% CO_2 in DMEM containing 100 units /ml of penicillin and streptomycin and supplemented with 10% FCS.

Puromycin resistant psvpaX and psvloxltr reporter cell lined were obtained by transfection of $2X10^6$ cells in a 10cm dish with 8µg of the reporter plasmids using Effectene transfection reagent (Qiagen) according to the manufacturer's instructions, and selection with the antibiotic was started 48 hours post transfection. The final stable clones were grown in presence of 3µg/ml of puromycin (Invivogen).

4.9. Transient assays for measurement of β-galactosidase activity

Hela cells were transfected with a 1:2 ratio recombinase expression plasmids and lacZ reporter plasmids in 6-well format by using Effectene transfection mix. A total of 0.6µg of DNA was used for transfections. Parallel transfections with recombined version of the reporter plasmids were carried out to express the final enzyme activity as a percentage of total activity in recombined reporter plasmids. Relative β–galactosidase activities were measured 48 hours post transfection using the Tropix Galacto-Light kit (Applied Biosystems, Bedford, MA) according to manufacturer's protocol. Recombination

efficiencies were corrected for total protein content in the cell lysate measured by BCA protein assay kit (Pierce, Rockford, IL).

The loxltr stable cell lines were transfected with 0.8μg with either mock, Cre or Neo-hTre plasmids using Effectene and were assayed for β–galactosidase activities 48 hours post transfection using Galacto-Light. The β–galactosidase activities were corrected for transfection efficiencies by measurement of renilla luciferase activity using assay kit from Biotium, Hayward, CA.

4.10. X-Gal staining of Hela cells

Hela cells were analysed for recombination and expression of β–galactosidase by X-Gal staining. The cells were washed with PBS 48 hours post-transfection, and fixed with 2% formaldehyde and 0.1% glutaraldehyde, permeabilized with 0.5%TritonX100 and stained overnight at 37°C with staining solution containing 1mg/ml X-gal, 2mM $MgCl_2$, 5mM potassium ferricyanide and 5mM potassium ferrocyanide.

4.11. PCR analysis of recombination in stable reporter Hela cell lines.

Stable loxltr and loxP Hela cell lines were transfected in 10 cm format with respective recombinase expression vectors using Effectene. Transfected cells were washed with PBS after 48 hours, and trypsinised. Hela genomic DNA was isolated using QIAamp DNA Blood mini kit (Qiagen) according to the manufacturer's instructions. Genomic DNA (3μg) was analysed for recombination by using primers P1 (5'-GCCTCGGCCTAGGAACAGT-3') and P2 (5'- CCGCCACATATCCTGATCTT-3') using the PCR program 30 cycles of 95°C, 30s; 62°C, 30s; 72°C, 40s in a 50μl PCR reaction containing 1X PCR buffer, 250μM of each dNTP, 3mM of $MgCl_2$ and 1.5units of BioTaq DNA polymerase. The PCR products were visualized on a 0.7% agarose gel.

4.12. Southern analysis

10μg of Hela genomic DNA was electrophoresed in a 0.7% agarose gel. The DNA containing gel was depurinated using 250mM HCL for 20minutes, followed by denaturation in 0.5MnaOH, 1.5M NaCl for 15 minutes, twice, and finally neutralized with 0.5M Tris-Hcl (pH7.5), 1.5M NaCl for 15 minutes twice. The gel was then shortly equilibrated in 20XSSC, before overnight blotting onto a nylon membrane by capillary transfer in 20XSSC. After transfer, the membrane was UV crosslinked. After crosslinking the membrane was washed in water and air-dried followed by 2 hour long prehybridisation in buffer containing DIG Easy Hybridization granules (Roche Diagnostics, Germany). A DIG-dUTP labeled DNA probe hybridizing to the lacZ region in the psvpaX or psvloxltr reporter plasmids was synthesized using PCR DIG Probe Synthesis Kit (Roche Diagnostics, Germany) according to the manufactures's instructions. After prehybridization, the membrane was hybridized overnight at 50°C in prewarmed hybridization buffer prepared from DIG Easy Hyb Granules and containing 2μl probe /ml hybridization buffer. Hybridization was also carried out by using RNA probe synthesized using DIG RNA labeling mix (Roche Diagnostics, Germany). After hybridization, the membrane was washed in low stringency buffer (2XSSC, 0.1%SDS) three times for 5 minutes each at room temperature, followed by washing in high stringency buffer (0.5XSSC, 0.1%SDS for DNA probe and 0.1XSSC, 0.1%SDS, for RNA probe), at 68°C twice for 15 minutes each.

After stringency washing, the blot was subjected to immunological detection using an anti-digoxigenin antibody conjugated to alkaline phosphatase (Roche Diagnostics, Germany) and chemiluminescent substrate CDP-Star (Roche Diagnostics, Germany).

4.13. Transient Tat transactivation assays using luciferase reporter vector and CAT reporter vector

Hela cells were transfected in 12 well format using Effectene transfection reagent for Tat activation assay with luciferase reporter vectors. The transfection mix contained 0.16μg HIV_T2_LUC or HIV_T1_LUC expressing firefly luciferase, 0.16μg of pIres-neo-hTre recombinase vector or the empty vector as negative control, 0,08μg of the Tat vector or pcDNA3 as negative control for Tat. The cells were lysed 48 hours post transfection and assayed for firefly luciferase activity by using Firefly and Renilla luciferase assay kit (Biotium, Inc., Hayward, CA). The activities were corrected for transfection efficiency by using psvΔpax vector as an internal control and measuring the β-galactosidase activity from the cell lysates.

The Tat activation assay with the HIV_T2_CAT vector was carried out in a similar manner. Transfections were carried out in 6-well format by using 0.2μg of the CAT vector, 0.4μg of the pIres-neo-hTre vector or the empty vector as negative control and 0.2μg Tat vector or pcDNA3 as negative control for Tat. The transfected cells were lysed 48 hours post transfection and the CAT activity was measured using the CAT ELISA kit (Roche Diagnostics, Germany) according to manufacturer's instructions. PCR analysis of recombination was carried out in Hela cells transfected with the HIV reporter vectors in presence or absence of the recombinase, using primers L1 (5'-GAAGGTGGGTTTTCCAGTCA-3') and L2 (5'- AGGGAAGAAAGCGAAAGGAG-3') using the PCR program program 30 cycles of 95°C, 30s; 60°C, 30s; 72°C, 40s in a 25μl PCR reaction containing 1X PCR buffer, 250μM of each dNTP, 3mM of $MgCl_2$ and 1.5units of BioTaq DNA polymerase. The PCR products were visualized on a 0.7% agarose gel. The PCR products were gel extracted and purified using Gel extraction kit (Qiagen) and sent for sequencing.

5. Results

The sequence specific manipulation of DNA by Cre recombinase represents a powerful approach to studying gene function. The relatively simple biochemical requirements of Cre, the very high sequence specificity towards loxP sites and the fact that Cre can mediate recombination both in prokaryotes and eukaryotes have contributed to the development of elegant strategies for genomic manipulations and Cre has been successfully used for conditional mutagenesis, gene replacement and deletion and chromosome engineering experiments (Nagy 2000; Sauer 2002). Applications of Cre however require appropriately positioned loxP sites in the targeted DNA regions. The ability to target for new target sites different from a canonical loxP would significantly increase the use of Cre. In the recent years several groups have applied directed evolution to generate Cre variants recognizing novel target sites (Buchholz and Stewart 2001; Rufer and Sauer 2002; Santoro and Schultz 2002). Such studies represent a step towards the custom evolution of recombinases recognising unique sites within mammalian genomes and serve as a useful starting point for the directed development of recombinases recognizing distinctly non-lox targets. To maximize the utility of evolved recombinases it is essential to demonstrate the feasibility of directing evolution towards more variant and asymmetric target sites and requires detailed study of the applied properties of such recombinases and their applications for therapeutic purposes.

5.1. Selection of a sequence from an HIV-1 LTR as a novel target site for directed evolution.

The first requirement was to select an HIV-1 LTR sequence with optimum sequence similarity to loxP, to facilitate the evolution process. A prerequisite for any directed evolution experiment is the detection of some amount of desired activity in the starting library, as the mutations in a library made by whole gene random mutagenesis has limited

diversity which is unlikely to increase the activity more than a factor of two (Moore and Arnold 1996).

We looked for sequences within the long terminal repeats (LTRs) of HIV-1 genomes. Several HIV-1 LTR sequences were scanned for similarities to the canonical loxP site. Among all the LTR sequences scanned, the sequence having the maximum sequence identity to loxP was selected as the novel target site to direct evolution of Cre and was termed loxltr. The sequence belonged to the 695 bp 5' LTR DNA of HIV-1 strain TZB0003 (locus AF096641) from Tanzania and is part of the modulatory U3 region of the LTR. The U3 region of an HIV-1 LTR is normally around 450 base pairs and contains mostly of cis-acting DNA elements, which are binding sites for cellular transcription factors. The selected *loxltr* site is a 34 bp asymmetric sequence and has 50% sequence similarity to *loxP* with four bases mismatch in the left element and six in the right element and a completely different spacer (Fig.7).

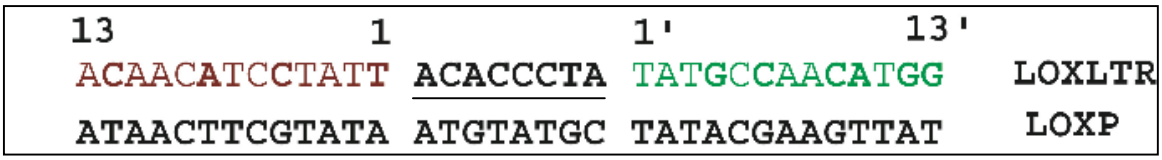


Fig.7The selected *loxltr* sequence. The bases in bold represent those different from loxP, the red sequence represents the left element or left half-site and the green the right element or right half site. The 8bp spacer sequence is underlined for *loxltr*.

5.2. Directed evolution strategy to generate loxltr specific recombinase.

Once the target site was selected, it was necessary to direct the evolution of Cre recombinase using an evolution strategy to create a variant recognizing the *loxltr* sequence.

A PCR based evolution strategy, **substrate linked protein evolution** (**SLiDE**) had been previously used successfully to generate Cre variants, which recognise and recombine a sequence chosen from the human chromosome 22, termed *loxH* (Buchholz and Stewart 2001). SliDE incorporates an important feature of directed RNA evolution in that there is physical linkage between the coding molecule undergoing evolution and the target.

The evolution vector used in SliDE is termed pEVO and is a single vector in which both the recombinase coding region and its target lox sites are present. The recombinase expression is regulated by an arabinose inducible promoter (the details of this vector are given in the Materials and methods section and Fig.6a). The vector is so designed that the downstream target sites flank an intervening region of around 700 bp which contain an *Nde*1 cleavage site.

The presence of the recombinase coding region and its substrate (target sites) on the same vector ensures that upon expression of the recombinase, its action would alter the substrate only upon successful recombination and lead to a deletion of the intervening region between the target site pair resulting in the loss of the *Nde*1 site, and a PCR using primer pair P1 and P4 (as shown in Fig.8) would selectively amplify only those recombinases which produce the desired change in the substrate from a whole library of recombinase variants. In this way large libraries can be conveniently screened and successful recombinase coding regions can be rapidly retrieved. The evolution cycle begins with cloning the recombinase or their mutated libraries into pEVO along with the required target sites. This vector is then transformed into *E.coli*, and recombinase expression induced with arabinose. After overnight culture the plasmids are harvested from the pool, digested with *Nde*1 and then PCR amplified with the primers P1 and P4 specified in Fig.8. The PCR product of the recombined form of the plasmid produces a 1.7 Kb band. The strategically placed restriction site ensures that plasmids not recombined are linearised by *Nde*1 and rendered unsuitable for PCR amplification thus removing the unsuccessful candidates from the pool. The PCR product is digested by *BsrG*1 and *Xba*1 and the successful recombinase coding sequences is subcloned back into the vector for the next cycle. Using *E.coli* for the evolution process reduces the time-scale of each cycle and also favours easy evaluation of the efficiency of evolution process. I decided to use this strategy for *in vitro* evolution of a *loxltr*- specific recombinase owing to the ease, reproducibility, technical simplicity and the availability of all required resources and vectors in my lab. Fig.8 gives a schematic description of the evolution strategy.

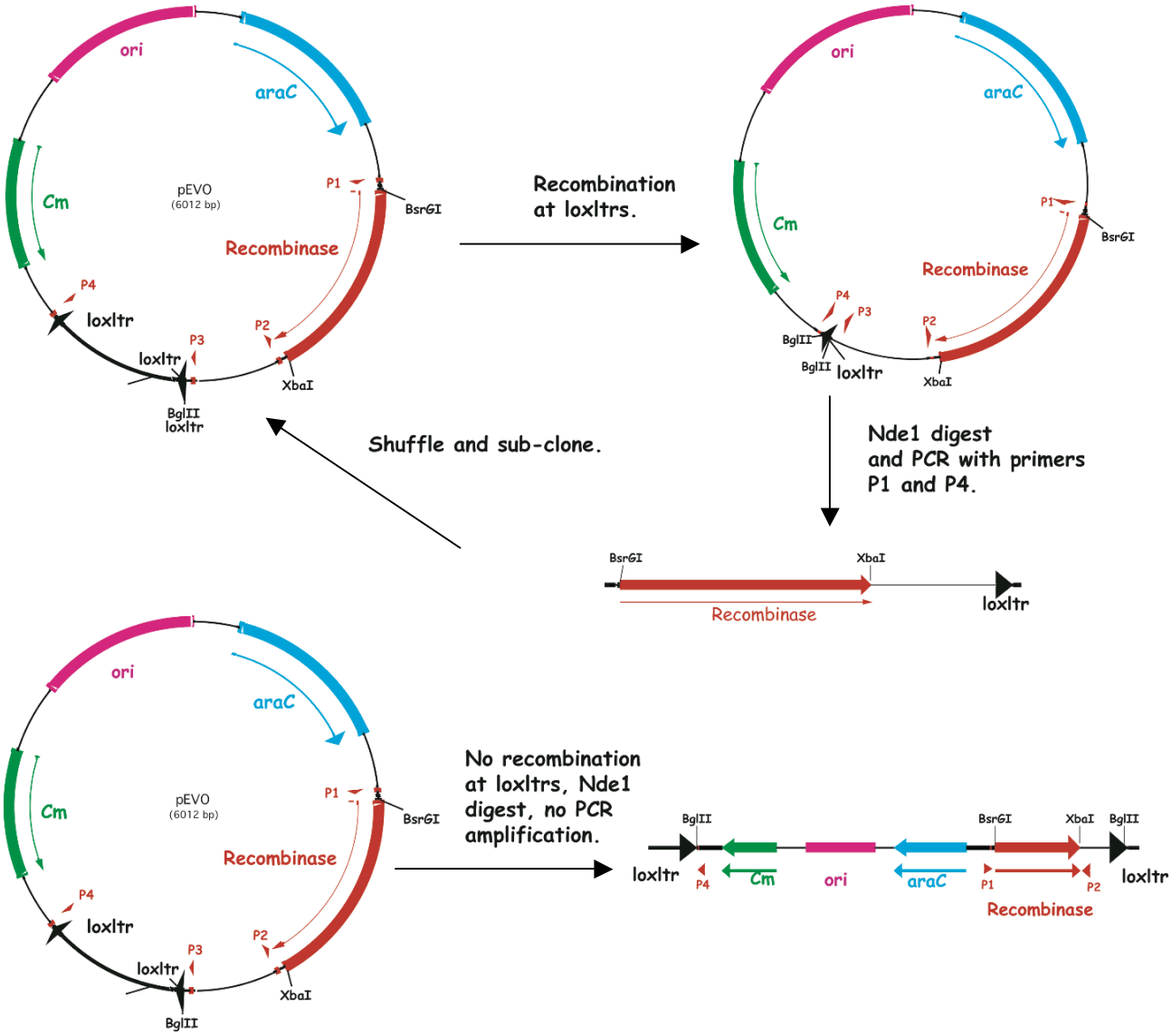


Fig.8. Schematics of Substrate linked directed evolution or SliDE.

5.3. Getting started with in vitro evolution.

The pEVO vector permits a qualitative and easy analysis of the extent of recombination by a *BsrG*1-*Xba*1 digest of the vector carrying the recombinase and its target sites. A 4.2 Kb band appears as a result of recombination which deletes the intervening DNA between the target lox sites. Intensity of this band reflects the extent of recombination in comparison to unrecombined upper band of 4.9 Kb. Using the pEVO-loxP-Cre vector as an example, recombination at loxP was evident even in absence of induction, an almost complete recombination occurred at an induction with 10μg/ml L-arabinose (fig.9).

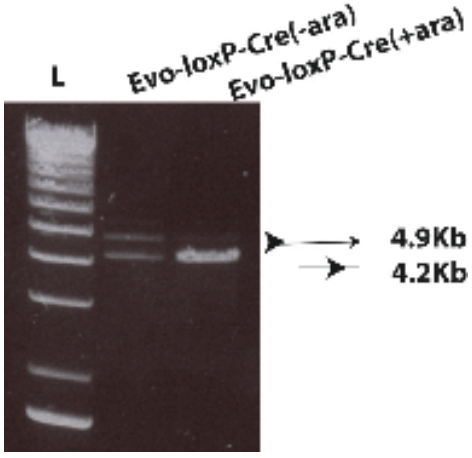


Fig.9. Cre mediated recombination at loxP in *E.coli*.. *BsrG*1 and *Xba*1 digest of pEVOloxP-Cre plasmids in presence (10 μg/ml) and absence of arabinose induction. The lower band of 4.2 Kb represents the recombined band.

The loxltr sequence was cloned into the pEVO vector in place of loxP by using oligos (described in materials and methods section) to generate the pEVO-loxltr vector. Firstly wt Cre was cloned into this vector to create pEVO-loxltr-Cre, transformed into *E.coli* DH5α, and recombinase expression induced with high concentration of arabinose (200μg/ml). Recombination at the new loxltr target sites would lead to a deletion of 690 bp of intervening DNA between the two sites. An *Nde*1 digestion would remove the un-

recombined vectors and subsequent PCR would produce a 1.7 kb band reflecting recombination. However Cre failed to recombine the loxltr sites and no PCR product was obtained (data not shown).

An archive of randomly mutagenised Cre libraries present in our lab was subsequently cloned into the pEVO-loxltr to produce pEVO-loxltr-lib and analysed for recombination in the same manner as for Cre. The starting mutant Cre library was unable to recognise the loxltr target and the expected PCR product was not obtained even at the high level of arabinose induction (data not shown). This result reflected that the asymmetry and the mutations in loxltr were not recognizable by the mutant Cre library to have any residual recombinational activity and it was not possible to begin directed evolution process.

5.4 Splitting up of loxltr into pallindromic subsets.

The rationale therefore was to split the original loxltr target into subsets which are pallindromic to facilitate the evolution process. Two additional pallindromic target sites, *loxltr1* and *loxltr2* were created based on the original aymmetric *loxltr* sequence, with the LE and RE sequences respectively used to form an inverted repeat sequence as is in *loxP* (fig.10).

The loxltrs 1 and 2 were cloned into the evolution vector to give pEVO-loxltr1and pEVO-loxltr2. Recombination at these new pallindromic targets were tested by subsequent cloning of Cre recombinase into the evolution vectors, pEVO-loxltr1 and pEVOloxltr2. As before recombinase expression was induced with 200μg/ml of L-arabinose but none of these two subsets were recombined by the Cre (data not shown). When the mutagenised Cre library was cloned into these vectors, and induced with arabinose, the same result was repeated and no recombination was observed (data not shown).

This led to the conclusion that the pallindromic nature of the sub-sets was not enough and that the mutations in these sites were still too many for the starting library to have any action, and this necessitated the further splitting of *loxltrs1* and *2* by evenly dividing the half-site mutations in the left element (LE) and right element (RE). The mutations of the

half-sites of loxltr1 and 2 were thus split equally to form four new subsets, loxltrs1a, 1b, 2a and 2b. The sequences of all the loxltr subsets are shown in Fig.10.

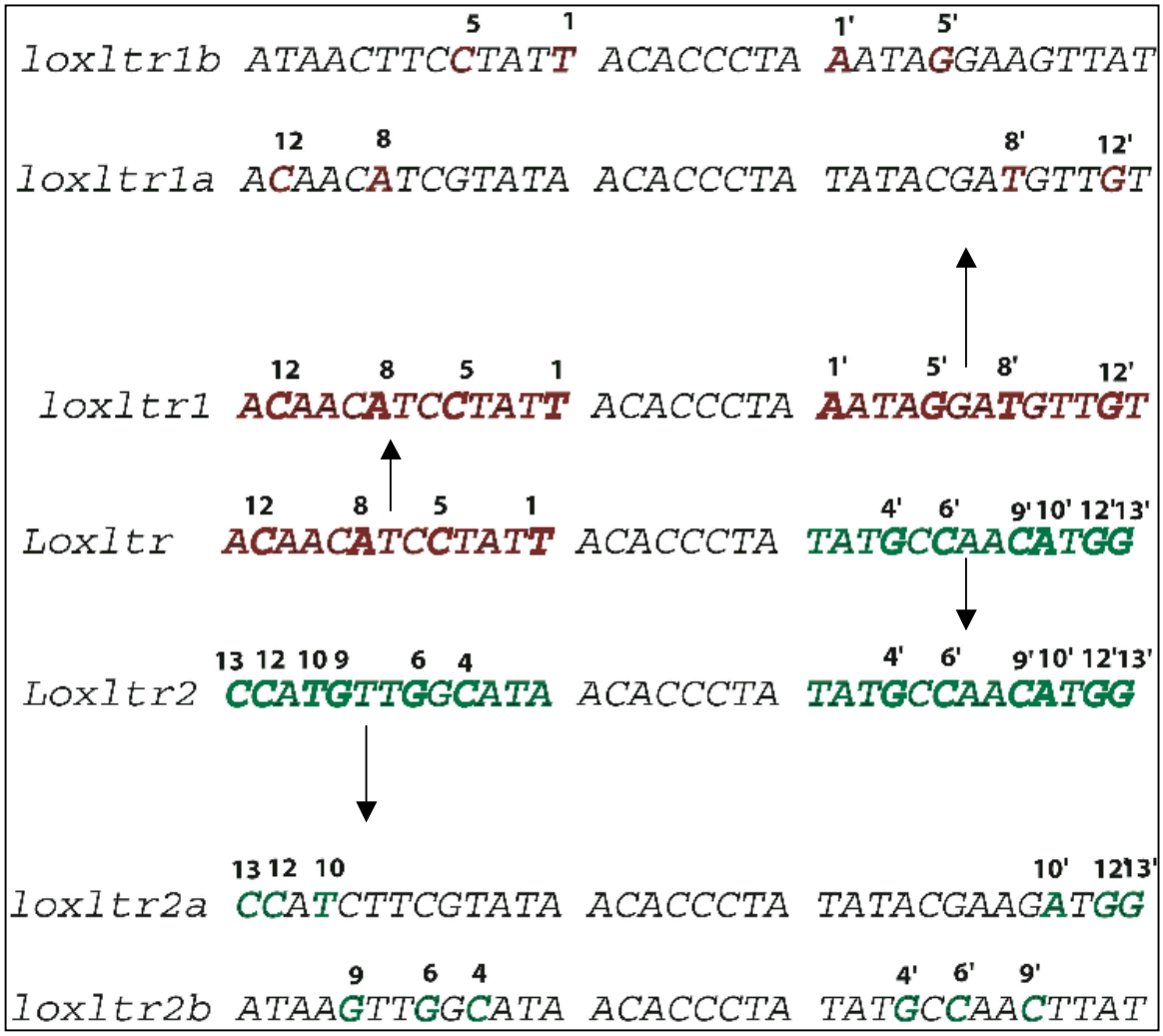


Fig.10. Splitting scheme of loxltr site. Left 13 bp and right 13 bp sequences of loxltr has been respectively used to form pallindromic 34 bp sub-sets loxltr1 and loxltr2. Both these target sites have the original spacer of loxltr. The loxltr1 half-sites are in red with the mutated residues with respect to loxP in bold, and the loxltr2 half-sites are similarly represented in green. The loxltrs1a and 1b are derived from loxltr1 and have two half-site mutations each shown in red, whereas loxltrs2a and 2b have three half-site mutations each shown in green and are split sub-sets of loxltr2. Arrows depict the manner in which the sub-sets originated from loxltr.

5.5 Directing evolution of recombinases for loxltrs 1a, 1b, 2a and 2b.

The four new loxltr sub-sets 1a, 1b, 2a and 2b were cloned in the pEVO vector. Cre recombined only loxltr1a at an arabinose induction of 200μg/ml, and the expected PCR product of 1.7Kb was produced, but the other three loxltr subsets were not recombined by wt Cre (data notshown).

Expression of the mutagenised library (lib1) with the same arabinose induction (200μg/ml) resulted in recombination at sites 1a, 1b, 2a and 2b (Fig.11). Plasmid DNA was isolated from the overnight cultures of evolution vectors, pEVO-loxltr1a-lib1, pEVOloxltr1b-lib1, pEVO-loxltr2a-lib1 and pEVO-loxltr2b-lib1, and digested with *Nde*1 followed by PCR (Fig.11a) and also with *BsrG*1-*Xba*1 to analyse for recombination at the respective targets (Fig.11b).

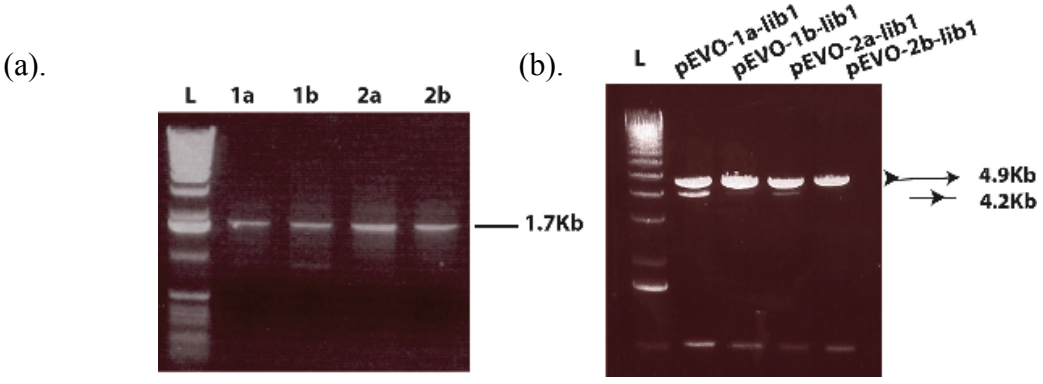


Fig.11. Activity of mutagenised Cre library on different loxltr subsets in the first evolution cycle.

(a). PCR product of 1.7 Kb reflecting recombination after *Nde*1 digest of the evolution vectors pEVO-loxltr1a-lib1, pEVOloxltr1b-lib1, pEVO-loxltr2a-lib1 and pEVO-loxltr2b-lib1.

(b). *BsrG*1-*Xba*1 of plasmid DNA to analyse for recombination at loxltrs 1a, 1b, 2a and 2b at the first evolution cycle.

The lower 4.2Kb band corresponding to the recombined product was evident for loxltrs1a and 2a from *BsrG*1-*Xba*1 digest. The 1.7Kb PCR product corresponding to each substrate was digested with *BsrG*1-*Xba*1 and the successful recombinase libraries were sub-cloned back into the respective evolution vectors to start the next cycle. This result showed that

splitting the mutations facilitated recognition by the mutagenised Cre library, and hence served as a starting point for subsequent directed evolution cycles. The aim now was to enrich through each cycle functional recombinases recognizing these four targets and then merging them in a combinatorial approach to carry out evolution using the next higher sub-sets loxltrs1 and 2 as targets.

5.6 DNA shuffling and stringent selection to generate functional recombinases.

The directed evolution strategy described here also takes into consideration two facts. First, recombinases capable of recognising the loxltr sites were shuffled at every third cycle of evolution. DNA shuffling is a process of randomly recombining different sequences of functional genes in a combinatorial fashion. This method searches much larger regions of sequence space with a bias towards functional sequences of improved fitness. The technique of DNA shuffling used here was that of random fragmentation and PCR reassembly (Stemmer, 1994). The recombinases in each library corresponding to each loxltr were randomly fragmented using sonication, and these fragments were reassembled by PCR (described in the materials and methods section) without primers through pairing of homologous sequences. This process creates a library of chimeric sequences containing crossovers between different parent sequences to create more genetic diversity based on recombination. The primerless PCR product is then used as a template for a regular PCR using defined primers flanking the recombinase coding region to amplify the reassembled strands to give a full-length product. The outcome of a typical shuffling reaction is shown in Fig.12.

The final 1Kb PCR product corresponding to the shuffled recombinase libraries were digested with *BsrG*1-*Xba*1 and sub-cloned back into the evolution vectors. We used a molecular breeding protocol where every third cycle of evolution was shuffled. DNA shuffling typically results in a reduced fitness in the library in the F1 generation, however, with the subsequent screening and selection there is rapid improvement in the F2 and later generations as a result of a combination of different mutations in the shuffled libraries (Fig.13).

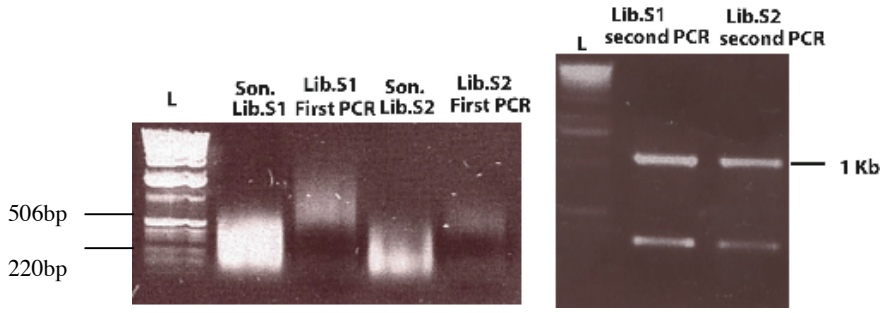


Fig.12. DNA shuffling. (a) Son.Lib.S1 and Son.Lib.S2 correspond to two sonicated recombinase libraries. Lib.S1 First PCR and Lib.S2 First PCR are the products of the primerless PCR, typically the smear moves up as recombinase coding regions are amplified depending on homology.
(c) PCR amplification using defined primers flanking the recombinase coding region produces a 1Kb band as shown. The lower band corresponds to a background band resulting from amplification of vector sequences without recombinase. L= 1Kb ladder.

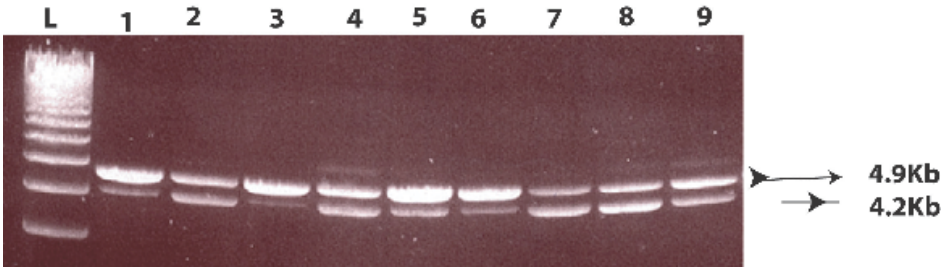


Fig.13. *Bsrg*1/*Xba*1 digestion of pEVO-loxltr1a libraries for first 9 cycles of evolution. Note that the cycles 3, 6 and 9 (shuffled cycles) display a reduced recombination efficiency reflected by the intensity of the lower 4.2Kb band, which rapidly improves in the subsequent cycles. The arabinose inductions for the recombinase libraries are listed in Table1.

Apart from shuffling, the stringency of selection was increased by concomitantly reducing the L-arabinose concentration used to induce expression of recombinases. For all the loxltr targets the recombinases were expressed with 200µg/ml L-arabinose in the first cycle. The next cycles had reduced induction levels to select out more functional recombinases and in this way the concentration of the inducer was made lower in the

following cycles till a minimal induction resulted in recombination. The arabinose induction used in each evolution cycle for all the loxltr subsets are shown in table1. Another very important aspect of the evolution cycles was the library size. Throughout the evolution process, the library size was maintained at 10,000 clones or above. In case of decreased library size the cycles were repeated till a satisfactory colony number was reached.

Table1. L-arabinose concentration (μg/ml) used for inducing recombinase expression for each evolution cycle corresponding to all the loxltr sub-sets and target loxltr.

Evolution Cycle number	L-Ara (μg/ml) for loxltr1a	L-Ara (μg/ml) for loxltr1b	L-Ara (μg/ml) for loxltr2a	L-Ara (μg/ml) for loxltr2b	L-Ara (μg/ml) for loxltr1	L-Ara (μg/ml) for loxltr2	L-Ara (μg/ml) for loxltr
1	200	200	200	200	200	200	200
2	200	200	200	200	200	200	200
3	200	200	200	200	200	200	200
4	200	200	200	200	200	200	200
5	200	200	200	200	200	200	150
6	200	200	200	200	200	200	150
7	75	75	100	200	100	100	125
8	10	50	50	100	100	100	100
9	10	10	10	50	100	100	50
10	10	10	10	50	75	75	50
11	0	0	0	10	50	50	50
12	0	0	0	5	10	10	50
13	0	0	0	5	10	10	30
14	0	0	0	2	10	10	30
15	0	0	0	0	10	10	30
16				0	5	5	30
17				0			30
18				0			25
19							25
20							15
21							15
22							15
23							10
24							5
25							5

5.7 Enrichment in libraries with efficient recombinase clones through cycles of evolution at different loxltrs.

The number of evolution cycles required to obtain highly efficient recombinases for each loxltr is not pre-determinable and varied with the evolvability of the starting mutagenised recombinase library with respect to the loxltr targets. The final aim was to have good library size and quality at minimum induction.

Among all the four starting *loxltr* targets, recombinases recombining loxltr2b was relatively difficult to evolve, compared to the other three. By cycle 9 the L-Ara concentration to induce recombinase expression was decreased to 10µg/ml for loxltrs 1a, 1b, and 2a and 50µg/ml for loxltr2b. Functional libraries were obtained even in absence of induction by cycle 10 for loxltrs1a, 1b, and 2a and by cycle 15 for loxltr2b. Successive cycles of evolution, shuffling, screening and selection led to an increase in recombination at the different loxltr subsets. Restriction analysis of recombinase libraries using *Bsrg*1/*Xba*1 digestion, at cycle 1 and cycle 10 are compared for loxltrs1a to 2b (Fig.14a). The tenth libraries for all the four target sites had distinct improvement in recombination efficiency compared to the first cycle. The evolution process was carried out for 15 cycles for loxltrs1a, 1b, and 2a and a total of 19 cycles were performed for loxltr2b. The evolution cycles were stopped when a satisfactory library quality was obtained at minimum possible induction.

5.8 Sequence analysis of evolved recombinases for loxltrs 1a, 1b, 2a and 2b.

Individual recombinases were sequenced to determine amino acid changes resulting in altered specificity. The comparative analyses of several recombinases with the same specificity would allow the identification of important residues which confer the change in specificity. Individual recombinases clones were randomly picked from evolution cycles 8, 9 and 10 for all the four loxltr subsets and analysed functionally using *Bsrg*1/*Xba*1 digestion and through sequence analysis.

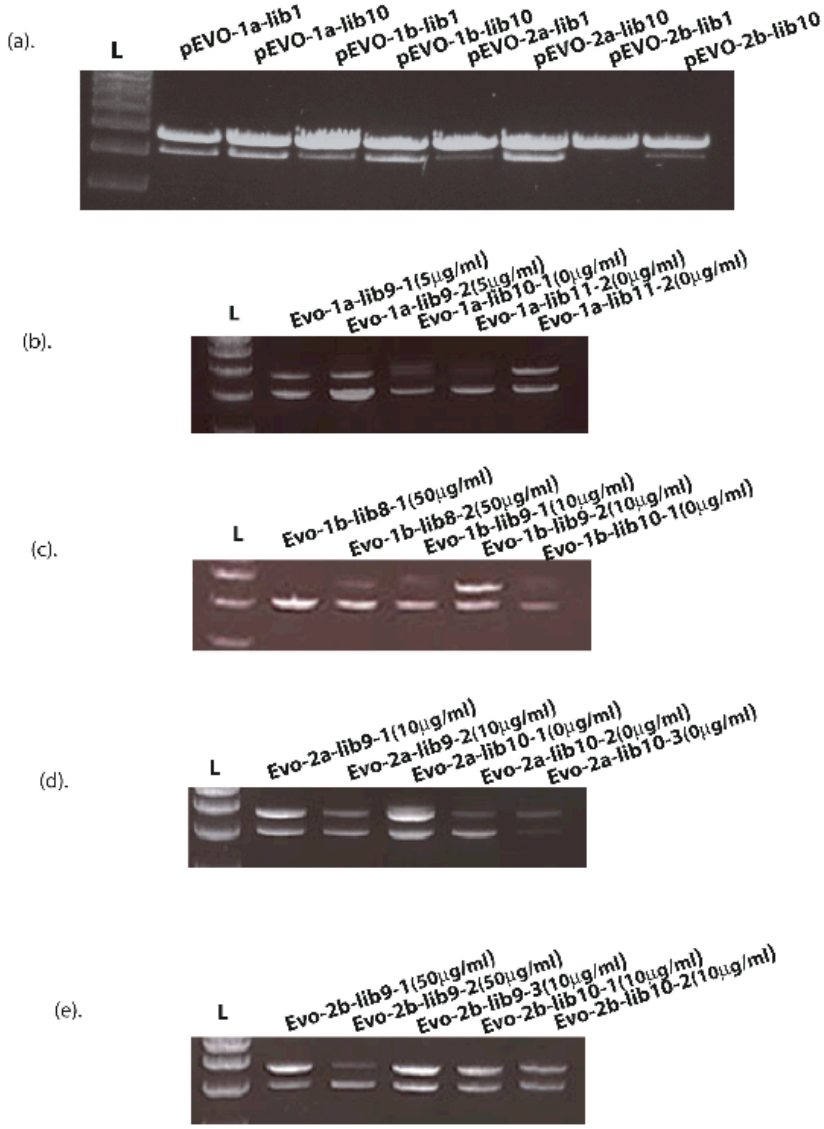


Fig.14. **Restriction analysis of recombination mediated by loxltr subset libraries and individual clones**. (a). *Bsrg*1/*Xba*1 digestion of recombinase libraries for loxltrs1a, 1b, 2a and 2b at cycle 1 and cycle 10 of evolution. The recombinase libraries are induced with 200μg/ml arabinose at cycle1 for loxltr1a, 1b, 2a and 2b and at cycle10 with 10μg/ml for loxltrs 1a, 1b, and 2a and 50μg/ml for loxltr2b. Improvement in recombination efficiency is evident at cycle 10 for all the loxltr subsets.

(b) to (e). *Bsrg*1/*Xba*1 digestion of some selected recombinase clones for loxltrs 1a, 1b, 2a and 2b, the respective cycle numbers and L-Ara induction concentrations are shown.

Table2. Amino acid mutations in some selected recombinases for loxltrs 1a to 2b. The residues in red are known to contact DNA from the wt Cre crystal structure. The hyprevariable residues are in bold for both contact and non-contact sites.

Evolved recombinases for loxltr1a, 1b, 2a and 2b.	Amino acid mutations.
1a.9.1(5µg/ml ara)	K25E, D29G, A127V, E138G, M149I, D341N
1a.9.2(5µg/ml ara)	M28A, M97I, V125I, I166T, G229D, D232V
1a.10.1(0µg/ml ara)	G82S, E150A, L171V, V204A, I320M
1a.11.2(0µg/ml ara)	A175G
1a.11.3(0µg/ml ara)	N3D, V7L, K62R, M97L, E138G, D329A
1a.11.7(0µg/ml ara)	D29A, V145I, E262Q, S305P, T316A
1b.8.1(50µg/ml ara)	N59T, I88V, I166V, I174M, S186T, S214I, P234L, V242A, **E262A**
1b.8.2(50µg/ml ara)	V7L, K62N, A112V, R146C, L161Q, F163L, G208R, R241P, **E262G**, D341G
1b.9.1(10µg/ml ara)	V7L, Q94L, S108G, E123K, N151T, T268A, N317T, I320M
1b.9.2(10µg/ml ara)	V7L, C155R, **E262Q**, N317T, I320S
1b.10.1(0µg/ml ara)	N3D, V7L, S108C, D143N, **E262Q**, N317T, I320S
1b.10.2(0µg/ml ara)	N3D, V7L, S114P, D189G, **E262G**, I306T
2a.9.1(10µg/ml ara)	L5P, L14F, Q156R, D189G, M193A, L215Q, I320S
2a.9.2(10µg/ml ara)	G93A, A112T, E138G, D143G, D278G, T316A
2a.10.1(0µg/ml ara)	V7L, Q94L, R118G, A127P, D153N, A178P, N319S
2a.10.2(0µg/ml ara)	N10H, P107I, T253S, T316A, I320S
2a.10.3(0µg/ml ara)	T6S, D21E, Q94L, L98I, E138G, G198S, A231V
2a.10.5(0µg/ml ara)	T19M, Q35R, E150V, G191E, G198S, S254C, D278G
2b.9.1(50µg/ml ara)	V7L, V16A, M30V, Q35P, N59S, A131V, D157E, A249T, R259C, **E262Q**, **N317T**, **I320S**, T332A
2b.9.2(50µg/ml ara)	V7L, R24H, M28I, A36V, F37L, D73Y, Y77H, P107L, A131V, T206A, **E262A**, D277G, **N317T**, **I320S**
2b.9.3(50µg/ml ara)	V7L, V16A, M30V, Q35P, S108G, G208R, **E262A**, **N317T**, **I320S**, G342W
2b.10.1(10µg/ml ara)	V7L, V16A, M30V, V71A, N111S, Q156K, A249T, Q255R, R259C, **E262Q**, **N317T**, **I320S**
2b.10.2(10µg/ml ara)	N3D, V7L, M30V, Q35P, S108G, A249T, **E262Q**, A267G, **N317T**, **I320S**
2b.10.3(10µg/ml ara)	V7L, V16A, M30V, Q35P, Y77H, G93C, Q156K, R259C, **E262Q**, **N317T**, **I320S**

At least 10 different clones of evolved recombinases for the four loxltr subsets were sequenced and functionally analysed by restriction digestion, the recombination competence at respective target site (Fig. 14b-14e) and the corresponding amino acid changes in the six best recombinases per loxltr subset are shown in Table2.

The average number of amino acid mutations for recombinases recognising loxltrs1a, 1b, 2a ranged from 6 –10 per clone though interestingly the clone 1a.11.2 had just one

mutation suggesting that one single mutation can be sufficient to allow better recombination at loxltr1a (Table2). Most of the mutations were at non-contact sites, and the mutations were not hypervariable especially for the sequenced clones of recombinases for loxltrs1a and 2a, hence it was not possible to assign a particular mutation to a particular loxltr. But there were some interesting mutations, which are M97L or M97I for loxltr1a recombinases, S108G/C and E262G/A/Q for loxltr1b recombinases and G198S for loxltr2a recombinases. The mutations have been mapped onto the wt Cre secondary structure to get an idea of their distribution (fig.15). In case of individual sequenced recombinase clones for loxltr2b, 70% of the sequences had Arg259 mutated to Cys, and all the clones contained Glu262 mutation to Gln or Ala. This is interesting because both the positions 259 and 262 make sequence specific contact with the lox site around base pairs 5-7 in the half-sites. Interestingly all the clones for loxltr2a had N317T and I320S mutations. These residues surround the helix M and have been previously implicated in broadening or shifting of mutant Cre specificity (Hartung and Kisters-Woike 1998). (The mutations will be discussed in greater detail in the Discussion section).

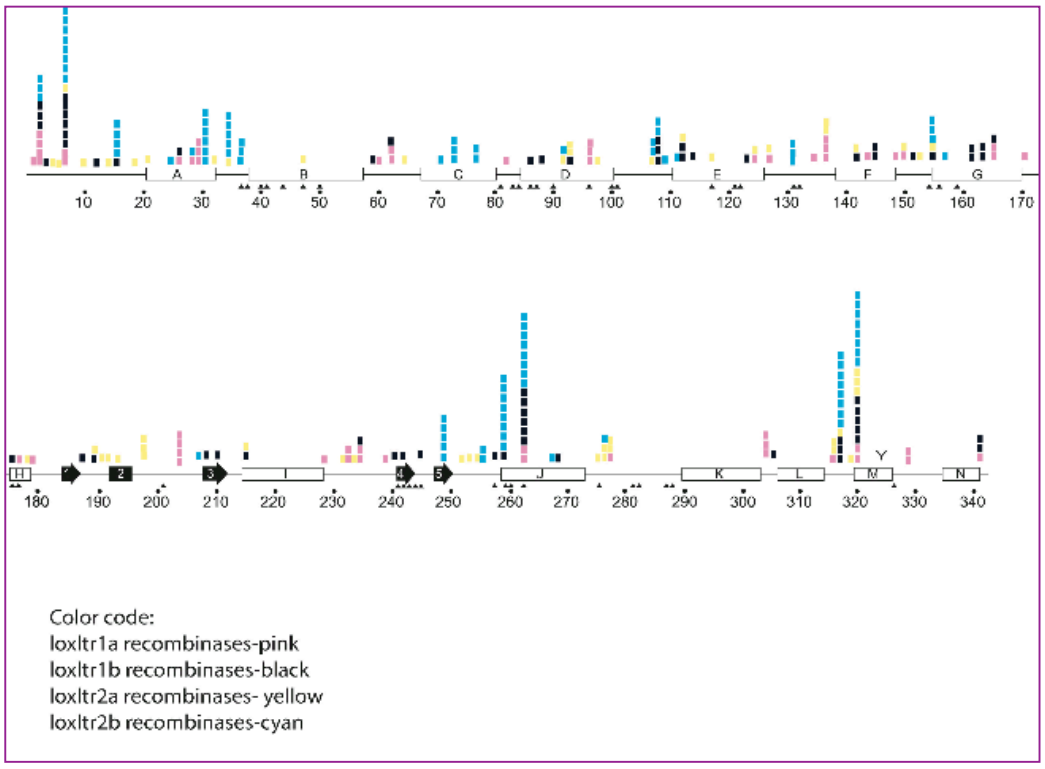


Fig.15. Mapping of mutations onto Cre secondary structure. The colour code of the mutations for the respective recombinases are shown in the figure. Amino acid mutations appearing more than once add to the bar. The secondary structural elements of Cre are shown as bars for α-helices (AtoN) and arrows for β-sheets (1to5). The catalytic tyrosine is denoted by "Y".

5.9 Generation of *loxltr* specific recombinases using combinatorial SliDE.

Once libraries of functional recombinases were generated which recombined these four subsets, the libraries recombining *loxltrs* 1a and 1b and *loxltrs* 2a and 2b were pooled and shuffled and the next phase of evolution was started to achieve recombination at loxltr1 and loxltr2 respectively. The reason of this combinatorial approach was due to the fact that the half-site mutations in loxltr1 and loxltr2 targets were additively present in the half-sites of loxltrs1a and 1b, and loxltrs2a and 2b, respectively. DNA shuffling of recombinase libraries was performed to recombine and unify those mutations in the recombinases which are responsible for conferring specificity to the lower sub-sets loxltr1a, 1b, 2a and 2b. The evolution protocol was the same as before. The first cycle with the shuffled libraries were induced with 200μg/ml arabinose. The expected PCR product after *Nde*1 digest, corresponding to recombination was obtained for both loxltr1 and loxltr2 when shuffled recombinase libraries for loxltrs (1a and 1b) and (2a and 2b) respectively were used. However no recombination was obtained when the randomly mutagenised Cre library (starting library) was used, which reflects that combining recombinase libraries which separately recognize and recombine the mutations in the split loxltr subsets, can be used to achieve recombination at loxltrs 1and 2, and thus we started evolving functional recombinase libraries for loxltrs 1and 2. The synergistic effect of the shuffled libraries of the lower subsets led to the generation of recombinases which recombined loxltrs1 and 2. The improvement in library quality was very slow for both loxltrs1 and 2, especially for the former. This could be attributed to the increased complexity in these targets compared to the former split targets. The libraries were induced with 200μg/ml arabinose till cycle11 for loxltr1 and cycle9 for loxltr2 to stabilize the pool of positive candidates in the recombinase libraries (Table1). A total of 20

evolution cycles were carried out for loxltr1 and 17 cycles for loxltr2 at a minimum possible induction of 5μg/ml arabinose.

Two recombinase clones, 1.17.4 (corresponding to a loxltr1 specific recombinase clone 4 of library17) and 2.12.14 (corresponding to loxltr2 specific recombinase clone 14 from library12) which displayed 50% recombination activity on their respective substrates were analysed for cross recombination activity and also for their activity on the native loxP target under induced (50μg/ml arabinose) and uninduced conditions (Fig.16). Both the recombinases retained recombinational activity on loxP even in absence of induction and displayed over 50% recombination at loxP under induction, but interestingly showed no cross recombination, that is 1.17.4 did not recombine loxltr2 and 2.12.14 did not recombine loxltr1. Thus the recombinases retained their recombination potential at loxP the native substrate, but did not have completely relaxed specificity as they did not show cross-recombination. This reflects that through the course of directed evolution, the recombinases acquire novel specificity but retain memory of their original targets, but they are not necessarily absolutely substrate ambiguous.

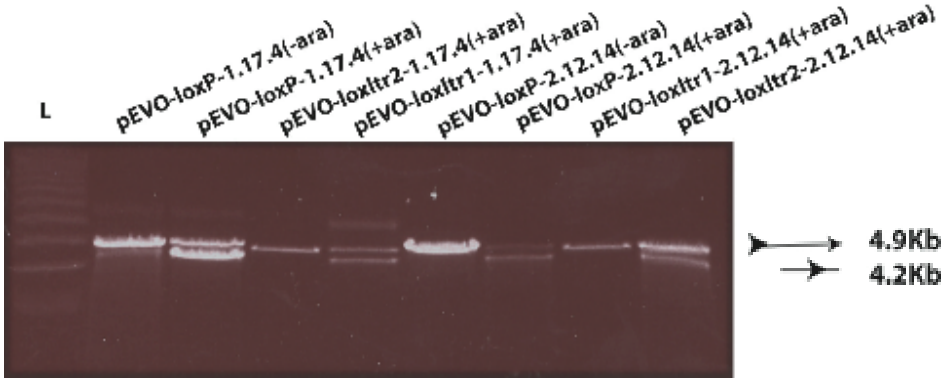


Fig.16. Restriction analysis of recombination mediated by loxltr1 specific recombinase, 1.17.4 on loxP and loxltr2, and loxltr2 specific recombinase, 2.12.14, on loxP and loxltr1.

The respective lanes have been labeled. The lower band represents the recombined band.

5.10. Sequence analysis of evolved recombinases for loxltr1 and loxltr2.

Individual recombinases were randomly picked from library17 for loxltr1 and library12 for loxltr2 for functional and sequence analysis. A total of 24 individual recombinase clones for loxltr1 and 20 individual clones for loxltr2 were analysed for recombination efficiency by *BsrG*1-*Xba*1 digestion. All the recombinases were induced with10µg/ml arabinose. Among the clones analysed for recombination at loxltr1, 14 showed above 50% recombination at the target site, and 17 recombinases displayed 50-60% recombination efficiency at loxltr2 (data not shown). All the sequenced recombinases had different sequences and no sequence was repeated reflecting that the evolution process was not complete and more variants were still being generated. Ten best clones were selected for sequence analysis for loxltr1 and loxltr2. Amino acid mutations of recombinases sequenced have been listed in Table3.

The number of amino acid mutations increased with the complexity of the target loxltr, and the average number of mutations in recombinases for loxltrs1 and 2 were between 10 to13. The sequenced loxltr1 recombinases had the glutamate262 mutated to either Gly/Arg/His/Ala. The other most prominent mutations were at positions 94, 108, 317 and 320.

All the loxltr2 recombinases had a chain of mutations appearing persistently N245Y, R259Y, E262Q, G263R of which the first three amino acids make contacts with the lox site in the wt Cre-loxP complex. The implications of these mutations will be discussed in greater detail in the Discussion section.

Table3. Amino acid mutations for some selected clones for loxltrs 1and 2. The recombinases were induced with 10μg/ml arabinose and dispayed between 50 to 60% recombination at their respective target sites. The residues in red are known to contact DNA from the wt Cre crystal structure. The hyprevariable residues are in bold for both contact and non-contact sites.

Evolved recombinases for loxltr1 and loxltr2.	**Amino acid mutations.**
1.16.3	L75F, V85A, **Q94L**, **S108G**, C155R, G198S, **E262G**, D278G, **N317H**, **I320S**
1.16.4	V85A, **Q94L**, **S108G**, G198S, D232E, **E262G**
1.16.11	V48I, R50Q, **Q94L**, V230A, S257T, **E262R**, **N317T**, **I320S**
1.16.14	V23A, L75F, V85A, **Q94L**, R101Q, **S108G**, G230A, **E262H**, **N317T**, **I320S**.
1.16.16	V7L, Q9H, N10S, V23A, D29E, R34H, K62E, L75F, V85A, **Q94L**, **S108G**, L164P, T206A, F239L, V247A, **E262H**, **N317T**, **I320S**
1.16.18	V85A, **Q94L**, **S108G**, D153N, G198S, D232E, **E262G**, **N317T**, **I320S**
1.17.2	V85A, **Q94L**, **S108G**, I174M, T206A, F239L, **E262A**, D278G, **N317T**, **I320S**
1.17.4	L75F, V85A, **Q94L**, **S108G**, G198S, **E262R**, D278G, **N317T**, **I320S**
2.12.1	T61, V7L, V16A, R24C, M30V, K57E, **G93C**, Q156K, I166V, **N245Y**, **R259Y**, **E262Q**, **G263R**, I272V, N317T, I320S
2.12.3	M44T, S51L, Y77H, **G93C**, P107L, **N245Y**, **R259Y**, **E262Q**, **G263R**.
2.12.13	N3D, V7L, V16A, Q35P, K57E, Y77H, **G93C**, P107L, A131V, S147L, Q156K, **N245Y**, **R259Y**, **E262Q**, **G263R**, N317T, I320S
2.12.14	V7L, V16A, V23A, M30V, Y77H, **G93C**, P107L, Q156K, R243G, **N245Y**, **R259Y**, **E262Q**, **G263R**.
2.12.16	V7L**, G93C**, Q156K, A175G, E222G, **N245Y**, **R259Y**, **E262Q**, **G263R**, N317T, I320S
2.12.18	Q35P, Y77H, **G93C**, Q156K, **N245Y**, **R259Y**, **E262Q**, **G263R,** N317T, I320S
2.12.9	M28I, M30V, Q35P, Y77H, **G93C**, A131V, **N245Y**, **R259Y**, **E262Q**, **G263R** K57E, **G93C**, **N245Y**, **R259Y**, **E262Q**, **G263R**, N317T, I320S

4.11 Tre recombinases.

The same combinatorial approach was followed up to achieve specificity for the final target loxltr from shuffled libraries of loxltr1 and loxltr2. Recombinase libraries pEVO-loxltr1-lib17 and pEVO-loxltr2-lib15 were mixed in equal proportions, shuffled and evolution was started with this library. The shuffled library was cloned into the pEVO-loxltr vector and recombinases induced with 200μg/ml of L-arabinose. The plasmid DNA from the overnight culture was digested with *Nde*1 and PCR was performed. The

expected 1.7 Kb band was obtained, which once again proved that combining two recombinase libraries (for loxltr1 and loxltr2) which separately recognize the mutations in the right and left arms in the final target did work to achieve recombination at loxltr. The PCR product was sub-cloned back into the loxltr evolution vector and the final stage evolution was started to achieve specificity at loxltr. The arabinose concentration used to induce recombinase expression at each cycle is shown in table1. A total of 25 cycles of evolution were performed for the final loxltr substrate till considerable recombination was obtained under low induction with 5μg/ml arabinose. The battery of recombinases recombining loxltr were termed **Tre** recombinases (for recognizing and recombining a sequence from a long 'terminal repeat').

The summary of this very extensive evolution process have been presented in Fig.17. The most noteworthy and novel feature of this strategy was that, it was possible to generate functional recombinases that recognize the asymmetric loxltr by combining recombinases which separately recognize the loxltr subsets with different mutations all present in the final substrate. Such a combinatorial approach can thus be applied to generate recombinases for highly variant asymmetric sites compared to the native or original target. A total of 126 evolution cycles were performed through all the subsets to achieve recombination at the final target loxltr.

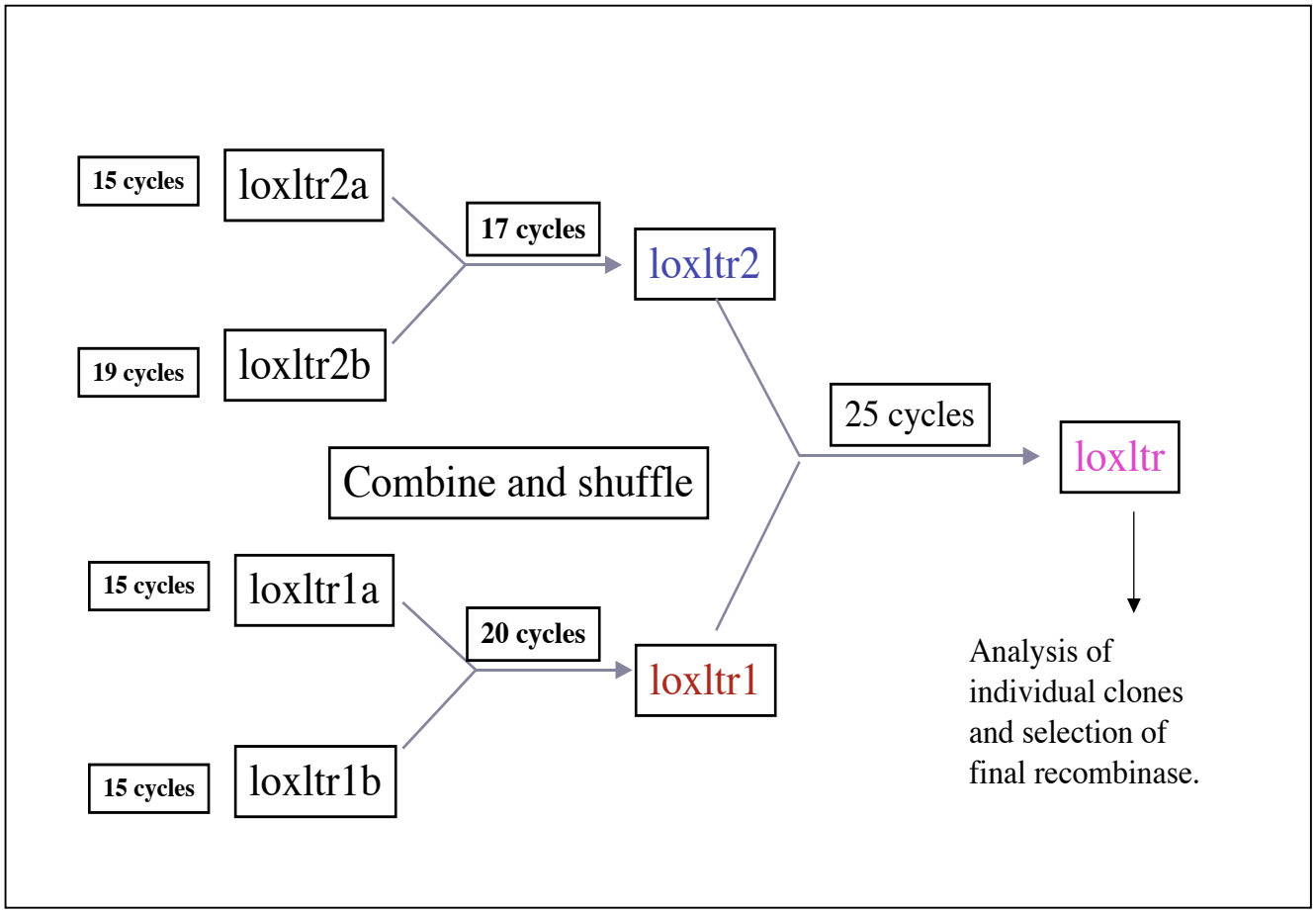


Fig.17. Summary of the directed evolution process. The number of evolution cycles performed for each loxltr subset is shown besides the respective targets, and the merging of specific subset libraries are shown with arrows.

5.12 Selection of final Tre recombinase.

Evolution of loxltr specific recombinases was halted after the 25th cycle at an induction of 5μg/ml arabinose. Fifty individual recombinases were randomly picked, induced with 5μg/ml arabinose and functionally analysed using *Bsrg*1/*Xba*1 digestion. Among the clones analysed, 17 clones displayed 40-50% recombination at loxltr at the induction used, the other clones showed less and were excluded. *Bsrg*1/*Xba*1 digestion of 10 recombinases from the 17 that had satisfactory recombination at loxltr is shown in Fig.18a. These clones were sequenced for mutation analysis and amino acid mutations of eight of the best sequenced recombinases is shown in Table 4. Again as for all other sequenced recombinases none of the sequences of the Tre recombinases were repeated reflecting the presence of diverse clones in the evolved library and also that the evolution process was not yet completed.

The best clone (clone 42, Fig.18a) screened from the 25th cycle, that recombined loxltr to 50% was selected for further analysis in *E.coli* and mammalian cells. The chosen recombinase has been termed as Tre25.

Tre25 was cloned into pEVO-loxP to produce pEVO-loxP-Tre. The specificity of Tre25 for the target loxltr and the native loxP was compared by transforming the two vectors, pEVO-loxP-Tre and pEVO-loxltr-Tre into *E.coli* DH5α, and then inducing the recombinase at different arabinose concentrations ranging from zero to 100μg/ml.

Tre25 recombined loxltr to a maximum extent of 80%, whereas at the same induction Tre25 recombined loxP to a maximum extent of 50% (Fig.18b and c). No recombination was observed at either substrate in absence of induction and no improvement in recombination efficiency was observed beyond 50μg/ml arabinose induction. Thus Tre25 retained some specificity at loxP though the efficiency of recombination was higher for loxltr.

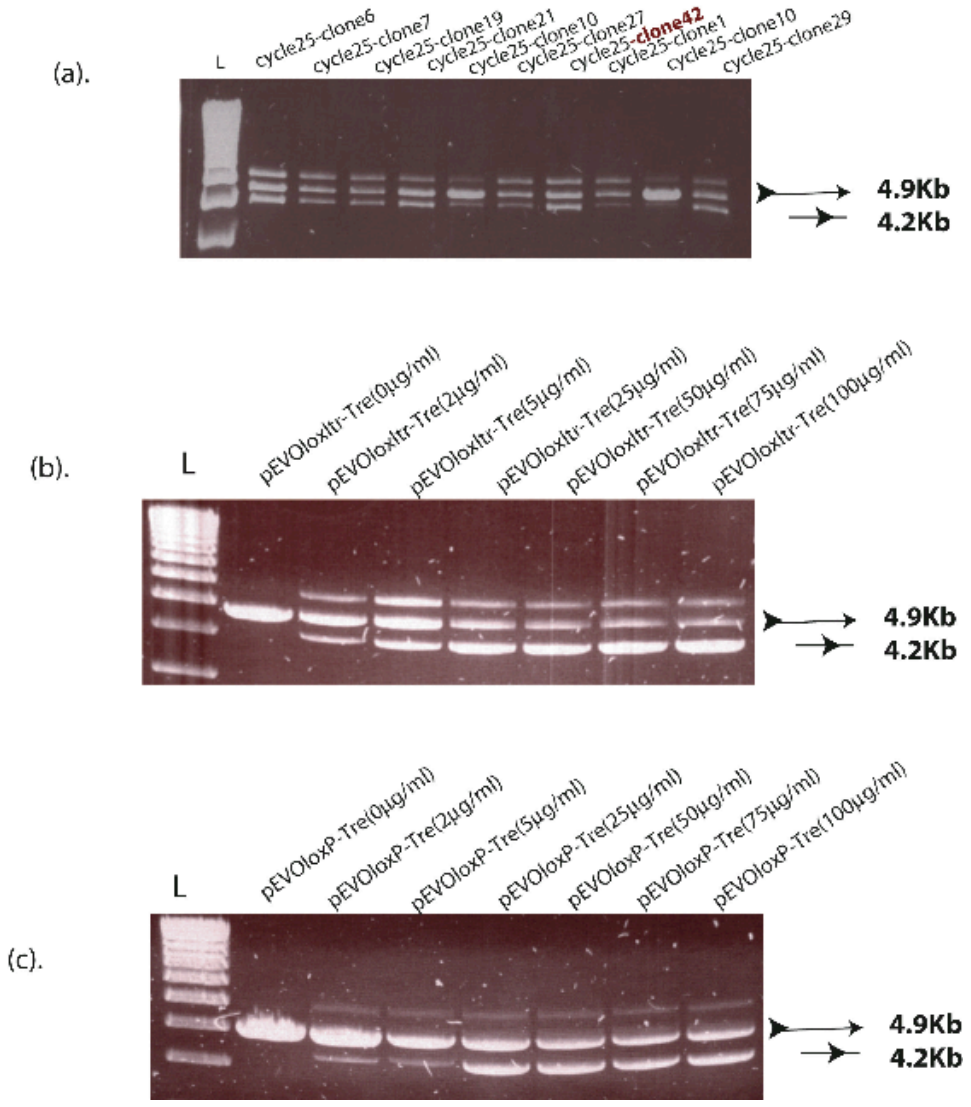


Fig.18. Screening of Tre recombinases by restriction analysis and recombination specificity of selected Tre25 at loxltr and loxP. (a). *Bsrg*1/*Xba*1 digestion of 10 different clones at 5µg/ml arabinose induction from the loxltr library after 25 evolution cycles. The clone 42 (termed as Tre25) marked in red was selected for further analysis.

(b) Recombination at loxltr target mediated by Tre25 (labeled as 'Tre' in the lanes) at different arabinose inductions. The inductions used are shown in the figure.

(c) Recombination at loxP target mediated by Tre25 (labeled as 'Tre' in the lanes) at different arabinose inductions.

5.13 Specificity analysis of Tre15 and Tre25.

Laboratory evolution can provide insight into the mechanisms of adaptive molecular evolution. Evolved enzymes tend to display broadened substrate specificity, which suggests that enzymes fluctuate through more specific or 'specialist' state to a more 'generalist' and gradually acquire a more 'specialist' state during evolution ((Matsumura and Ellington 2001). To test whether this trend can be seen in our *in vitro* evolution, we analysed the specificity of a recombinase from an earlier evolution cycle (cycle15) with the final selected Tre25 from cycle 25.

Tre15 and Tre25 were cloned into the pEVO versions of loxP and all the loxltr substrates and analysed for recombination by restriction analysis at 50μg/ml induction. Tre15 recombined loxltr to almost 50% whereas Tre25 recombined to almost 80%. Tre15 displayed much more relaxed specificity and recombined loxltrs1, 2, 1b, 2a, and 2b, whereas at the same induction Tre25 recombined only loxltr2b (data not shown). Both the variants recombined loxP with Tre15 recombining loxP a greater extent in comparison to Tre25. This result is interesting as it shows that as evolution proceeded from cycle 15 to cycle 25, the Tre recombinases slowly lost their specificities for most of the loxltrs, though they retained recombination potential at loxP.

5.14 Sequence analysis of evolved Tre recombinases and identification of mutation hotspots.

The Tre recombinases had around 15 mutations per sequenced clone. There were several mutations which were lost during course of evolution to the higher target loxltr, and some mutations which rose later and became fixed suggesting that epistatic interactions occur during in vitro evolution.

The distribution of mutations of the recombinases for loxltr, loxltr1 amd loxltr2 revealed a clustering of residues through the course of evolution. High density of hypervariable codons were present in helix J followed by helix D (Fig.19). The 4-5 linker and L-M linker had clustering of mutations at positions 245 and 317 respectively. Helix J amino

terminus interacts with the major groove where Arg259 forms hydrogen bonds with N7 and O6 of guanine 10 (Guo et al., 1997). The structural determinants of substrate specificity are difficult to predict with sequence comparison alone, yet the mutations do give an idea of the plasticity residues contributing to directed molecular evolution.

Three clusters of mutational hotspots are evident in case of the sequenced Tre recombinases, the first at positions 85, 94 and 108, which were also present in 90% of the sequenced clones of loxltr1 recombinases (Fig.19). The fact that these codons were hypervariable in the Tre recombinases suggest that this cluster of mutations were contributed by the loxltr1 specific recombinase library. A second cluster of hypervariable codons were at positions 245, 259, 262 and 263. These residues were also hypervariable in clones of loxltr2 recombinases. The codons 259 and 262 were also hypervariable in loxltr2b recombinases and the 262 codon was also mutated in sequenced recombinases of loxltr1b (tables2, 3 and 4). It is possible that the first and second mutation clusters are responsible for facilitating recombination at loxltr1 and loxltr2 respectively. Shuffling of loxltr1 and loxltr2 recombinase libraries for directing evolution of recombinases for loxltr have combined these mutations and sequestered them in the Tre recombinases. The final clustering are at positions 317 and 320, which were recurrently mutated in all sequenced recombinase clones of loxltr2b, 1 and 2.

Apart from positions 259 and 262 the other residue of interest which was mutated in all the Tre and loxltr2 recombinases was Asn 245. This residue is the last in a stretch of amino acids which contact DNA in the wild type recombinase and was consistenly mutated to Tyr suggesting that this mutation was important for recognition of loxltr and also the subset loxltr2. The amino acid mutations of the selected Tre recombinase (Tre25.42) is shown in Table 4. None of the sequenced Tre recombinases had identical sequences, a property observed also for the other loxltr subset recombinases.

Table4. Amino acid mutations for some selected recombinase clones for loxltr. The recombinases were induced with 5µg/ml arabinose. The recombination proficiency of the clones can be seen from fig.12a. The residues in red are known to contact DNA from the wt Cre crystal structure. The hyprevariable residues are in bold for both contact and non-contact sites.

Tre recombinases from cycle25.	**Amino acid mutations.**
Tre25.42 (selected for cell culture)	V7L, Q9H, N10S, V16A, M30V, Q35P, K43E, Y77H, G93C**, Q94L**, A131T, I166V, K244R**, N245Y**, **R259Y**, **E262Q**, **G263R**, **N317T**, **I320S**.
Tre25.6	E22G, F37S, A84V, V85A, **Q94L**, **S108G**, K132N, I166V, A175S, **N245Y**, **R259Y**, Q281R, **N317T**, **I320S**.
Tre25.7	L5Q, V7L, P12S, L14S, P15L, V23A, A80V, V85A, **Q94L**, **S108G**, G198S, **N245Y**, **R259Y**, **E262Q**, **G263R**, T268A, **N317T**, **I320S**.
Tre25.27	T6I, V7L, Q9H, N10S, V23A, V85A, **Q94L**, K132N, I166V, A175S, D232G, **N245Y**, **R259H**, **E262Q**, **G263R**, **N317T**, **I320S**.
Tre25.29	L59W, Y77H, V85A, **Q94L**, **S108G**, **N245Y**, **R259Y**, **E262Q**, **G263R**, **N317T**, **I320S**.
Tre25.1	V7L, Q9H, N10S, K57E, V85A, **Q94L**, **S108G**, G198S, **N245Y**, **R259H**, **E262Q**, **G263R**, Q281R, **N317T**, **I320S.**
Tre25.10	V7L, Q9H, N10S, V23A, V85A, I88V, **Q94L**, **S108G**, M149I, I166V, **N245Y**, **R259Y**, **E262Q**, **G263R**, **N317T**, **I320S**.
Tre25.21	V7L, Q9H, N10S, V23A, R34H, V85A, **Q94L**, **S108G,** K132N, I166V, A175S, I225V, **N245Y**, **R259Y**, **E262Q**, **G263R**, **N317T**, **I320S**.

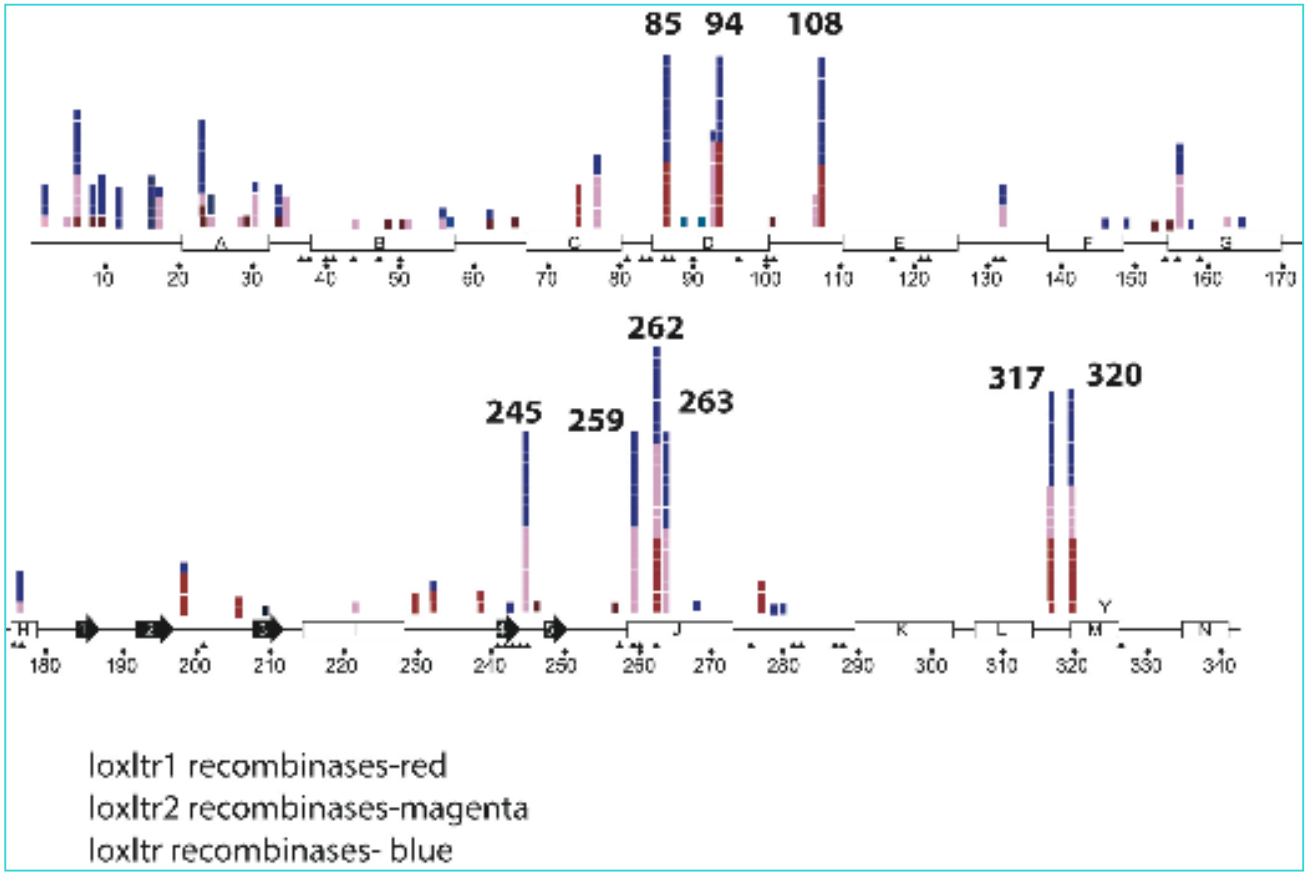


Fig.19. Mapping of mutations onto Cre secondary structure. The colour code of the mutations for the respective recombinases are shown in the figure. The positions showing hypervariability and sequestering of mutations from recombinases of loxltr1 and 2 are marked.

5.15 Analysis of Tre recombinase activity in mammalian cells.

Once the evolution process was stopped, the next important requirement was to investigate whether our evolved recombinase was active in mammalian cells. In this section and the coming sections the selected Tre25 recombinase would be referred to as just Tre recombinase for simplicity.

To evaluate the activity of Tre in mammalian cells, a lacZ based reporter assay was used. The reporter plasmid for analyzing recombinase activity contains a puromycin resistance gene (pac) flanked by loxltr or loxP sites, followed by a *lacZ* gene, such that recombination at target sites would turn on l*acZ* expression by removal of an intermediate stop sequence and cells subsequently stained with X-gal would be blue (Fig.20). The loxltr was cloned into this reporter vector to give psvltr, the loxP version is psvpaX.

Cre or Tre recombinase was cloned in a mammalian expression vector, containing the human cytomegalovirus major immediate early promoter/enhancer, a synthetic intron known to enhance the stability of the transcript, an ECMV IRES followed by an antibiotic resistance marker. The IRES facilitated the expression of the selection marker. The recombinases were cloned into the pIRES-Hyg version, and the vectors were re-named as pIRES-hyg-Cre and pIRES-hyg-Tre.

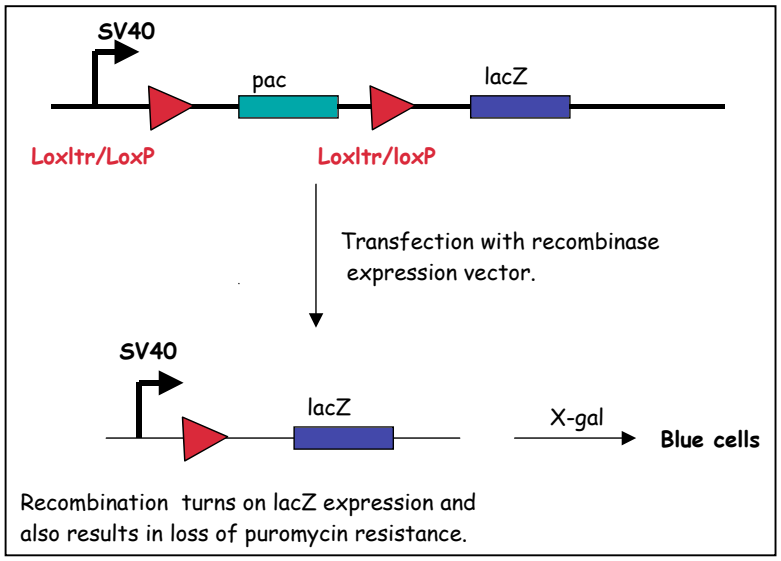
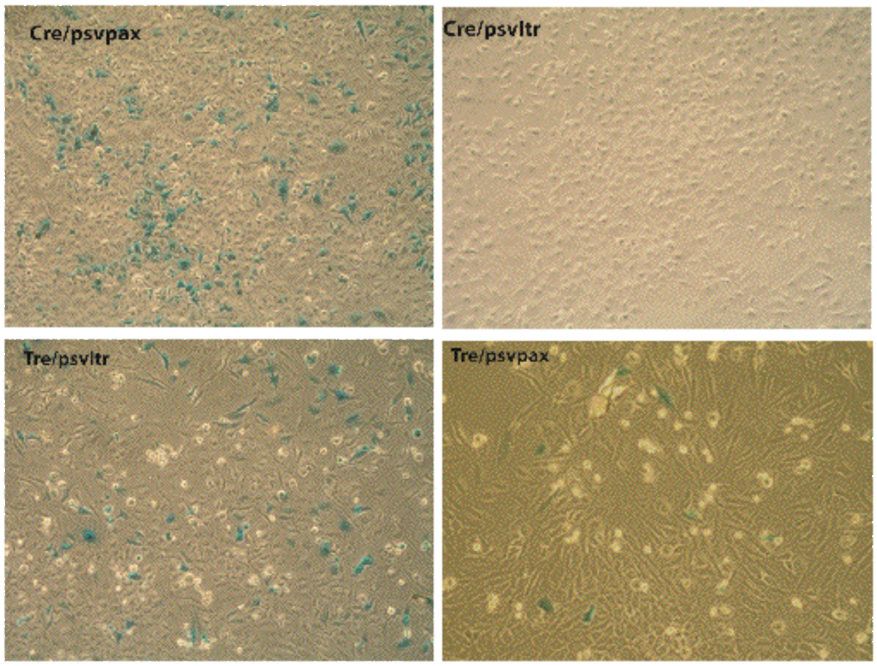


Fig. 20. Reporter assay for evaluating recombinase activity in Hela cells.

a)



b)

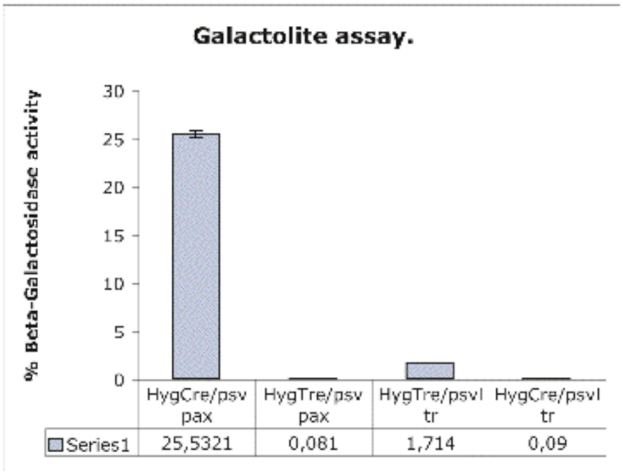


Fig.21. Assay for transient recombinase activity in Hela cells. (a) Hela cells were co- transfected with Cre or Tre recombinase and reporter plasmids. The cells were fixed and stained with X-gal 48 hours post transfection. Recombination deletes a 1.1 kb stop region (denoted as 'pac' for puromycin acetyl transferase, that prevents expression of lacZ. in fig.9) and results in turning on lacZ expression reflected by blue cells.
(b) Measurement of β-galactosidase activity in cell lysates of Hela cells co-transfected with recombinase and reporter plasmids. The β-galactosidase activity is presented as a percentage of the enzyme activity of cells after parallel transfection with the recombined forms of reporter plasmids psvΔpax and psvΔltr.

Hela cells were co-transfected with pIreshyg versions of recombinase plasmids and psvpax or psvltr reporter plasmids. Recombinase activity was evaluated 48 hours post transfection using X-Gal staining and also measurement of β-galactosidase activity. X-Gal stained blue cells are evident for Cre/psvpaX, Tre/psvltr and Tre/psvpaX transfections, but not for Cre/psvltr (Fig.21a). The β-galactosidase activity for transfection of Hela cells in parallel with recombined psvΔpax or psvΔltr was set to 100%. Co-transfection of pIreshyg-Cre with psvpaX generated almost 30% of β-galactosidase activity within 48 hours post-transfection (Fig.21b). As in *E. coli* assays, Cre did not recombine loxltr and Tre displayed relaxed specificity and recombined also loxP, but the recombination activity of Tre was higher for loxltr substrate than for loxP, as shown by the staining and measurement of β-galactosidase activity. The transient assay shows that the evolved recombinase, Tre, mediates recombination in Hela cells at a low but measurable activity in comparison to the wild-type Cre recombinase.

5.16 Tre mediated recombination at loxltr sites integrated in the genome.

The above results testify to the fact that Tre can mediate recombination in mammalian cells, however the more important question is whether it can mediate recombination in the genomic context, which is an absolute prerequisite for application of evolved recombinases as genomic surgery tools.

The loxltr and loxP versions of the reporter vector were used to generate stable puromycin resistant Hela cell lines to investigate whether Tre can recombine its target in a genomic context. As for the transient assays the stable loxltr or loxP cell lines were transfected with either Cre or Tre recombinase vectors, and transfected cells were analysed for recombination after 48 hours. Two stable cell lines each for loxltr and loxP were used for the assay. Initially the cell lines were transfected and stained with X-Gal to test for recombination, but only one of the loxltr cell lines (loxltrHela1) showed blue cells upon transfection with pIres-Hyg-Tre, whereas the loxP stable cell lines or the other loxltr cell line did not show blue cells. The fact that there were no blue cells in these cell lines could be due to silencing of the SV40 promoter driving the expression of the lacZ

gene or that somehow that plasmid had integrated in a manner thst destroyed the lacZ gene. I used PCR to detect recombination in stable cell lines. Transfection of the loxP stable Hela cell lines with pIres-Hyg-Cre gave the expected 724 bp band for recombination but no band was observed when loxltr stable cell lines were transfected with Cre (Fig.22).

Transfection of the Tre recombinase expression vector resulted in recombination only at loxltr target but, surprisingly not at loxP as detected by PCR, though the transient assays showed Tre mediated recombination at loxP. It could be that the threshold of recombinational activity of Tre at loxP were too low, and hence beyond detectable limits.

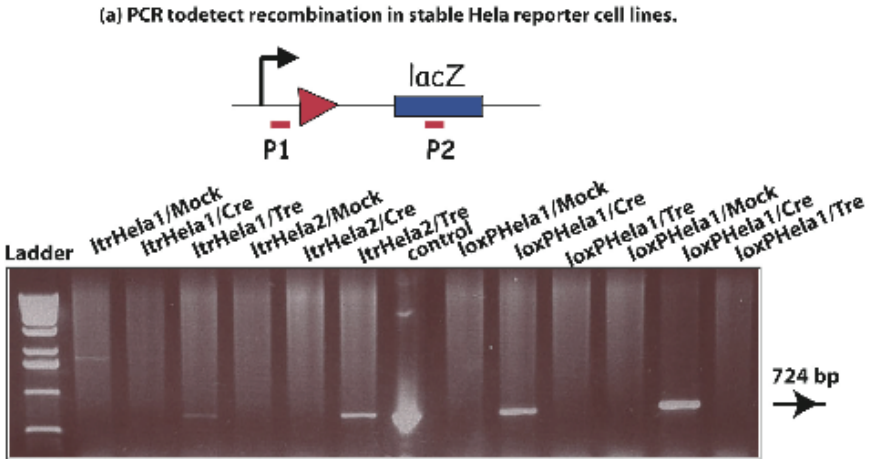


Fig.22. PCR detection of recombination in stable reporter cell lines. Recombination gives a 724 bp band after PCR using primers as shown in the schematic representation.

The loxP and loxltr Hela stables were transfected with mock (pIres-Hyg vector), pIres-Hyg-Cre, and pIres-Hyg-Tre vectors, and 48 hours post transfection genomic DNA was isolated from transfected cells and PCR was performed.

PCR detection of recombination in stable Hela reporter cell lines showed that Tre does recombine a loxltr target integrated into the genome.

To quantify the extent of recombination achieved in this assay we designed a probe to be used for Southern blot analysis (Fig.23). A DIG-labelled DNA probe corresponding to the lacZ region was generated. Genomic DNA was isolated from stable cell lines after transfection with respective recombinase vectors and were probed after overnight

digestion with *Avr*11 and *Cla*1. The reporter plasmids psvpaX and psvΔpax were similarly digested and hybridised with the same probe as positive controls. Recombination should result in a 950bp band in comparison to 2kb before recombination (schematics in Fig.23).

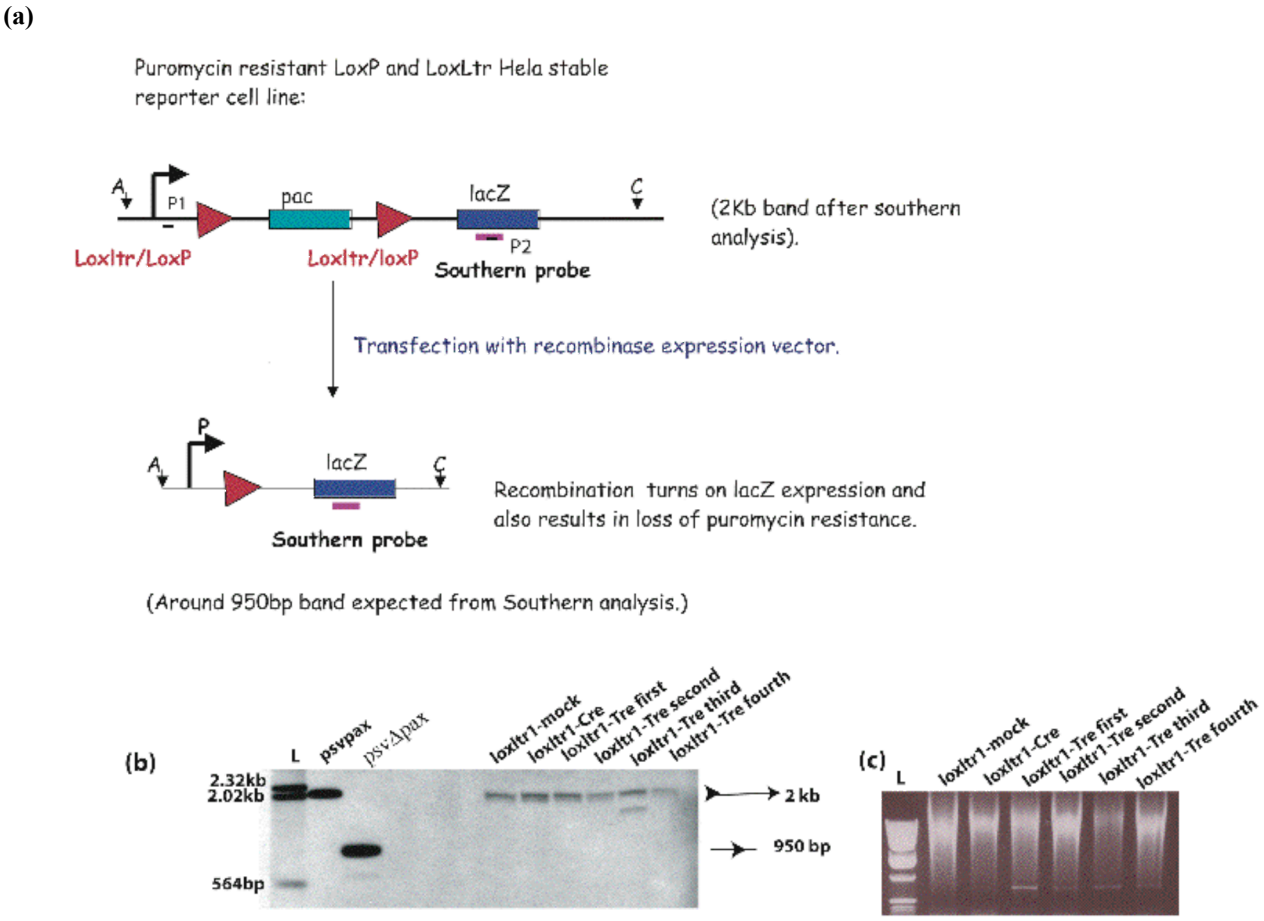


Fig.23. Southern blot in stable loxltr reporter Hela cell line.

(a) Schematics of southern blot detection of recombination in Hela stable cell lines loxltrHela1.

(b) Southern analysis of the loxltr1 cell lines transfected in four consecutive rounds with Tre expression vector as explained in text. DIG labeled DNA probe was used to hybridise digested Hela genomic DNA transferred onto nylon membrane.

(c) PCR detection of recombination using primers P1 and P2.

Southern hybridization of loxltr Hela1 DNA 48 hours post transfection with either mock, Cre or Tre recombinase vectors did not show the lower band for recombination. Thus it was decided to do four consecutive rounds of transfection of loxltr Hela cells with Tre vector to enrich for recombined cells. The loxltr Hela cells were transfected with Tre plasmid and 48 hours post transfection, the cells were passaged into two dishes, one for isolation of genomic DNA, and the other batch of transfected cells were used for re-transfecting with Tre plasmid.

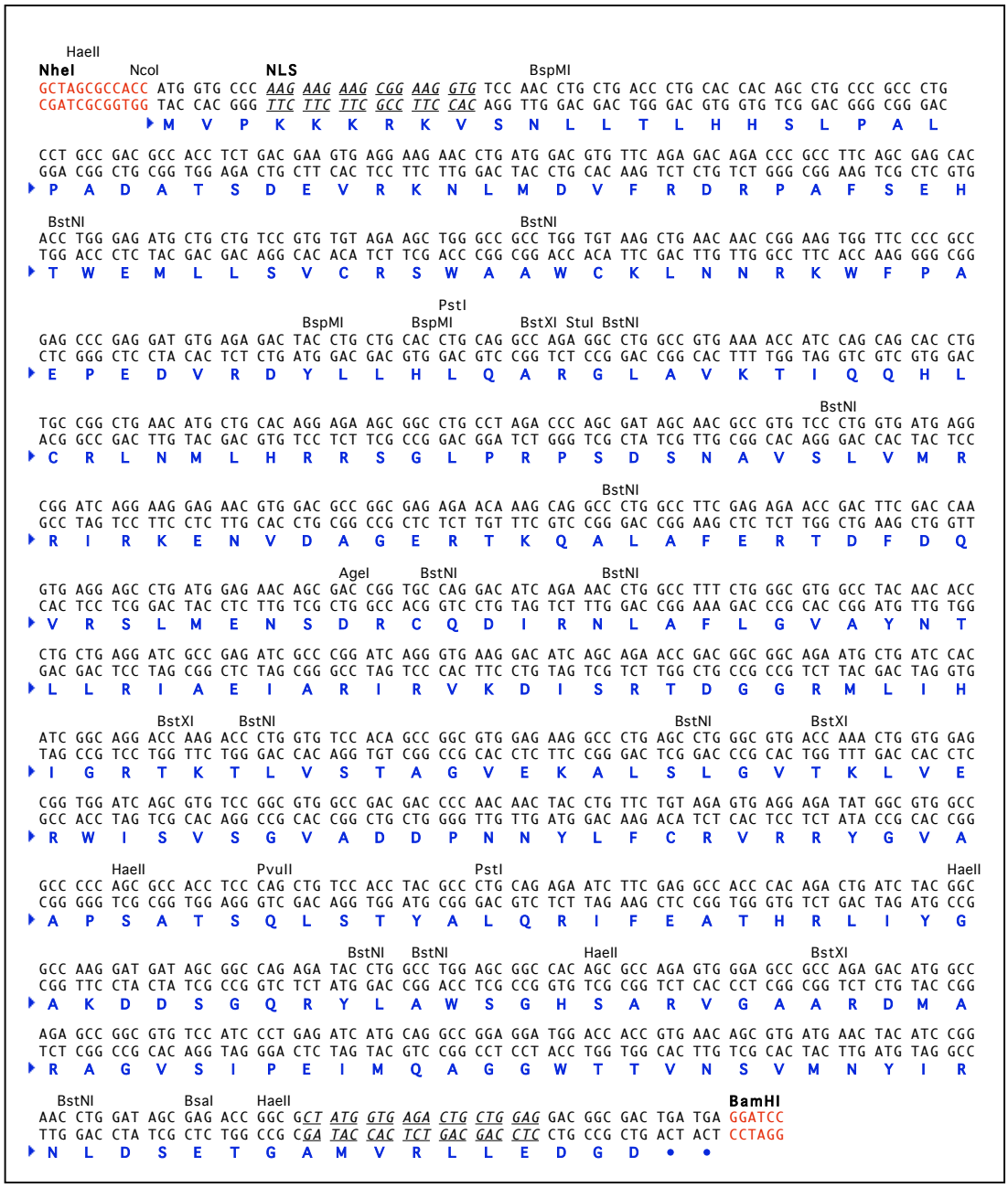
After four such rounds, the genomic DNA from the four batches were analysed for recombination using both PCR and southern hybridization. Though recombination was detected by PCR, southern blot failed to show any recombined band reflecting that recombination occurs, albeit at a low rate.

Tre mediates recombination in mammalian cells as shown by transient co-transfection assay and also recombines its target integrated into the genome as shown by PCR of reporter stable cell lines after transfection with the recombinase. The rate of recombination mediated by Tre was low. Improved recombination efficiency in mammalian cells was recently demonstrated for Cre recombinase by optimizing the codon usage to human codon usage preferences (Shimshek et al. 2002).

5.17 Codon-optimisation of Tre.

To improve the activity of Tre in mammalian cells, the codon usage of the recombinase was adapted to human codon usage preferences by introducing silent base mutations. The altered version is termed as hTre (humanized Tre). Apart from codon-optimisation an optimal Kozak consensus sequence was introduced for efficient translational initiation, an SV40 T-NLS was included for N-end rule stability (Varshavsky 1997), and two termination codons were introduced for efficient termination. Previously the *cre* gene was similarly modified to be more eukaryotic like (Shimshek et al. 2002). The encoded iCre (for improved Cre) was detected at about 1.6 fold higher concentration than conventional Cre when expressed from an identical vector, and the efficiency of iCre at DNA recombination was almost1.8 fold higher than conventional Cre. We hoped that the codon optimized hTre would be more efficient than Tre.

Sequence of codon-optmised altered hTre:



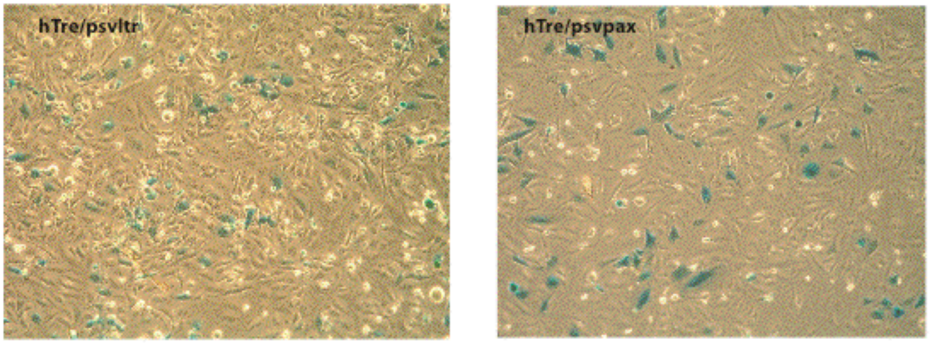
5.18 Transient co-transfection assay in Hela cells with optimized hTre.

The hTre expression vector was similar in design to the one used for Tre as before, only the selection marker was neomycin. The resulting vector was termed pIres-neo-hTre.

I used the lacZ based reporter assay to evaluate the activity of hTre in Hela cells in the same way as for the unoptimised version of Tre.

The codon optimized version of Tre expression vector was cotransfected into Hela cells separately along with either loxltr (psvltr) and loxP (psvpax) reporter vectors. As before the recombination activity was evaluated using X-Gal staining and measurement of β-galactosidase activity from Hela cell lysates 48 hours post-transfection.

a)



b)

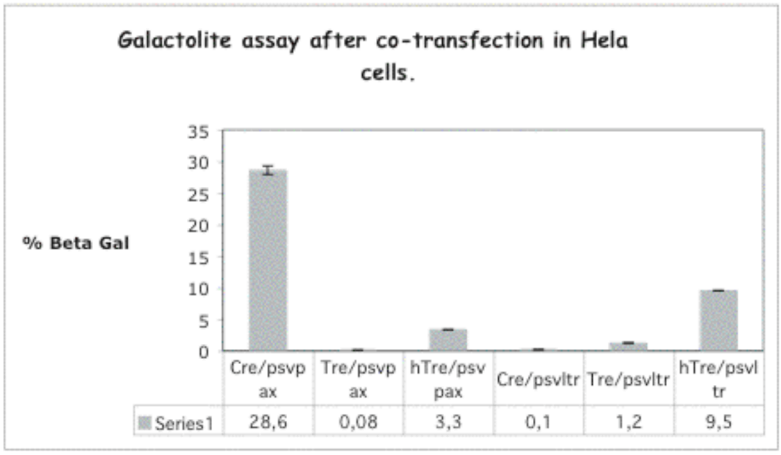


Fig 24. Assay for transient hTre recombinase activity in Hela cells. (a) Hela cells were co- transfected with hTre recombinase and loxP or loxltr reporter plasmids. The cells were fixed and stained with X-gal 48 hours post transfection.

(b). Measurement of β-galactosidase activity in cell lysates of Hela cells co-transfected with recombinase and reporter plasmids. The β-galactosidase activity is presented as a percentage of the enzyme activity of cells after parallel transfection with the recombined forms of reporter plasmids psvΔpax and psvΔltr.

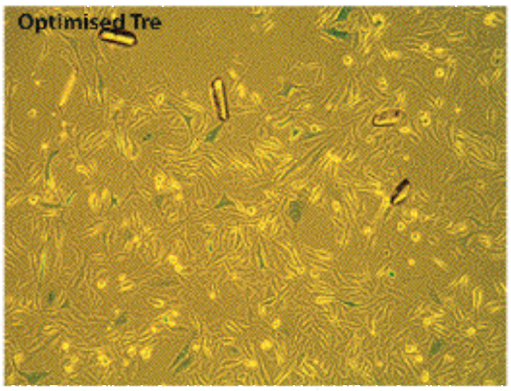
The transient assay using hTre with psvpax and psvltr plasmids showed an improved activity of hTre in mediating recombination in comparison to Tre (Fig.24). X-Gal staining of Hela cells co-transfected with hTre and loxltr plasmids clearly showed an increased proportion of blue cells, and also measurement of β-galactosidase activity from the cell lysates gave almost 9 times higher value for hTre and psvltr plasmids than when Tre was transfected with psvltr demonstrating the improved activity of hTre in Hela cells.

5.19 Activity of optimized hTre in the genomic context

The stable loxltr reporter cell line, loxltrHela1, was transfected with hTre vector, and 48 hours post-transfection the cells were fixed and stained with X-Gal. Parallel transfections were carried out with the mock vector, pIres-neo, the Cre vector and the hTre vector, and β-galactosidase activity in cell lysates were measured. There was a distinct increase in β-galactosidase activity in cells transfected with hTre, in comparison to Tre, and efficient recombination was also demonstrated by X-Gal staining (Fig.25).

These results showed that codon optimization of the evolved recombinase resulted in improved recombination efficiency. The loxltr and loxP stable cell line were transfected with hTre expression vector and neomycin pools were generated by selection for the integration of the hTre vector into the genome of loxltr and loxP stable cell lines. These cell lines were tested for recombination at loxltr and loxP. PCR results showed that the optimized Tre did recombine loxP and loxltr (data not shown), whereas the unoptimised version had not recombined loxP in the genomic context. Thus codon-optimisation did increase activity of the recombinase both at loxltr and at loxP. However the Southern blot in these stable cell lines did not work and failed to give recombined band. The reason for failure of Southern blot could be that the threshold of recombinase activity was still too low to be detectable by Southern. It could also be that the cell lines had multiple integrants of the lacZ reporter construct, or that the DIG labeled probe was not the best probe to use in this case. Hence we required another assay to detect extent of recombination in stable cell lines with the loxltr integrated in the genome.

a)



b)

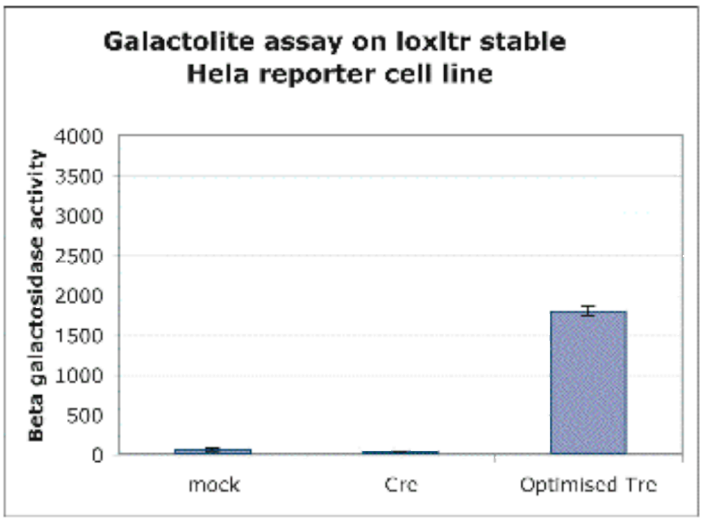


Fig.25. Recombination mediated by hTre in stable loxltr Hela reporter cell line. (a) The stable loxltr cell line was transfected with hTre expression vector and cells were stained with X-gal 48 hours post-transfection. (b). Measurement of β-galactosidase activity in cell lysates of the stable loxltr Hela cell line after transfection with mock, Cre and hTre vectors.

5.20 Repression of Tat mediated transactivation from an HIV-1 LTR using a model system.

To address the question whether recombination mediated by hTre can prevent Tat transactivation from an HIV-1 LTR promoter, we used an HIV-1 LTR based reporter system. The HIV-1 Tat protein is an essential viral regulatory factor which is involved in the transactivation of genes involved in the HIV long terminal repeat and plays a critical

role in viral replication. The loxltr sequence was cloned in an HIV-luciferase reporter vector (HIV_T2_LUC) within the HIV-1 HXB3 LTRs of the vector (Fig.26) and Tat activation assays were performed. As a control, the vector containing just one 5' loxltr sequence (HIV_T1_LUC) was also used in the Tat based assay.

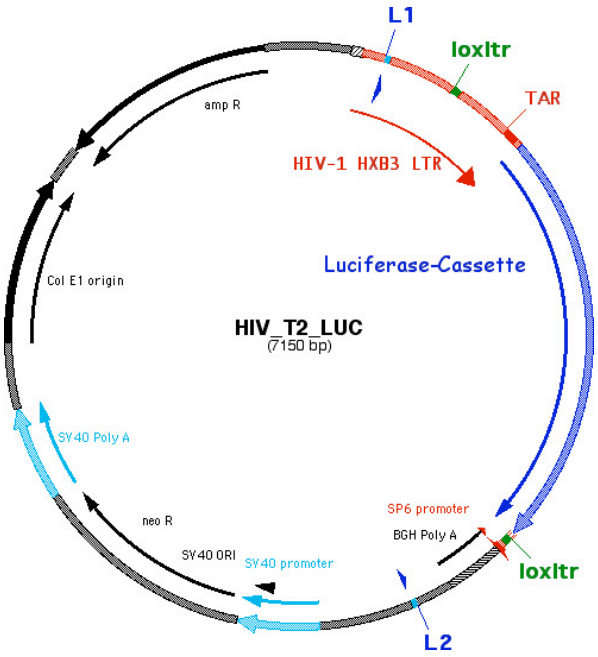


Fig.26. Map of HIV_T2_LUC vector. The vector has the loxltr sequence cloned within the HIV-1 HXB3 LTR. The 5' LTR sequence is complete with the TAR region where Tat binds and transactivates firefly luciferase expression. L1 and L2 denote the primer pair used for PCR detect for recombination. Recombination via the loxltr sites will delete the luciferase coding region.

Tat activation of the HIV_ T2_LUC vector results in expression of firefly luciferase driven by the LTR promoter. In absence of Tat there is very low level of luciferase expression. When Hela cells were co-transfected with hTre vector along with the Tat vector and HIV-T2_LUC vector, there was a three-fold decrease in luciferase activity as shown by the luciferase readouts (Fig27a). When the same experiment was performed with the HIV-T1_LUC vector which does not contain the second loxltr site, there was no decrease in luciferase expression (Fig27b).

To prove that the decrease in luciferase expression was a result of recombination and not blocking the Tat activity or transcription from the LTR promoter by the recombinase, we

performed PCR analysis for assaying for recombination from Hela DNA after Tat transactivation. Hela cells were co-transfected in a parallel experiment with the Luc_T1 or the LucT2 vectors along with the Tat vector and the hTre expression vector. As control, cells were also transfected with the luciferase vectors and Tat vector but not the recombinase vector. DNA was isolated from Hela cells 48 hours post transfection, and PCR was performed using primers in the upstream and downstream LTR sequences (primers L1 and L2, Fig.26). Recombination results in a band size of 756 bp due to removal of the luciferase coding sequence. In absence of recombination the band size is 2.7 kb. PCR of Hela DNA transfected with the HIV_T1_LUC vector containing just one loxltr site produced only the higher band both in presence and absence of recombinase, whereas when the HIV-T2_LUC vector was used, PCR of Hela DNA resulted both in the upper and lower bands, in presence of hTre indicating that hTre mediated recombination at the loxltr sites (Fig27c). Gel extraction of the PCR fragments followed by sequencing confirmed the precise excision of the loxltr1flanked sequence when the Luc-T2 vector was used (data not shown).

To investigate whether recombination occurs in a more physiologically relevant context of an LTR, Hela cells were transfected with the HIV_T2_LUC vectors to make stable cell lines. However, it was not possible to obtain a stable cell line. Most likely the high expression of luciferase was not tolerated by the cells. To overcome this problem another reporter system was tested. The HIV-1 LTR vector has the same backbone, but instead of firefly luciferase has chloramphenicol actetyl transferase (CAT) as the reporter gene (Fig28). Tat transactivation at the LTR region leads to high expression of CAT activity which can be measured by CAT ELISA.

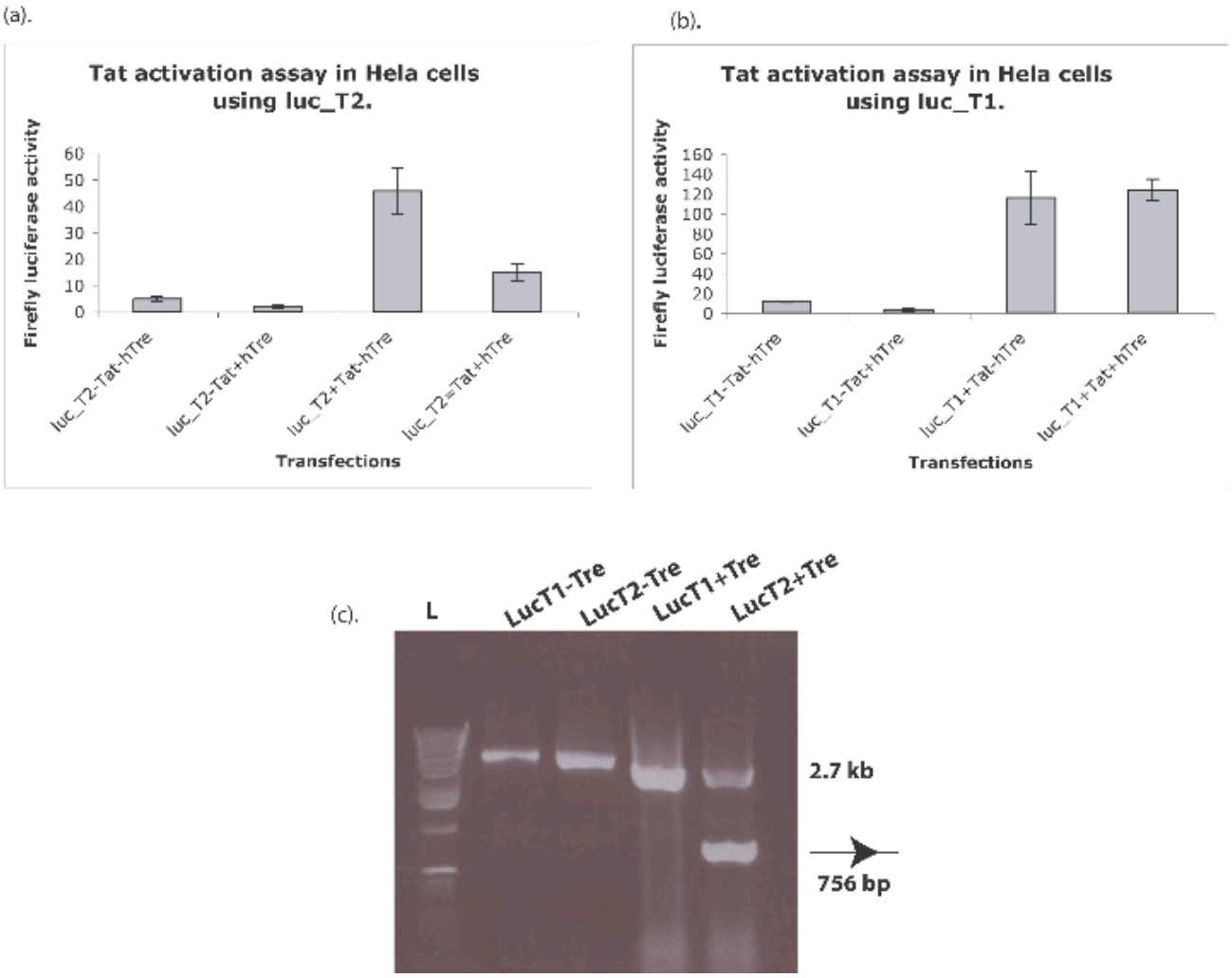


Fig.27. Tat transactivation assay in Hela cells. (a) Hela cells were transfected with HIV-T2-luciferase vector with the loxltr1 cloned in both the LTR regions of the vector, in presence and absence hTre. The firefly luciferase activity was measured from the transfected cell lysate 48 hours post transfection. In presence of hTre the firefly luciferase readout was lower.

(b) Transfection of Hela cells with the control version of the luciferase vector, HIV_T1_LUC with only one loxltr cloned did not show a lower FF luciferase activity in presence of hTre.

(c) PCR detection of recombination in Hela cells transiently transfected with HIV_T2_LUC or HIV_T1_Luc in presence and absence of hTre vectors. The lower band represents the recombined fragment after loss of the luciferase coding region. The lower band is absent in PCR from cells transfected with the HIV_T1_LUC and hTre vectors.

5.21 Transient Tat transactivation assay in Hela cells using the HIV_CAT_T2.

Hela cells were transfected with the HIV-1 LTR based CAT vector along with Tat in presence and absence of the recombinase and CAT Elisa was performed 48 hours post transfection to get the CAT activity of the transfected cells. As with the luciferase vector, recombination resulted in a lower readout of CAT activity (Fig.29), which was confirmed using PCR analysis (data not shown).

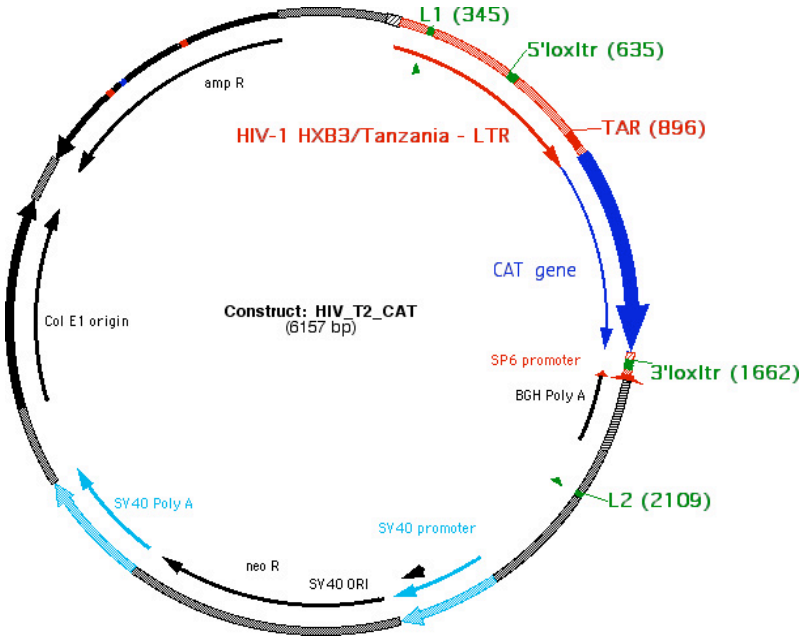


Fig.28. Map of the HIV_T2_CAT vector. The vector has the loxltr sequence cloned within the HIV-1 HXB3 LTR. The 5' LTR sequence is complete with the TAR region where Tat binds and transactivates chloramphenicol acetyl transferase or CAT expression. L1 and L2 denote the primer pair used for PCR detect recombination. Recombination via the loxltr sites will lead to deletion of the CAT coding region.

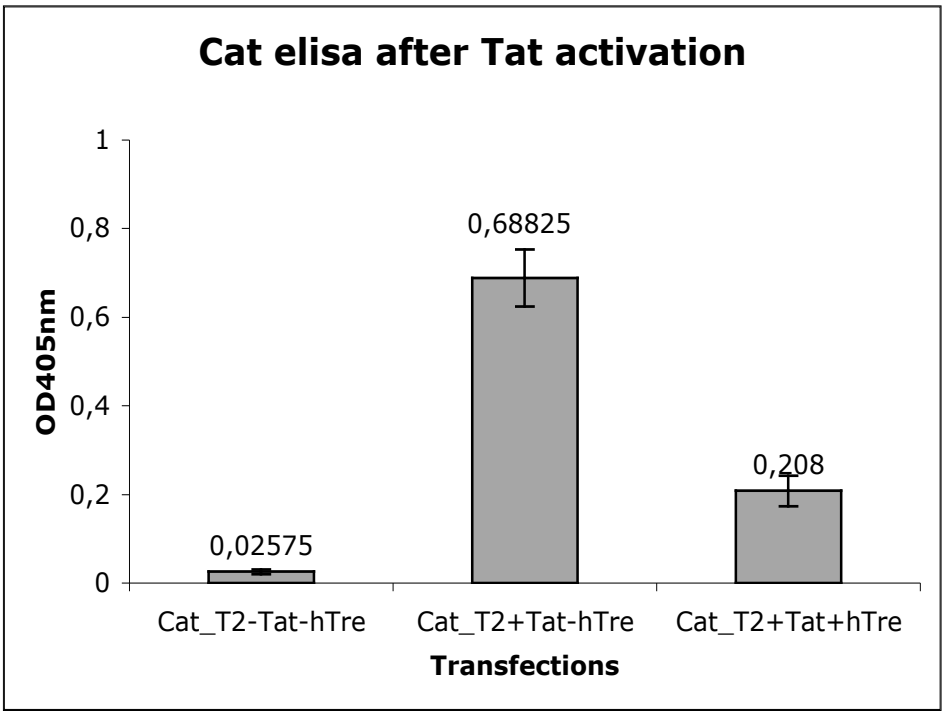


Fig. 29. Transient Tat activation assay in Hela cells transfected with the HIV_T2_CAT vectors in presence and absence of hTre.

The transient Tat transactivation assays worked well and recombination resulted in around three fold reduction in the measured CAT activity, a result consistent with the data from the luciferase assays.

5.22 Recombination mediated by hTre in HIV_T2_CAT stable Hela cell lines.

It was however important to determine the extent of recombination mediated by hTre at the loxltr target stably integrated into the genome. A PCR assay was developed by using primers L1 and L2 in the upstream and downstream LTRs of the HIV_T2-CAT vector that would allow easy and fast estimation of the recombination efficiency. This PCR was used to test for recombination in case of both the luciferase vectors (Fig.27c) and the

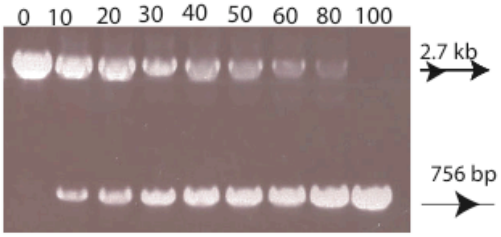
CAT vectors (which have identical plasmid backbones with only different reporter genes) and gave higher and lower band reflecting unrecombined and recombined plasmids.

To obtain the recombined version of the HIV_T2_CAT vector, the plasmid was grown in presence of Tre in *E.coli* DH5α (preparation of the recombined version is described in materials and methods section). The HIVΔCAT clones were screened using PCR with primers L1 and L2.

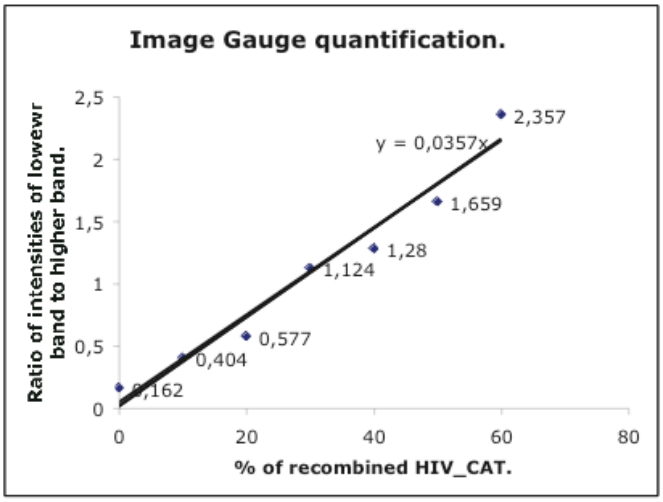
The HIVΔCAT vector was now used in combination with the unrecombined HIV_T2_CAT vector in different ratios to perform PCR with primers L1 and L2 (Fig.29a and b). Decreasing percentage of the unrecombined vector with respect to recombined vector correlated with increasing intensity of the lower 756 bp band for recombination and decreasing intensity of the higher band of 2.7 Kb. The respective band intensities of the upper and lower band for each combination of HIVΔCAT/HIV_T2_CAT was quantified using Image Gauge analysis and the ratio of band intensities with respect to percentage of recombined CAT vector was used to make a standard curve. The transient co-transfection assays with the HIV-1 LTR reporter vectors resulted in a higher recombination rate of around 40-50% (Fig.27c) as can be estimated from the intensity of the lower band.

To quantify the extent of recombination mediated at the loxltr target stably integrated into the genome using PCR analysis, the HIV_T2_CAT vector was transfected into Hela cells and selected with neomycin. Three different stable CAT Hela cell lines were transfected with hTre expression vector or mock vector and genomic DNA was harvested 48 hours post-transfection and analysed by PCR using primers L1 and L2. The stable cell line 7 responded best to the PCR analysis (Fig.29c). The PCR data was quantified using Image Gauge analysis and it was found that the extent of recombination mediated by hTre was to the extent of 9.5% in the Cat stable Hela cell line7 (Fig.29c). The CAT stable cell lines resemble an integrated proviral construct with a full length HIV-1 LTR flanking the reporter gene. In this experimental setup hTre recombined to an extent of around 10% after a single round of transfection.

(a).



(b).



(c).

Fig.29.(a). HIV_T2_CAT and HIVΔCAT vectors were combined in different ratios and PCR was performed using the L1 and L2 primers. The intensity of the lower 756 bp band increases with the increasing percentage of HIVΔCAT plasmid (shown on the gel) compared to the unrecombined counterpart.

(b). The PCR results were quantified using Image Gauge analysis and the data was used to get the trendline.

(c) PCR analysis of recombination in HIV_T2_CAT stable cell lines. Out of the three cell lines, cell lines 5 and 7 were positive for the integrated construct and also for recombination.

6. Discussion

6.1 Exploring evolvability of Cre recombinase.

Evolvability can be linked to the ability to innovate. Proteins have multiple and unforeseeable biochemical activities or biological functions. Numerous terms are used to describe the ability of a single protein to exhibit more than one specificity. Cross-reactivity refers to the activities of a single protein that overlap with the original activity, for instance, with substrates that resemble the original target of the protein.

Directed evolution of Cre recombinase and other recombinases like Flp to generate novel variants with new substrate-specificity, have proven that these proteins are amenable to changes. The Tre recombinases have been generated using an extensive and combinatorial PCR based laboratory evolution strategy traversing six loxltr subsets. Each sub-set was a different target, and it was possible to generate functional recombinases for each of these targets. The idea was to build recombinase libraries for each pallindromic subset with their distinct set of mutations, and recombine them by DNA shuffling to assimilate the beneficial mutations conferring the change in specificity for each subset. The final loxltr specific Tre recombinases were generated by combining two libraries for loxltr1 and loxltr2, which separately recognized and recombined the pallindromic variants derived from loxltr left and right arms. The process facilitated the generation of a single recombinase which is able to recombine the asymmetric loxltr substrate *in vivo* in *E. coli* and also in mammalian cells. This is the first example of tailoring Cre to recombine an asymmetric target on an HIV-1 LTR using directed evolution and bypassing the requirement for rational design.

Recently it has been shown that asymmetric and chimeric lox sites generated by combining the half-sites of loxP and a variant, the loxM7 were recombined in an *in vitro* recombination assay by using an equimolar mixture of wtCre and a Cre variant, CM2 (Saraf-Levy et al. 2006). CM2 was generated using both positive selection for recombination of loxM7 and negative selection against recombination at loxP site. The other variant CM1 used in this study was generated only using positive selection and

displayed relaxed specificity for both loxM7 and loxP, whereas CM2 had almost 40-fold higher efficiency for recombination at loxM7 than at loxP (Santoro and Schultz 2002). Crystal structure analysis of these two variants have revealed the importance of solvent network and water molecule bridging in mediating the specificity of these two recombinase (Baldwin et al. 2003). CM1 was also found to be efficient in mediating recombination at the chimeric asymmetric substrates. Cre or CM2 alone were not able to recombine the asymmetric targets.

Despite the availability of structural information about Cre recombinase, and other site-specific recombinases, it must be stated that *de novo* design of such a loxltr specific recombinase would be quite difficult based on rational design approaches alone, as the sequence space of a 343 amino acid protein like Cre would be enormous (20^{343}). However the information available from the laboratory evolved Cre variants, X-ray crystal structures and mutagenesis studies made on Cre to identify the connectivities in structure and function can lead to a point where slowly rational design and directed evolution can be combined to generate a recombinase variant desired.

6.2. Comparative analysis of mutations in evolved Cre recombinase variants and their target sites.

Structural, biochemical and directed evolution information of Cre variants can be explored to identify the residues important for functional plasticity. The plasticity residues are those that primarily govern substrate specificity and are considered separately from the residues important for catalytic activity. Structural analysis, random mutagenesis and phylogenetic analysis have been used to predict whether a residue is a plasticity residue or not (Matsumura and Ellington 2001; Aharoni et al. 2004; Aharoni et al. 2005; Yoshikuni, Ferrin, and Keasling 2006).

Since the first published crystal structure data of Cre-loxA synapse (Guo, Gopaul, and van Duyne 1997), a lot of information is now available regarding the structural details and the interactions of the recombinase with the target lox site (Van Duyne 2001; Martin et al. 2002; Ennifar et al. 2003). Cre has two distinct domains, the amino terminal domain from amino acid residues 20 to 129, and the larger carboxy terminal domain from residue

132 to 341, separated by a short linker. The N-terminal domain consists of five α-helical segments (A to E), among which the DNA is contacted primarily via the B and D helices. The carboxy terminal domain of Cre consists of a nine-helix domain (helices F-N) packed against a small β-sheet (strands 1to 3). The primarily helical C-terminal domain interacts with the entire 13 base- pair inverted repeat region and the first two bases of the strand exchange or spacer region. To analyse the target recognition specificity of Cre, mutants of Cre with amino acid substitutions in the presumptive DNA binding region in the C-terminal domain were constructed and analysed for their binding to loxP using gel retardation assay (Hartung and Kisters-Woike 1998). The same study revealed that binding of wt Cre was sensitive to mutations in positions 2 and 6 of the inverted repeats of loxP. Band shift assays using symmetrical loxP sites and also symmetrical single base substitutions in positions 2 or 6 of the inverted repeats using Cre mutants revealed that there was broadening of specificity in case of mutants with substitutions in the region surrounding helix M (positions 317-323) and was best illustrated by CreN317A, CreN319C or CreN319P mutants and they could bind both loxP2 and loxP6 variants along with loxP (Hartung and Kisters-Woike 1998).

Interestingly the sequenced recombinases for loxltr2b, loxltr1, loxltr2 and final target loxltr consistently contained the mutations N317T and I320S (tables 2, 3 and 4). Mutations at positions 317 and 320 were also present in seven of the published sequences of Fre recombinases (Buchholz and Stewart 2001). These residues have been implicated in positioning of the loxP spacer sequence for cleavage in the precleaved complex (Guo, Gopaul, and Van Duyne 1999). The presence of the mutations at these positions in recombinases for different target sites suggest that amino acid changes that get fixed in a population or that were identified in independent screens likely represent residues that might indeed be functional in broadening specificity.

Analysis of the Cre crystal structure showed that the base pairs 5 to 7 in inverted repeats of loxP are proximal to residues 258 to 266 in helix J. The central guanine in position 6 of loxP is recognized by Arg 259 in the major groove, where it forms hydrogen bonds to the N7 and O6 of guanine 6. The complementary C nucleotide to this G6 nucleotide is linked by hydrogen bonding and solvent network by Thr258, this network also includes the Glu262 carboxylate group (Martin et al. 2002).

The asymmetric loxltr target has a different distribution of mutations in its right arm and left arm. The G'6 is a C in the right arm of loxltr, and also in loxltr2b, and loxltr2 subsets, derived from the right 13 base pairs sequence, and the G5 is a C5 in the left arm, and also in loxltr1b and loxltr2 derived from the left 13 base pairs (Fig.7 and Fig.10). All the sequenced loxltr2 recombinases and 90% of sequenced loxltr recombinases showed consistently the mutation R259Y. R259C was present in around 40% of the sequenced recombinases for loxltr2b and R259H in 10% of loxltr recombinases. However among the recombinases sequenced for loxltr1a, loxltr1b and loxltr1 the Arg259 was unchanged.

Mutation at position 259 were also present in sequences of evolved Cre recombinase variants which recognized loxM7 target in which the residues 5,6 and 7 in the loxP inverted repeats were mutated (Santoro and Schultz 2002). In this case two recombinases one with relaxed specificity for loxM7 and loxP, and the other with around 40-fold higher activity for loxM7, displayed a R259S mutation. The noticeable feature is the change from basic Arg to a polar residue (Tyr, Ser and Cys). It could be possible that changes in these amino acids from a basic to a polar nature somehow modify the surrounding hydrogen bonding and solvent network to facilitate binding and recognition of the mutated target sites. The individual role of the amino acid mutations cannot be predicted correctly without the crystal structure, or at least modeling of the mutations with respect to available crystal structure of the wild type recombinase, nevertheless it is possible to at least get an idea of the probable plasticity residues which are players in determining the substrate specificity of an enzyme.

The other most hypervariable codon was E262. One of the common features in the target sites used to evolve recombinase variants till now, that is, loxM7, loxK2, loxH and the loxltr substrates have mutations within residues 5 to 8 in the inverted repeat sites (Fig.30).

All sequenced recombinases for loxltr and the loxltr2 sub-set displayed an E262Q mutation and few clones for loxltr1b, loxltr2b and loxltr1 had the glutamate mutated to Ala or Gln or Arg or His. The Fre recombinases also had Glu at 262 mutated to Gln, or Ala or Gly (Buchholz and Stewart 2001). This residue was mutated to Gly in one of the loxM7 recombinases and five Cre mutants recognizing loxK2 sites (Rufer and Sauer 2002; Santoro and Schultz 2002). Systematic replacement of the only Glu at 262 with all

the remaining 19 amino acids revealed that Tyr, Gly, Ala, Gln, Ser or Trp resulted in detectable recombination at the loxK2 substrate, whilst retaining 100% recombination proficiency at the native loxP substrate (Rufer and Sauer 2002). The data from the E262 single mutants show that that this Glu262 is an optimal choice for imposing change in substrate specificity while allowing undiminished recombination at loxP.

Fig.30. Sequence of loxP and mutant lox variants used as target sites in directed evolution experiments till now. The bases in violet are mutated with respect to loxP. The first seven bases in the inverted repeats have been labeled.

6.3. Regions of potential plasticity in the Cre recombinase mutants.

Mapping of mutations in the Tre recombinases and also the loxltr1 and loxltr2 recombinases show a wide distribution of mutations throughout the protein, with a high density of mutations in the extreme N-terminus and on helices D, J and M, and 4-5 β hairpin and L-M linker (fig.19). Such mutation maps can be informative about the regions within the protein which are flexible and can tolerate mutations. In the directed evolution study of Cre to recognize loxH, the resultant Fre recombinases also showed presence of hypervariable codons on helices D and J and M. The N-terminus was again very tolerant to mutations with positions 3 and 7 being very persistent, though numerous mutations near the N- terminus suggest the flexibility of this region of the protein to mutations (Buchholz and Stewart 2001).

Recently pentapeptide insertional mutagenesis of Cre was performed to make a functional map of the recombinase (Petyuk et al. 2004). Pentapeptide scanning mutagenesis is different from directed evolution studies yet both systems allow the probing of regions of flexibility in a protein. The study identified 8 major clusters of active insertions and 6 clusters of inactivating insertions. There were regions of no insertions like helices B, E and J. The loxltr recombinases however showed persistent mutations between positions 259 and 263 on helix J. The B and E helices in our case also showed either no or single mutation. The insertions in the above study concentrated mainly to the extreme N-terminus, and loop between helices J and K. It was found that four non-conserved linker and loop regions do not tolerate mutations, A-B, E-F, 3-I and M-N. The more structured and presumably functionally important regions like the helix A, 4-5 β hairpin, which is evolutionarily flexible and the J-K loop which contact the DNA in the minor groove were tolerant to mutations. In general the inactivating insertions occurred in regions more buried and the non-inactivating insertions were in regions which were more solvent accessible. The evolved Tre and loxltr1 and loxltr2 recombinases showed highly recurssive mutations at positions 317 and 320 in the L-M region, but the A-B, E-F, 3-I and M-N regions have either very low or no mutations showing that these regions allow little to no changes to maintain function and also is in agreement with the data from the pentapeptide mutagenesis study. The extreme N-terminus (including helix A) for the Tre recombinases was also very tolerant to mutations. The outcome of mutation distribution depends on the mutagenesis protocol, but such comparative analysis would give a clearer vision of regions in the recombinases which might be further explored for their function in enzyme catalysis and specificity. In general adaptive mutations that affect substrate specificity are difficult to identify. Preponderance of apparently neutral mutations in molecular evolution (Kimura 1983) makes it difficult to identify adaptive mutations through sequence comparison alone. The relationship between structure and function of proteins are often unclear even when an X-ray crystal structure is available. An important determinant of substrate specificity is the conformational dynamics of a protein, and mutations which affect this property can occur anywhere in the protein (James and Tawfik 2003). However the role of seemingly interesting mutations that get fixed over the course of evolution can be dissected by using site-directed mutagenesis to create

single or multiple mutants and characterize the mutations which are synergistic with respect to catalytic efficiency and specificity.

Based on the sequence information from the loxltr recombinases and other published Cre variants and mutagenesis studies, it could be worthwhile to focus on the regions like those in and around helices D, J, M, the 4-5 β hairpin and L-M linker, which have proved to be flexible to changes, for future directed evolution experiments.

6.4. Modelling of mutations in the selected Tre25 recombinase onto wild type Cre crystal structure.

The major amino acid mutations in the selected Tre recombinase, Tre25.42 (table4) have been modeled onto published Cre crystal structure (PDB entry 1Crx) synapsed to lox site (Guo, Gopaul, and van Duyne 1997). The individual role of the amino acid mutations cannot be predicted correctly without the crystal structure, however modeling of the mutations with respect to available crystal structure of the wild type recombinase could give an idea of how specific mutations might interact with the target site and contribute to recognition of the loxltr target site.

The recognition of a specific nucleotide sequence by a DNA binding protein is determined by the numerous atomic interactions between the amino acids of the protein and nucleotides of the DNA. Systematic analysis of different DNA binding proteins revealed common interactions between specific amino acids and nucleotides (Pabo and Sauer 1992; Suzuki and Yagi 1994; Mandel-Gutfreund, Schueler, and Margalit 1995; Luscombe, Laskowski, and Thornton 2001).

The studies have concluded that though there are some typically favoured interactions between amino acids and DNA bases, there is no simple code that can describe the specific recognition of target sites by proteins (Pabo and Sauer 1984; Matthews 1988).

The synergistic effect of hydrogen bonds, van der Waals contact and water mediated interactions contribute to recognition of specific DNA sequence by a protein.

Inspection of the sequences of the Tre recombinases reveal hypervariable clusters of mutations that were combined during shuffling of loxltr1 and loxltr2 recombinases for directing evolution at loxltr. Among these mutations the residues at positions 259, 262

and 263 all occur in the major groove of DNA. This is important as there are different hydrogen bond donors and acceptors in the major and minor groove of DNA. This cluster of mutations interact with the right arm of the asymmetric loxltr site.

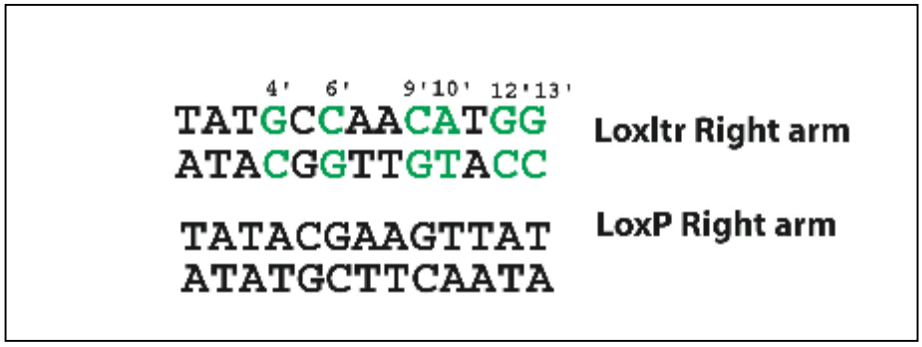


Fig.31. Sequences of the right arm of loxltr and loxP sequences.

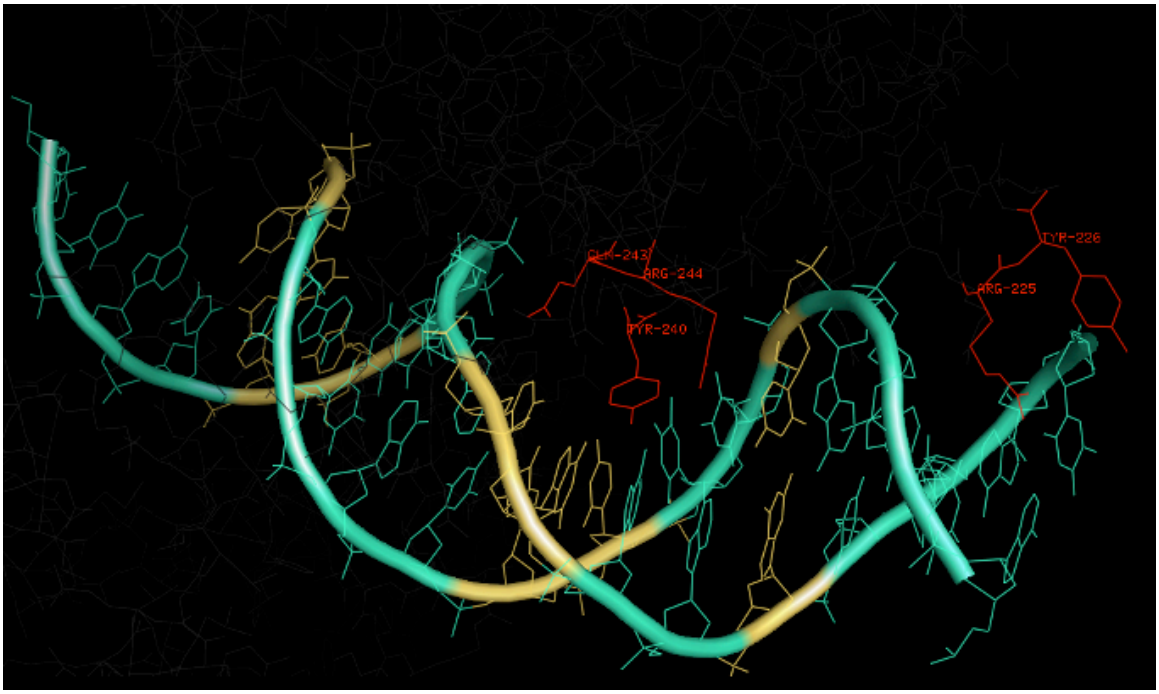


Fig.32. Modelling of amino acid mutations in the Tre25.42 recombinase onto the wild type Cre crystal structure (1Crx). The first 19 amino acids have not been resolved in the crystal structure and so the numbers of the residues modeled are actually: Y240 is actually residueY259, Q243 is Q262, R244 is R263, R225 is R244 and Y226 is Y245. The AT and GC base pairs are shown in green and orange respectively.

R259Y mutation: This residue is in proximity to A7' and also to preceeding G6' that has been mutated to a C in loxltr sequence (Fig.31 and 32). There are several possibilities as to how this tyrosine might interact with the DNA. On one hand this aromatic residue might intercalate with the DNA base pair stacks and lead to a kink in the loxltr, which can change the size and width of the major groove. It is known that aromatic amino acids and also proline and histidine are able to produce extensive ring stacking interactions with DNA. When no base is exposed then the side chains of these residues are positioned with the plane of the ring facing the DNA, which maximizes the contact surface area. Intercalation with adjacent base stacks lead to kinks in DNA as for integration host factor. There could also be hydrogen bonding with the DNA in the following ways:
(a). The phenolic group can directly form hydrogen bond with the two proximal hydrogen bond donors, the N4 of C6' and also with N6 of A7'. In the modeled structure of Tre25.42, the N6 of adenine directly faces the planar ring of tyrosine. The mutation from G6' to C6' would also expose the N4 of cytosine towards the edge of the planar ring (fig.32).
(b). Alternatively a water molecule can be trapped between the DNA and the tyrosine and provide a bridging interaction for hydrogen bond acceptors for N7 and O6 of guanine, N7 of adenine and O4 of thymine.

E262Q mutation: This residue is close to the unchanged cytosine 5'and the mutated G4'. Glutamine is known to show a preference for adenine and also guanine. From the modeled structure the amino group of glutamine would be capable of interacting with the N7 of G4' and the carboxy group is oriented towards the sugar moiety in the backbone of DNA and cannot form hydrogen bonding interactions with the base itself.

G263R mutation: This mutation is predicted to have major impact on the DNA sequence specificity, as a small non-interactor has been replaced by a bulky and positively charged arginine. Arginine is known to make the highest number of hydrogen bonds especially with guanine, and is capable of making bidentate interactions with bases. In a bidentate

bond, arginine interacts with both N7 and O6 atoms of guanine. Arginine is used most frequently in such interactions because of the length of the side-chain, the capacity to interact in different conformations and the ability to produce good hydrogen bonding geometries.

From the modeled structure (Fig.32), the Arg residue is in close contact with the non-mutated A8' and the following cytosine9' in loxltr (guanine in loxP, Fig.31). Arginine is close to the guanine in the complementary strand and could make hydrogen bonds with the N7 and O6 acceptors. The other potential contact is with O4 of thymine in the complementary strand.

The other recurring mutations that are close to the DNA include the K244R and N245Y mutations. Both residues are very close to the last two bases in the loxltr right arm, that is the G12'and G13'.

K244R mutation: In comparison to lysine, arginine provides two hydrogen bond acceptor sites and could make bidentate interactions with the N7 and O6 of the guanine bases.

N245Y mutation: This mutation lies at the end of the loxltr and is difficult to interpret. The side-chain of the amino acid is in close proximity to the backbone of DNA, rather than with any base. An intercalation is suggested by the orientation of the residue, though it could also involve a hydrogen bond network with the DNA. In general, DNA backbone interactions are important for stability rather than specificity.

Modelling of the mutations in the Tre25.42 recombinase on the wild type Cre crystal structure suggests that recognition of the asymmetric site come from the interactions of the mutations on one side of the binding surface of the protein which interacts with the right arm, the mutations in the other half are not in proximity to the DNA and their contribution are relatively difficult to interpret.

6.5. Relaxed specificity of evolved recombinases.

An important feature of any directed evolution experiment is the specificity of the generated enzyme variant. It is known that directed evolution tends to produce enzyme variants with broadened and relaxed substrate specificity (Matsumura and Ellington 2001; Aharoni et al. 2005). As has been discussed in section 4.9, the evolved recombinases for loxltr1 and loxltr2 analysed, did not display cross-reactivity on the other target although they did recombine loxP (Fig.16). This could be explained by the fact that the loxltr1 and loxltr2 were different sites. The Tre recombinases recombining loxltr which consists of all the mutations of all the subsets, could be expected to recognize and recombine the six loxltr sub-sets, however while both the Tre15 and Tre25 recombinases retained their recombinational proficiency at loxP it was interesting to note that Tre 15, had much higher cross-reactivity on the loxltr subsets, but after 10 more cycles of screening and selection, the Tre25 recombinase recombined only loxltr2b. This shows that evolution proceeded to make the recombinase more specific and less promiscuous. Transitions of a protein from a specialized enzyme into a generalized enzyme and ultimately a new 'respecialised' enzyme is a feature of molecular evolution.

The fact that evolved recombinases retain their native activity at loxP was also shown for the previously evolved Cre variants for loxH, loxK2 and loxM7. The observation that substantial changes in the specificity of a protein need not come at the expense of its native function has also been made in many other directed evolution experiments (Kumamaru et al. 1998; Wan et al. 1998; Glieder, Farinas, and Arnold 2002; Xia et al. 2002; Aharoni et al. 2005). An evolved protein can acquire increased fitness for a new function without losing its original function.

Tre recombines loxP to an extent of 50% at an induction when it recombines loxltr to about 80% (Fig.18). Increasing the evolution cycles would have likely increased further the activity at loxltr, but the decrease in the activity at loxP is not predictable, as there was no counterselection of recombinases against recombination at loxP. However, it is an interesting possibility that the evolved recombinases would lose their activity on loxP

even without counterselection by continuing with evolution cycles based on positive selection alone.

In an extensive literature survey on directed evolution of proteins made by Aharoni et al., it was shown that in comparison to the wild type counterpart, the original or native activity of an evolved protein decreased by a factor of around 3.2 on average, whereas the increase in novel activity was many times more. Taking Fre17 as an example, a recombinase generated after 20 evolution cycles (Buchholz and Stewart 2001), they have shown that there was 1.15 fold change in the native activity or recombination at loxP in comparison to wt Cre, whereas the fold change relative to wt Cre at loxH was 220 times more. These results were from a comparison based on the choice of evolved recombinases for integration at either loxP or loxH. Fre 17 was evolved only in presence of positive selection pressure, whereas, another variant Fre22, which underwent counterselection, had lost its activity on loxP while retaining 80 fold more activity on loxH compared to wt Cre.

When Fre17 and Fre22 were compared on the basis of deletion at either loxP or loxH, it was found that Fre17 recombined loxP with around 46 fold more efficiency than Fre22. The more specific Fre22 was around two fold more efficient at recombining loxH in comparison to Fre17. This is in agreement with the fact that counterselection does result in a respecialised enzyme.

The Tre recombinases were generated using positive selection pressure for recombining loxltr. To increase the speed at which recombination activity on loxP is lost one could include counterselection in future evolution experiments to decrease the native activity. Counterselection may be implemented by use of a pEVO10 like vector with intertwined loxP and loxltr sites as had been used in the evolution of Fre recombinases (Buchholz and Stewart 2001). The final loxltr recombinase library can then be enriched for more specificity at the new target in comparison to the original target.

6.6. Activity of codon-optimised Tre recombinase (hTre) in mammalian cells.

The transient and stable β-galactosidase based reporter assays in Hela cells revealed that hTre is active in Hela cells and also in the genomic context when the reporter construct was integrated into the genome (Fig.24 and 25).

The X-Gal staining of Hela cells and measurement of β-galactosidase activity after co-transfection with the recombinase and reporter vectors revealed that hTre recombined loxltr and loxP in mammalian cells. But it was less efficient at recombining loxP than loxltr, a result consistent with the finding in *E.coli*. According to the results of transient β-galactosidase activity in Hela cells, wtCre was around 9 folds more active in recombining loxP in comparison to hTre, which was around 95 folds more active in recombining loxltr.

Recombination at loxltr target integrated into the genome was proven by PCR, measurement of β-galactosidase activity and staining of loxltr stable cell line. As for the unoptimised Tre, transfection of hTre recombinase vector into the stable cell lines followed by Southern did not work.

6.7. Evolved recombinase and application in anti-retroviral therapy.

We used HIV-1 LTR based reporter vectors (HIV_T2_LUC and HIV_T2_CAT) to address the question whether Tre can recombine loxltr in the context of a real LTR and repress Tat-mediated transactivation of the reporter genes. Such HIV-1 LTR vectors are routinely used to assay for inhibitors of Tat mediated transcription activation of reporter constructs.

The transient Tat activation assay in Hela cells showed that recombination at the loxltr target does decrease the reporter gene activity by almost three fold (Fig.27 and 29). PCR and sequencing gave proof that the repression in reporter gene activity was due to site-specific recombination. The luciferase vector with only one loxltr sequence (HIV_T1_Luc) did not show decrease of luciferase activity in presence of Tre. The result of transient Tat activation assay using the CAT reporter gave data consistent with the luciferase assays.

Further proof that our evolved recombinase can recombine loxltr in the context of a true LTR integrated into the genome came from the experiments in the CAT Hela stable cell lines. Tre recombined the CAT Hela stable cell lines, and recombination was detectable by PCR (Fig.29). The same PCR was used to quantify extent of recombination at the loxltr target in the genome. The results from the PCR of DNA isolated from CAT Hela 7 cell line after transfection with hTre, were compared to the PCR performed with different ratios of HIV_CAT and recombined HIVΔCAT vector. The results indicated that hTre recombined loxltr integrated into the genome to an extent of 10%. If Cre mediated recombination at loxP target integrated in the genome represents 100% recombination, then hTre has one-tenth of recombination efficiency at loxltr compared to Cre at loxP.
To obtain a more active recombinase additional evolution cycles would be necessary. The activity of the evolved hTre recombinase would not be sufficient for efficient removal of an integrated provirus from the genome, however, the study provides proof of concept that evolution can be used for engineering Cre for therapeutic site- specific recombination based anti -retroviral therapy.

6.8. From proof of principle to application- an outlook on directed evolution for genetic surgery.

The riches of protein sequence space can provide answers to complex questions. Laboratory evolution is not quick work, but in the recent years researchers have devised numerous assays to accelerate this process and maximize the power of evolution that will permit the generation of novel proteins in a relatively short time-scale.
Zinc-finger proteins represent a great example of how protein engineering and directed evolution can be used for therapeutic purposes. Zinc fingers can be linked in tandem to recognize DNA sequences of different lengths with high fidelity. Such modular design provides a large number of possibilities for specific recognition of DNA. Over the last few years zinc fingers have been engineered to target longer DNA sequences. By adding functional groups to the engineered DNA binding domains silencing or activation of gene expression can be achieved. Some therapeutic applications of engineered zinc finger proteins which are being developed include gene correction by homologous recombination triggered by introducing double stranded breaks using zinc finger

nucleases for treating monogenic disorders like SCID (Urnov et al. 2005), disruption of HSV infective cycle (Papworth et al. 2003), activation of VEGF in a human cell line and animal model (Rebar et al. 2002), inhibition of HIV-1 gene expression (Reynolds et al. 2003).

We have used directed evolution to engineer Cre to recombine an HIV-1 LTR sequence. Such strategies represents a novel anti-retroviral approach which can be used to target not only HIV but several other retroviruses associated with an array of different malignancies, immunodeficiencies and neurological disorders that afflict a large number of different creatures ranging from fish (WDSV), birds (ALV, RAV, REV) and mammals (MMTV, MLV, BLV, SIV, FIV, HTLV, HIV).

Retroviruses are characterized by their ability to integrate into the host genome and persist indefinitely as a provirus flanked on both sides by viral long terminal repeats which direct the expression of viral genes. Current treatment regimens to combat retroviral diseases in humans, chiefly ATLL (adult T-cell leukemia/ lymphoma) inflicted by HTLV-1 and AIDS caused by HIV-1 are chemotherapeutic and based primarily on enzyme inhibitors. ATLL therapies include nucleoside analogues inhibiting adenosine deaminase, topoisomerase inhibitors and interferons. Treatment of HIV-1 target three steps in the replication cycle of the virus (reverse transcriptase, protease and fusion). Recently clinical efficacy has been demonstrated for new targets like the HIV-1 co-receptors, gp120, integrase and virion maturation. These developments have encouraged the pursuit of novel inhibition strategies and non- traditional targets (Chiu et al. 2005; Hauber et al. 2005). However till date none of these therapies actually remove the provirus from the host genome which makes eradication of the virus impossible and leads to gradual development of viral resistance. Excision of the proviral DNA from infected cells could be achieved by applying the principle of site-specific recombination mediated through sequences in the flanking LTR repeats of the pro-viral genome. Our study shows that evolved recombinases like Tre can be generated by laboratory evolution and hence provides the first example of a strategy targeting the proviral form of a retrovirus.

Realisation of such a recombination mediated genetic surgery to at least pre-clinical phases can be made possible by taking into consideration three issues: (a) by improving the evolution approach to generate more active and specific recombinases for the most

conserved sequence within a retroviral LTR, (b) to prove that evolved recombinases can be functional in the more physiologically relevant cell types and (c) to use the best delivery vehicle to introduce the recombinase,

One of the most important aspect of any directed evolution experiment is the diversity of the staring mutagenised library. *In vitro* methods for creating genetic diversity are very powerful, yet a laborious process when needed to be applied iteratively. Though evolution in bacteria has advantages like relatively short time scale of individual cycle and technical simplicity, yet the process requires generation of a huge library of mutant genes, transfection into cells, screening for improved phenotype, amplification, recovery and sequencing of the best performers.

A big breakthrough in enhancing diversity was made around two years back when a group of researchers used somatic hypermutation to evolve new variants of RFP which showed a shift in fluorescence toward the infrared (Wang et al. 2004). Somatic hypermutation (SHM) is a process by which B cells specifically mutate immunoglobulins when activated by antigens. SHM uses activation-induced cytidine deaminase and error prone DNA repair to introduce point mutations into the rearranged V regions of Ig at a rate 10^6 times higher than the rest of the genome. SHM has been used to repair premature stop codons in non Ig genes. Tsien's group transfected B cells with monomeric RFP and generated evolved mRFP mutants with far-red emissions and enhanced photostability through a process of iterative SHM in a FACs based assay. SHM mediated protein evolution in B cells does not require the much more laborious *in vitro* mutagenesis and screening procedures and allows much more sequence space to be searched. This tool should not be limited to fluorescent proteins, and should be used for the evolution of new enzymes. It would be interesting to develop an SHM based evolution protocol for evolving novel HIV ltr specific recombinases.

Enzymes engineered through laboratory evolution often demonstrate lower catalytic rates and efficiency in comparison to their wild type counterparts. This can be improved by implementing selection and counterselection assay schemes. Recently directed evolution of recombinases have capitalized on counterselection to increase the novel specificity of enzymes and concomitantly decreasing their native or original specificity (Buchholz and Stewart 2001). It is thus very important to include negative selection in the evolution

strategy, and cycle the interesting candidates or the final libraries through such rounds of counterselection.

Tre was evolved to recognize a loxltr sequence with 50% sequence similarity to loxP. A comparison of loxltr to sequences within 30 other HIV LTRs was made and it was found that the loxltr had a sequence similarity ranging from 70 to 90%. The generated loxltr specific libraries can be used as the starting library to direct evolution for other loxltr like sequences.

Finally it is very important to be able to deliver the therapeutic proteins to the right destination. Controlled delivery of Cre recombinase to specific target tissues have been made by using recombinant viral vectors. E1- deleted first generation adenoviral Cre vectors have been used to mediate recombination in tissue culture (Anton and Graham 1995); (Kanegae et al. 1996), in mice for gene targeting in the liver (Akagi et al. 1997); (Chang et al. 1999a), in the brain (Wang, Krushel, and Edelman 1996) in tumors (Sato et al. 1998), and in prostate (Leow, Wang, and Gao 2005). These studies demonstrated very efficient in vivo delivery of the Cre and can be further extended to other evolved recombinases as well.

Recently protein transduction have been developed as an efficient method to deliver biologically active proteins directly into mammalian cells (Schwarze, Hruska, and Dowdy 2000). The 11 amino acid basic peptide fragment from HIV-TAT has been used to deliver β-galactosidase and the enzyme activity was detected in liver, kidney, brain and spleen of mice after i.p. injection (Schwarze et al. 1999). Since then a wide variety of full length Tat fusion proteins ranging from 15 to 121 kDa in size and spanning diverse functional classes have been effectively delivered into cells with biological activity. A few examples include TAT-p16, TATp27 (Nagahara et al. 1998), TAT-Cre (Peitz et al. 2002); (Joshi, Hashimoto, and Koni 2002), TAT-HIV protease (Vocero-Akbani et al. 1999). Both primary and transformed cell types including peripheral blood lymphocytes, fibroblasts, keratinocytes, bone marrow stem cells, osteoclasts, T cells have been transduced with TAT-proteins. The TAT-Cre fusion protein transduced very efficiently in fibroblasts, murine embryonic stem cells and primary splenocytes. High transduction would ensure also higher rate of recombination and can be used to increase recombination efficiency for evolved recombinases like Tre. One important advantage of

using protein transduction to deliver recombinases in comparison to viral delivery is that viral vectors are often found to be toxic due to expression of viral genes in the vector resulting in triggering of host immune response and loss of transduced cells. Also in spite of the incorporation of tissue-specific promoters in first generation vectors, there is loss of specificity and not all tissues are well infected. Thus by combining all the existing tools in hand, it would be possible to engineer solutions to problems in human therapeutics by using evolved proteins with desired characteristics.

7. References

1. Abremski, K., B. Frommer, A. Wierzbicki, and R. H. Hoess. 1988. Properties of a mutant Cre protein that alters the topological linkage of recombination products. J Mol Biol **202**:59-66.
2. Abremski, K., and R. Hoess. 1985. Phage P1 Cre-loxP site-specific recombination. Effects of DNA supercoiling on catenation and knotting of recombinant products. J Mol Biol **184**:211-220.
3. Abremski, K., and R. Hoess. 1984. Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. J Biol Chem **259**:1509-1514.
4. Abremski, K. E., and R. H. Hoess. 1992. Evidence for a second conserved arginine residue in the integrase family of recombination proteins. Protein Eng **5**:87-91.
5. Aharoni, A., L. Gaidukov, O. Khersonsky, Q. G. S. Mc, C. Roodveldt, and D. S. Tawfik. 2005. The 'evolvability' of promiscuous protein functions. Nat Genet **37**:73-76.
6. Aharoni, A., L. Gaidukov, S. Yagur, L. Toker, I. Silman, and D. S. Tawfik. 2004. Directed evolution of mammalian paraoxonases PON1 and PON3 for bacterial expression and catalytic specialization. Proc Natl Acad Sci U S A **101**:482-487.
7. Akagi, K., V. Sandig, M. Vooijs, M. Van der Valk, M. Giovannini, M. Strauss, and A. Berns. 1997. Cre-mediated somatic site-specific recombination in mice. Nucleic Acids Res **25**:1766-1773.
8. Albert, H., E. C. Dale, E. Lee, and D. W. Ow. 1995. Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. Plant J **7**:649-659.
9. Anton, M., and F. L. Graham. 1995. Site-specific recombination mediated by an adenovirus vector expressing the Cre recombinase protein: a molecular switch for control of gene expression. J Virol **69**:4600-4606.
10. Araki, K., M. Araki, and K. Yamamura. 2002. Site-directed integration of the cre gene mediated by Cre recombinase using a combination of mutant lox sites. Nucleic Acids Res **30**:e103.
11. Argos, P., A. Landy, K. Abremski, J. B. Egan, E. Haggard-Ljungquist, R. H. Hoess, M. L. Kahn, B. Kalionis, S. V. Narayana, L. S. Pierson, 3rd, and et al. 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. Embo J **5**:433-440.
12. Arin, M. J., and D. R. Roop. 2004. Inducible mouse models for inherited skin diseases: implications for skin gene therapy. Cells Tissues Organs **177**:160-168.
13. Awatramani, R., P. Soriano, C. Rodriguez, J. J. Mai, and S. M. Dymecki. 2003. Cryptic boundaries in roof plate and choroid plexus identified by intersectional gene activation. Nat Genet **35**:70-75.
14. Baldwin, E. P., S. S. Martin, J. Abel, K. A. Gelato, H. Kim, P. G. Schultz, and S. W. Santoro. 2003. A specificity switch in selected cre recombinase variants is mediated by macromolecular plasticity and water. Chem Biol **10**:1085-1094.
15. Blakely, G., G. May, R. McCulloch, L. K. Arciszewska, M. Burke, S. T. Lovett, and D. J. Sherratt. 1993. Two related recombinases are required for site-specific recombination at dif and cer in E. coli K12. Cell **75**:351-361.

16. Boder, E. T., K. S. Midelfort, and K. D. Wittrup. 2000. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. Proc Natl Acad Sci U S A **97**:10701-10705.
17. Bouhassira, E. E., K. Westerman, and P. Leboulch. 1997. Transcriptional behavior of LCR enhancer elements integrated at the same chromosomal locus by recombinase-mediated cassette exchange. Blood **90**:3332-3344.
18. Buchholz, F., P. O. Angrand, and A. F. Stewart. 1998. Improved properties of FLP recombinase evolved by cycling mutagenesis. Nat Biotechnol **16**:657-662.
19. Buchholz, F., Y. Refaeli, A. Trumpp, and J. M. Bishop. 2000. Inducible chromosomal translocation of AML1 and ETO genes through Cre/loxP-mediated recombination in the mouse. EMBO Rep **1**:133-139.
20. Buchholz, F., and A. F. Stewart. 2001. Alteration of Cre recombinase site specificity by substrate-linked protein evolution. Nat Biotechnol **19**:1047-1052.
21. Cao, T., M. A. Longley, X. J. Wang, and D. R. Roop. 2001. An inducible mouse model for epidermolysis bullosa simplex: implications for gene therapy. J Cell Biol **152**:651-656.
22. Chang, B. H., W. Liao, L. Li, M. Nakamuta, D. Mack, and L. Chan. 1999a. Liver-specific inactivation of the abetalipoproteinemia gene completely abrogates very low density lipoprotein/low density lipoprotein production in a viable conditional knockout mouse. J Biol Chem **274**:6051-6055.
23. Chang, C. C., T. T. Chen, B. W. Cox, G. N. Dawes, W. P. Stemmer, J. Punnonen, and P. A. Patten. 1999b. Evolution of a cytokine using DNA family shuffling. Nat Biotechnol **17**:793-797.
24. Chen, Y., U. Narendra, L. E. Iype, M. M. Cox, and P. A. Rice. 2000. Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. Mol Cell **6**:885-897.
25. Cheng, Q., B. M. Swalla, M. Beck, R. Alcaraz, Jr., R. I. Gumport, and J. F. Gardner. 2000. Specificity determinants for bacteriophage Hong Kong 022 integrase: analysis of mutants with relaxed core-binding specificities. Mol Microbiol **36**:424-436.
26. Chiu, Y. L., V. B. Soros, J. F. Kreisberg, K. Stopak, W. Yonemoto, and W. C. Greene. 2005. Cellular APOBEC3G restricts HIV-1 infection in resting CD4+ T cells. Nature **435**:108-114.
27. Coco, W. M., L. P. Encell, W. E. Levinson, M. J. Crist, A. K. Loomis, L. L. Licato, J. J. Arensdorf, N. Sica, P. T. Pienkos, and D. J. Monticello. 2002. Growth factor engineering by degenerate homoduplex gene family recombination. Nat Biotechnol **20**:1246-1250.
28. Collins, E. C., R. Pannell, E. M. Simpson, A. Forster, and T. H. Rabbitts. 2000. Inter-chromosomal recombination of Mll and Af9 genes mediated by cre-loxP in mouse development. EMBO Rep **1**:127-132.
29. Colloms, S. D., R. McCulloch, K. Grant, L. Neilson, and D. J. Sherratt. 1996. Xer-mediated site-specific recombination in vitro. Embo J **15**:1172-1181.

30. Colloms, S. D., P. Sykora, G. Szatmari, and D. J. Sherratt. 1990. Recombination at ColE1 cer requires the Escherichia coli xerC gene product, a member of the lambda integrase family of site-specific recombinases. J Bacteriol **172**:6973-6980.
31. Crameri, A., S. A. Raillard, E. Bermudez, and W. P. Stemmer. 1998. DNA shuffling of a family of genes from diverse species accelerates directed evolution. Nature **391**:288-291.
32. Doi, N., and H. Yanagawa. 1999. STABLE: protein-DNA fusion system for screening of combinatorial protein libraries in vitro. FEBS Lett **457**:227-230.
33. Ennifar, E., J. E. Meyer, F. Buchholz, A. F. Stewart, and D. Suck. 2003. Crystal structure of a wild-type Cre recombinase-loxP synapse reveals a novel spacer conformation suggesting an alternative mechanism for DNA cleavage activation. Nucleic Acids Res **31**:5449-5460.
34. Esposito, D., and J. J. Scocca. 1997. The integrase family of tyrosine recombinases: evolution of a conserved active site domain. Nucleic Acids Res **25**:3605-3614.
35. Farley, F. W., P. Soriano, L. S. Steffen, and S. M. Dymecki. 2000. Widespread recombinase expression using FLPeR (flipper) mice. Genesis **28**:106-110.
36. Feil, R., J. Brocard, B. Mascrez, M. LeMeur, D. Metzger, and P. Chambon. 1996. Ligand-activated site-specific recombination in mice. Proc Natl Acad Sci U S A **93**:10887-10890.
37. Feng, Y. Q., M. C. Lorincz, S. Fiering, J. M. Greally, and E. E. Bouhassira. 2001. Position effects are influenced by the orientation of a transgene with respect to flanking chromatin. Mol Cell Biol **21**:298-309.
38. Feng, Y. Q., J. Seibler, R. Alami, A. Eisen, K. A. Westerman, P. Leboulch, S. Fiering, and E. E. Bouhassira. 1999. Site-specific chromosomal integration in mammalian cells: highly efficient CRE recombinase-mediated cassette exchange. J Mol Biol **292**:779-785.
39. Flowers, C. C., C. Woffendin, J. Petryniak, S. Yang, and G. J. Nabel. 1997. Inhibition of recombinant human immunodeficiency virus type 1 replication by a site-specific recombinase. J Virol **71**:2685-2692.
40. Ghosh, K., C. K. Lau, F. Guo, A. M. Segall, and G. D. Van Duyne. 2005. Peptide trapping of the Holliday junction intermediate in Cre-loxP site-specific recombination. J Biol Chem **280**:8290-8299.
41. Glieder, A., E. T. Farinas, and F. H. Arnold. 2002. Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. Nat Biotechnol **20**:1135-1139.
42. Gopaul, D. N., F. Guo, and G. D. Van Duyne. 1998. Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination. Embo J **17**:4175-4187.
43. Gossen, M., and H. Bujard. 1992. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc Natl Acad Sci U S A **89**:5547-5551.
44. Griffiths, A. D., and D. S. Tawfik. 2003. Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. Embo J **22**:24-35.

45. Gu, H., J. D. Marth, P. C. Orban, H. Mossmann, and K. Rajewsky. 1994. Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. Science **265**:103-106.
46. Guo, F., D. N. Gopaul, and G. D. Van Duyne. 1999. Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. Proc Natl Acad Sci U S A **96**:7143-7148.
47. Guo, F., D. N. Gopaul, and G. D. van Duyne. 1997. Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. Nature **389**:40-46.
48. Guzman, L. M., D. Belin, M. J. Carson, and J. Beckwith. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol **177**:4121-4130.
49. Hanes, J., C. Schaffitzel, A. Knappik, and A. Pluckthun. 2000. Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. Nat Biotechnol **18**:1287-1292.
50. Hartung, M., and B. Kisters-Woike. 1998. Cre mutants with altered DNA binding properties. J Biol Chem **273**:22884-22891.
51. Hauber, I., D. Bevec, J. Heukeshoven, F. Kratzer, F. Horn, A. Choidas, T. Harrer, and J. Hauber. 2005. Identification of cellular deoxyhypusine synthase as a novel target for antiretroviral therapy. J Clin Invest **115**:76-85.
52. Herrera, P. L. 2000. Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages. Development **127**:2317-2322.
53. Hiraga, K., and F. H. Arnold. 2003. General method for sequence-independent site-directed chimeragenesis. J Mol Biol **330**:287-296.
54. Hoess, R. H., and K. Abremski. 1985. Mechanism of strand cleavage and exchange in the Cre-lox site-specific recombination system. J Mol Biol **181**:351-362.
55. Hoess, R. H., A. Wierzbicki, and K. Abremski. 1986. The role of the loxP spacer region in P1 site-specific recombination. Nucleic Acids Res **14**:2287-2300.
56. Hoess, R. H., M. Ziese, and N. Sternberg. 1982. P1 site-specific recombination: nucleotide sequence of the recombining sites. Proc Natl Acad Sci U S A **79**:3398-3402.
57. Holzenberger, M., C. Lenzner, P. Leneuve, R. Zaoui, G. Hamard, S. Vaulont, and Y. L. Bouc. 2000a. Cre-mediated germline mosaicism: a method allowing rapid generation of several alleles of a target gene. Nucleic Acids Res **28**:E92.
58. Holzenberger, M., R. Zaoui, P. Leneuve, G. Hamard, and Y. Le Bouc. 2000b. Ubiquitous postnatal LoxP recombination using a doxycycline auto-inducible Cre transgene (DAI-Cre). Genesis **26**:157-159.
59. Jacob, J., and D. Baltimore. 1999. Modelling T-cell memory by genetic marking of memory T cells in vivo. Nature **399**:593-597.
60. James, L. C., and D. S. Tawfik. 2003. Conformational diversity and protein evolution--a 60-year-old hypothesis revisited. Trends Biochem Sci **28**:361-368.

61. Joseph, N. M., Y. S. Mukouyama, J. T. Mosher, M. Jaegle, S. A. Crone, E. L. Dormand, K. F. Lee, D. Meijer, D. J. Anderson, and S. J. Morrison. 2004. Neural crest stem cells undergo multilineage differentiation in developing peripheral nerves to generate endoneurial fibroblasts in addition to Schwann cells. Development **131**:5599-5612.
62. Joshi, S. K., K. Hashimoto, and P. A. Koni. 2002. Induced DNA recombination by Cre recombinase protein transduction. Genesis **33**:48-54.
63. Kanegae, Y., K. Takamori, Y. Sato, G. Lee, M. Nakai, and I. Saito. 1996. Efficient gene activation system on mammalian cell chromosomes using recombinant adenovirus producing Cre recombinase. Gene **181**:207-212.
64. Kellendonk, C., F. Tronche, A. P. Monaghan, P. O. Angrand, F. Stewart, and G. Schutz. 1996. Regulation of Cre recombinase activity by the synthetic steroid RU 486. Nucleic Acids Res **24**:1404-1411.
65. Kimura, M. 1983. Rare variant alleles in the light of the neutral theory. Mol Biol Evol **1**:84-93.
66. Kolb, A. F. 2001. Selection-marker-free modification of the murine beta-casein gene using a lox2272 [correction of lox2722] site. Anal Biochem **290**:260-271.
67. Kolkman, J. A., and W. P. Stemmer. 2001. Directed evolution of proteins by exon shuffling. Nat Biotechnol **19**:423-428.
68. Kuhn, R., F. Schwenk, M. Aguet, and K. Rajewsky. 1995. Inducible gene targeting in mice. Science **269**:1427-1429.
69. Kumamaru, T., H. Suenaga, M. Mitsuoka, T. Watanabe, and K. Furukawa. 1998. Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase. Nat Biotechnol **16**:663-666.
70. Kumar, S., C. S. Chen, D. J. Waxman, and J. R. Halpert. 2005. Directed evolution of mammalian cytochrome P450 2B1: mutations outside of the active site enhance the metabolism of several substrates, including the anticancer prodrugs cyclophosphamide and ifosfamide. J Biol Chem **280**:19569-19575.
71. Kwon, H. J., R. Tirumalai, A. Landy, and T. Ellenberger. 1997. Flexibility in DNA recombination: structure of the lambda integrase catalytic core. Science **276**:126-131.
72. Lakso, M., B. Sauer, B. Mosinger, Jr., E. J. Lee, R. W. Manning, S. H. Yu, K. L. Mulder, and H. Westphal. 1992. Targeted oncogene activation by site-specific recombination in transgenic mice. Proc Natl Acad Sci U S A **89**:6232-6236.
73. Landy, A. 1989. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. Annu Rev Biochem **58**:913-949.
74. Landy, A. 1993. Mechanistic and structural complexity in the site-specific recombination pathways of Int and FLP. Curr Opin Genet Dev **3**:699-707.
75. Lee, G., and I. Saito. 1998. Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. Gene **216**:55-65.

76. Lee, L., and P. D. Sadowski. 2001. Directional resolution of synthetic holliday structures by the Cre recombinase. J Biol Chem **276**:31092-31098.
77. Lee, Y., and J. Park. 1998. A novel mutant loxP containing part of long terminal repeat of HIV-1 in spacer region: presentation of possible target site for antiviral strategy using site-specific recombinase. Biochem Biophys Res Commun **253**:588-593.
78. Leow, C. C., X. D. Wang, and W. Q. Gao. 2005. Novel method of generating prostate-specific Cre-LoxP gene switching via intraductal delivery of adenovirus. Prostate **65**:1-9.
79. Liu, P., N. A. Jenkins, and N. G. Copeland. 2003. A highly efficient recombineering-based method for generating conditional knockout mutations. Genome Res **13**:476-484.
80. Liu, P., N. A. Jenkins, and N. G. Copeland. 2002. Efficient Cre-loxP-induced mitotic recombination in mouse embryonic stem cells. Nat Genet **30**:66-72.
81. Lobe, C. G., K. E. Koop, W. Kreppner, H. Lomeli, M. Gertsenstein, and A. Nagy. 1999. Z/AP, a double reporter for cre-mediated recombination. Dev Biol **208**:281-292.
82. Logie, C., and A. F. Stewart. 1995. Ligand-regulated site-specific recombination. Proc Natl Acad Sci U S A **92**:5940-5944.
83. Luscombe, N. M., R. A. Laskowski, and J. M. Thornton. 2001. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucleic Acids Res **29**:2860-2874.
84. Mahan, S. D., G. C. Ireton, C. Knoeber, B. L. Stoddard, and M. E. Black. 2004. Random mutagenesis and selection of Escherichia coli cytosine deaminase for cancer gene therapy. Protein Eng Des Sel **17**:625-633.
85. Mandel-Gutfreund, Y., O. Schueler, and H. Margalit. 1995. Comprehensive analysis of hydrogen bonds in regulatory protein DNA-complexes: in search of common principles. J Mol Biol **253**:370-382.
86. Mao, X., Y. Fujiwara, A. Chapdelaine, H. Yang, and S. H. Orkin. 2001. Activation of EGFP expression by Cre-mediated excision in a new ROSA26 reporter mouse strain. Blood **97**:324-326.
87. Mao, X., Y. Fujiwara, and S. H. Orkin. 1999. Improved reporter strain for monitoring Cre recombinase-mediated DNA excisions in mice. Proc Natl Acad Sci U S A **96**:5037-5042.
88. Martin, S. S., E. Pulido, V. C. Chu, T. S. Lechner, and E. P. Baldwin. 2002. The order of strand exchanges in Cre-LoxP recombination and its basis suggested by the crystal structure of a Cre-LoxP Holliday junction complex. J Mol Biol **319**:107-127.
89. Matsumura, I., and A. D. Ellington. 2001. In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. J Mol Biol **305**:331-339.
90. Matthews, B. W. 1988. Protein-DNA interaction. No code for recognition. Nature **335**:294-295.
91. McCulloch, R., L. W. Coggins, S. D. Colloms, and D. J. Sherratt. 1994. Xer-mediated site-specific recombination at cer generates Holliday junctions in vivo. Embo J **13**:1844-1855.

92. Meyers, E. N., M. Lewandoski, and G. R. Martin. 1998. An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination. Nat Genet **18**:136-141.
93. Moore, J. C., and F. H. Arnold. 1996. Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. Nat Biotechnol **14**:458-467.
94. Nagahara, H., A. M. Vocero-Akbani, E. L. Snyder, A. Ho, D. G. Latham, N. A. Lissy, M. Becker-Hapak, S. A. Ezhevsky, and S. F. Dowdy. 1998. Transduction of full-length TAT fusion proteins into mammalian cells: TAT-p27Kip1 induces cell migration. Nat Med **4**:1449-1452.
95. Nagy, A. 2000. Cre recombinase: the universal reagent for genome tailoring. Genesis **26**:99-109.
96. Nagy, A., C. Moens, E. Ivanyi, J. Pawling, M. Gertsenstein, A. K. Hadjantonakis, M. Pirity, and J. Rossant. 1998. Dissecting the role of N-myc in development using a single targeting vector to generate a series of alleles. Curr Biol **8**:661-664.
97. Ness, J. E., S. Kim, A. Gottman, R. Pak, A. Krebber, T. V. Borchert, S. Govindarajan, E. C. Mundorff, and J. Minshull. 2002. Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. Nat Biotechnol **20**:1251-1255.
98. Novak, A., C. Guo, W. Yang, A. Nagy, and C. G. Lobe. 2000. Z/EG, a double reporter mouse line that expresses enhanced green fluorescent protein upon Cre-mediated excision. Genesis **28**:147-155.
99. Nunes-Duby, S. E., H. J. Kwon, R. S. Tirumalai, T. Ellenberger, and A. Landy. 1998. Similarities and differences among 105 members of the Int family of site-specific recombinases. Nucleic Acids Res **26**:391-406.
100. Nunes-Duby, S. E., L. Matsumoto, and A. Landy. 1987. Site-specific recombination intermediates trapped with suicide substrates. Cell **50**:779-788.
101. Orban, P. C., D. Chui, and J. D. Marth. 1992. Tissue- and site-specific DNA recombination in transgenic mice. Proc Natl Acad Sci U S A **89**:6861-6865.
102. Otey, C. R., M. Landwehr, J. B. Endelman, K. Hiraga, J. D. Bloom, and F. H. Arnold. 2006. Structure-guided recombination creates an artificial family of cytochromes p450. PLoS Biol **4**:e112.
103. Pabo, C. O., and R. T. Sauer. 1984. Protein-DNA recognition. Annu Rev Biochem **53**:293-321.
104. Pabo, C. O., and R. T. Sauer. 1992. Transcription factors: structural families and principles of DNA recognition. Annu Rev Biochem **61**:1053-1095.
105. Papworth, M., M. Moore, M. Isalan, M. Minczuk, Y. Choo, and A. Klug. 2003. Inhibition of herpes simplex virus 1 gene expression by designer zinc-finger transcription factors. Proc Natl Acad Sci U S A **100**:1621-1626.
106. Pargellis, C. A., S. E. Nunes-Duby, L. M. de Vargas, and A. Landy. 1988. Suicide recombination substrates yield covalent lambda integrase-DNA complexes and lead to identification of the active site tyrosine. J Biol Chem **263**:7678-7685.
107. Peitz, M., K. Pfannkuche, K. Rajewsky, and F. Edenhofer. 2002. Ability of the hydrophobic FGF and basic TAT peptides to promote cellular uptake of recombinant Cre recombinase: a tool for efficient genetic engineering of mammalian genomes. Proc Natl Acad Sci U S A **99**:4489-4494.

108. Petyuk, V., J. McDermott, M. Cook, and B. Sauer. 2004. Functional mapping of Cre recombinase by pentapeptide insertional mutagenesis. J Biol Chem **279**:37040-37048.

109. Ramirez-Solis, R., P. Liu, and A. Bradley. 1995. Chromosome engineering in mice. Nature **378**:720-724.

110. Rebar, E. J., Y. Huang, R. Hickey, A. K. Nath, D. Meoli, S. Nath, B. Chen, L. Xu, Y. Liang, A. C. Jamieson, L. Zhang, S. K. Spratt, C. C. Case, A. Wolffe, and F. J. Giordano. 2002. Induction of angiogenesis in a mouse model using engineered transcription factors. Nat Med **8**:1427-1432.

111. Reetz, M. T., B. Brunner, T. Schneider, F. Schulz, C. M. Clouthier, and M. M. Kayser. 2004. Directed evolution as a method to create enantioselective cyclohexanone monooxygenases for catalysis in Baeyer-Villiger reactions. Angew Chem Int Ed Engl **43**:4075-4078.

112. Reynolds, L., C. Ullman, M. Moore, M. Isalan, M. J. West, P. Clapham, A. Klug, and Y. Choo. 2003. Repression of the HIV-1 5' LTR promoter and inhibition of HIV-1 replication by using engineered zinc-finger transcription factors. Proc Natl Acad Sci U S A **100**:1615-1620.

113. Ringrose, L., V. Lounnas, L. Ehrlich, F. Buchholz, R. Wade, and A. F. Stewart. 1998. Comparative kinetic analysis of FLP and cre recombinases: mathematical models for DNA binding and recombination. J Mol Biol **284**:363-384.

114. Rodriguez, C. I., F. Buchholz, J. Galloway, R. Sequerra, J. Kasper, R. Ayala, A. F. Stewart, and S. M. Dymecki. 2000. High-efficiency deleter mice show that FLPe is an alternative to Cre-loxP. Nat Genet **25**:139-140.

115. Rufer, A. W., and B. Sauer. 2002. Non-contact positions impose site selectivity on Cre recombinase. Nucleic Acids Res **30**:2764-2771.

116. Santoro, S. W., and P. G. Schultz. 2002. Directed evolution of the site specificity of Cre recombinase. Proc Natl Acad Sci U S A **99**:4185-4190.

117. Saraf-Levy, T., S. W. Santoro, H. Volpin, T. Kushnirsky, Y. Eyal, P. G. Schultz, D. Gidoni, and N. Carmi. 2006. Site-specific recombination of asymmetric lox sites mediated by a heterotetrameric Cre recombinase complex. Bioorg Med Chem **14**:3081-3089.

118. Sato, Y., K. Tanaka, G. Lee, Y. Kanegae, Y. Sakai, S. Kaneko, H. Nakabayashi, T. Tamaoki, and I. Saito. 1998. Enhanced and specific gene expression via tissue-specific production of Cre recombinase using adenovirus vector. Biochem Biophys Res Commun **244**:455-462.

119. Sauer, B. 1996. Multiplex Cre/lox recombination permits selective site-specific DNA targeting to both a natural and an engineered site in the yeast genome. Nucleic Acids Res **24**:4608-4613.

120. Sauer, B. 1992. Identification of cryptic lox sites in the yeast genome by selection for Cre-mediated chromosome translocations that confer multiple drug resistance. J Mol Biol **223**:911-928.

121. Sauer, B. 2002. Cre/lox: one more step in the taming of the genome. Endocrine **19**:221-228.

122. Sauer, B., and N. Henderson. 1990. Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. New Biol **2**:441-449.

123. Sauer, B., and N. Henderson. 1988. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. Proc Natl Acad Sci U S A **85**:5166-5170.

124. Schnutgen, F., N. Doerflinger, C. Calleja, O. Wendling, P. Chambon, and N. B. Ghyselinck. 2003. A directional strategy for monitoring Cre-mediated recombination at the cellular level in the mouse. Nat Biotechnol **21**:562-565.

125. Schwarze, S. R., A. Ho, A. Vocero-Akbani, and S. F. Dowdy. 1999. In vivo protein transduction: delivery of a biologically active protein into the mouse. Science **285**:1569-1572.

126. Schwarze, S. R., K. A. Hruska, and S. F. Dowdy. 2000. Protein transduction: unrestricted delivery into all cells? Trends Cell Biol **10**:290-295.

127. Sclimenti, C. R., B. Thyagarajan, and M. P. Calos. 2001. Directed evolution of a recombinase for improved genomic integration at a native human sequence. Nucleic Acids Res **29**:5044-5051.

128. Seibler, J., B. Zevnik, B. Kuter-Luks, S. Andreas, H. Kern, T. Hennek, A. Rode, C. Heimann, N. Faust, G. Kauselmann, M. Schoor, R. Jaenisch, K. Rajewsky, R. Kuhn, and F. Schwenk. 2003. Rapid generation of inducible mouse mutants. Nucleic Acids Res **31**:e12.

129. Sherratt, D. J., and D. B. Wigley. 1998. Conserved themes but novel activities in recombinases and topoisomerases. Cell **93**:149-152.

130. Shimshek, D. R., J. Kim, M. R. Hubner, D. J. Spergel, F. Buchholz, E. Casanova, A. F. Stewart, P. H. Seeburg, and R. Sprengel. 2002. Codon-improved Cre recombinase (iCre) expression in the mouse. Genesis **32**:19-26.

131. Sieber, V., C. A. Martinez, and F. H. Arnold. 2001. Libraries of hybrid proteins from distantly related sequences. Nat Biotechnol **19**:456-460.

132. Soong, N. W., L. Nomura, K. Pekrun, M. Reed, L. Sheppard, G. Dawes, and W. P. Stemmer. 2000. Molecular breeding of viruses. Nat Genet **25**:436-439.

133. Soriano, P. 1999. Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat Genet **21**:70-71.

134. Srinivas, S., T. Watanabe, C. S. Lin, C. M. William, Y. Tanabe, T. M. Jessell, and F. Costantini. 2001. Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. BMC Dev Biol **1**:4.

135. Stemmer, W. P. 1994. Rapid evolution of a protein in vitro by DNA shuffling. Nature **370**:389-391.

136. Sternberg, N., B. Sauer, R. Hoess, and K. Abremski. 1986. Bacteriophage P1 cre gene and its regulatory region. Evidence for multiple promoters and for regulation by DNA methylation. J Mol Biol **187**:197-212.

137. Stoop, A. A., E. Eldering, T. R. Dafforn, R. J. Read, and H. Pannekoek. 2001. Different structural requirements for plasminogen activator inhibitor 1 (PAI-1) during latency transition and proteinase inhibition as evidenced by phage-displayed hypermutated PAI-1 libraries. J Mol Biol **305**:773-783.

138. Subramanya, H. S., L. K. Arciszewska, R. A. Baker, L. E. Bird, D. J. Sherratt, and D. B. Wigley. 1997. Crystal structure of the site-specific recombinase, XerD. Embo J **16**:5178-5187.

139. Suzuki, M., and N. Yagi. 1994. DNA recognition code of transcription factors in the helix-turn-helix, probe helix, hormone receptor, and zinc finger families. Proc Natl Acad Sci U S A **91**:12357-12361.

140. Tawfik, D. S., and A. D. Griffiths. 1998. Man-made cell-like compartments for molecular evolution. Nat Biotechnol **16**:652-656.

141. Testa, G., Y. Zhang, K. Vintersten, V. Benes, W. W. Pijnappel, I. Chambers, A. J. Smith, A. G. Smith, and A. F. Stewart. 2003. Engineering the mouse genome with bacterial artificial chromosomes to create multipurpose alleles. Nat Biotechnol **21**:443-447.

142. Urnov, F. D., J. C. Miller, Y. L. Lee, C. M. Beausejour, J. M. Rock, S. Augustus, A. C. Jamieson, M. H. Porteus, P. D. Gregory, and M. C. Holmes. 2005. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature **435**:646-651.

143. Utomo, A. R., A. Y. Nikitin, and W. H. Lee. 1999. Temporal, spatial, and cell type-specific control of Cre-mediated DNA recombination in transgenic mice. Nat Biotechnol **17**:1091-1096.

144. Van Duyne, G. D. 2001. A structural view of cre-loxp site-specific recombination. Annu Rev Biophys Biomol Struct **30**:87-104.

145. Varshavsky, A. 1997. The N-end rule pathway of protein degradation. Genes Cells **2**:13-28.

146. Vocero-Akbani, A. M., N. V. Heyden, N. A. Lissy, L. Ratner, and S. F. Dowdy. 1999. Killing HIV-infected cells by transduction with an HIV protease-activated caspase-3 protein. Nat Med **5**:29-33.

147. Voziyanov, Y., A. F. Stewart, and M. Jayaram. 2002. A dual reporter screening system identifies the amino acid at position 82 in Flp site-specific recombinase as a determinant for target specificity. Nucleic Acids Res **30**:1656-1663.

148. Wan, L., M. B. Twitchett, L. D. Eltis, A. G. Mauk, and M. Smith. 1998. In vitro evolution of horse heart myoglobin to increase peroxidase activity. Proc Natl Acad Sci U S A **95**:12825-12831.

149. Wang, L., W. C. Jackson, P. A. Steinbach, and R. Y. Tsien. 2004. Evolution of new nonantibody proteins via iterative somatic hypermutation. Proc Natl Acad Sci U S A **101**:16745-16749.

150. Wang, Y., L. A. Krushel, and G. M. Edelman. 1996. Targeted DNA recombination in vivo using an adenovirus carrying the cre recombinase gene. Proc Natl Acad Sci U S A **93**:3932-3936.

151. Wojciak, J. M., D. Sarkar, A. Landy, and R. T. Clubb. 2002. Arm-site binding by lambda -integrase: solution structure and functional characterization of its amino-terminal domain. Proc Natl Acad Sci U S A **99**:3434-3439.

152. Woods, K. C., S. S. Martin, V. C. Chu, and E. P. Baldwin. 2001. Quasi-equivalence in site-specific recombinase structure and function: crystal structure and activity of trimeric Cre recombinase bound to a three-way Lox DNA junction. J Mol Biol **313**:49-69.

153. Wunderlich, F. T., H. Wildner, K. Rajewsky, and F. Edenhofer. 2001. New variants of inducible Cre recombinase: a novel mutant of Cre-PR fusion protein exhibits enhanced sensitivity and an expanded range of inducibility. Nucleic Acids Res **29**:E47.

154. Xia, G., L. Chen, T. Sera, M. Fa, P. G. Schultz, and F. E. Romesberg. 2002. Directed evolution of novel polymerase activities: mutation of a DNA polymerase into an efficient RNA polymerase. Proc Natl Acad Sci U S A **99**:6597-6602.

155. Yoshikuni, Y., T. E. Ferrin, and J. D. Keasling. 2006. Designed divergent evolution of enzyme function. Nature **440**:1078-1082.

156. Zha, D., A. Eipper, and M. T. Reetz. 2003. Assembly of designed oligonucleotides as an efficient method for gene recombination: a new tool in directed evolution. Chembiochem **4**:34-39.

157. Zhang, Y., F. Buchholz, J. P. Muyrers, and A. F. Stewart. 1998. A new logic for DNA engineering using recombination in Escherichia coli. Nat Genet **20**:123-128.

158. Zhang, Y., J. P. Muyrers, G. Testa, and A. F. Stewart. 2000. DNA cloning by homologous recombination in Escherichia coli. Nat Biotechnol **18**:1314-1317.

159. Zhao, H., L. Giver, Z. Shao, J. A. Affholter, and F. H. Arnold. 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat Biotechnol **16**:258-261.

160. Zheng, B., M. Sage, E. A. Sheppeard, V. Jurecic, and A. Bradley. 2000. Engineering mouse chromosomes with Cre-loxP: range, efficiency, and somatic applications. Mol Cell Biol **20**:648-655.

161. Zinyk, D. L., E. H. Mercer, E. Harris, D. J. Anderson, and A. L. Joyner. 1998. Fate mapping of the mouse midbrain-hindbrain constriction using a site-specific recombination system. Curr Biol **8**:665-668.