Summer 2016

Genome engineering in mammalian cells by Flp and Cre DNA recombinase variants

Riddhi Shah

GENOME ENGINEERING IN MAMMALIAN CELLS BY FLP AND CRE DNA RECOMBINASE VARIANTS

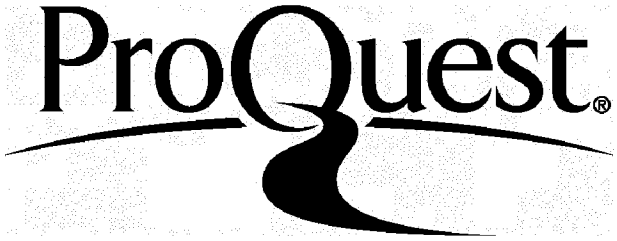
by

Riddhi Shah, BS, MS, MSNT

A Dissertation Presented in Partial Fulfillment
of the Requirements of the Degree
Doctor of Philosophy

COLLEGE OF APPLIED AND NATURAL SCIENCES
LOUISIANA TECH UNIVERSITY

August 2016



LOUISIANA TECH UNIVERSITY

THE GRADUATE SCHOOL

May 9, 2016
Date

We hereby recommend that the dissertation prepared under our supervision by Riddhi Shah, BS, MS, MSNT

entitled Genome Engineering in Mammalian Cells by Flp and Cre DNA Recombinase Variants

be accepted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Molecular Sciences and Nanotechnology

Yuri Vizigyanov
Supervisor of Dissertation Research

Head of Department
School of Biological Sciences
Department

Recommendation concurred in:

Advisory Committee

Approved:

Director of Graduate Studies

Approved:

Dean of the Graduate School

Dean of the College

GS Form 13a
(6/07)

ABSTRACT

Genome engineering relies on DNA modifying enzymes that are able to locate a DNA sequence of interest and initiate a desired genome rearrangement. Currently, the field predominantly utilizes site-specific DNA nucleases that depend on the host DNA repair machinery to complete a genome modification task. We show here that genome engineering approaches that employ self-sufficient, versatile site-specific DNA recombinase Flp and Cre can be developed into promising alternatives. We demonstrate that a Flp variant evolved to recombine an *FRT*-like sequence *FL-IL10A*, which is located upstream of the human interleukin-10 gene, can target this sequence in the model setting and native HEK293 cells. Similarly, Cre variant evolved to recombine at loxP-like sequence LL-69058, which is located upstream of the beta globin gene, can target this sequence in the model setting of CHO cells and human HEK293 cells.

The target-specific Flp variant is able to perform the integration reaction but the efficiency of the integration reaction in human cells can be enhanced by 'humanizing' the Flp variant gene and by adding the NLS sequence to the recombinase. Cre variant displays a poor replacement activity in the mammalian cells and thus is fused with TAL DNA-binding domain to enhance their performance. TAL-Cre variant is able to perform efficient replacement reaction, when paired with another recombinase (dual RMCE). The Cre variant replacement activity is observed only when they are fused to the TAL

DNA-binding domain. The TAL-Cre fusion has higher replacement efficiency compared to no replacement activity with just Cre variant.

APPROVAL FOR SCHOLARLY DISSEMINATION

The author grants to the Prescott Memorial Library of Louisiana Tech University the right to reproduce, by appropriate methods, upon request, any or all portions of this Dissertation. It is understood that "proper request" consists of the agreement, on the part of the requesting party, that said reproduction is for his personal use and that subsequent reproduction will not occur without written approval of the author of this Dissertation. Further, any portions of the Dissertation used in books, papers, and other works must be appropriately referenced to this Dissertation.

Finally, the author of this Dissertation reserves the right to publish freely, in the literature, at any time, any or all portions of this Dissertation.

Author RIDDHI SHAH

Date 08/01/2016

DEDICATION

This dissertation is dedicated to my parents, Jitendra S. Shah and Asha J. Shah.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

This dissertation could not have been finished without the help and support from many professors, research staff, graduate students, colleagues and my family. It is my great pleasure to acknowledge people who have given me guidance, help and encouragement.

I would like to first thank Dr. Yuri Voziyanov for his constant guidance, personal attention, suggestions for my graduate study and research. He gave me opportunities to participate in this research and to work in his laboratory.

Special thanks go to my advisory committee members, professors Dr. Jamie Newman, Dr. Rebecca Giorno, Dr. William Wolf and Dr. Patrick Hindmarsh, for their endless, important suggestions and for their time and effort in service on my doctoral committee despite their already heavy loads of responsibility.

I would like to express my sincere appreciation to Dr. Jamie Newman who gave me enormous valuable discussions and hands-on help in many aspects of this research program, and for her caring and concern for me throughout my dissertation.

It has been my pleasure to work with graduate students and colleagues in the School of Biological Sciences; they are Feng Li, Ngozi, Anusha Enumalai and Rawieh. We discussed together, exchanged ideas, and helped each other in many aspects of graduate school studies. Those made my years at Louisiana Tech the most memorable period in my life.

Finally, I would express my gratitude to my parents and my sister, Mansi, whose support and constant encouragement helped me through the hard times of this program. My deepest appreciation is expressed to them for their love, understanding, and inspiration. Without their blessings and encouragement, I would not have been able to finish this work.

CHAPTER 1

INTRODUCTION

1.1 Background

Human genetic disorders are caused by abnormalities in the genome which are then passed to offspring. Some of the genetic disorders that are caused by mutation in a single gene include sickle cell disease (SCD), cystic fibrosis, neurofibromatosis (NF), Tay-Sachs disease, Duchenne muscular dystrophy (DMD) and platelet type bleeding disorder as well as many others. Gene therapy is hoped to cure genetic disorders by adding the wild type gene or replacing the mutated gene with a wild type gene, restoring proper function in the cells. Initially, viral vectors were used as gene delivery vehicles but this approach poses potential risk of inserting the vector randomly in the genome, potentially activating the tumor promoting pathways. An alternative to viral vector based gene therapy could be the correction of the mutated gene using site-specific recombinases. These enzymes have been shown to integrate, excise, invert genes, and replace gene fragments in mammalian cells.

My research aims to use site-specific recombinases for integration and replacement of genes in the human genome with the goal of curing a genetic disease. Recombinases bind to specific DNA sequences called recombination targets and catalyze DNA rearrangement. For example, integrase from the bacteriophage lambda, specific to *attP* and *attB* target sequences, catalyze integration of the viral genome into the bacterial

chromosome. XerCD from the *Escherichi coli* is specific to the dif target sequence in the bacterial genome and also *dif*-like target sequences *cer* (plasmid ColE1), and *psi* (plasmid SC101) sites. XerCD plays an important role by resolving chromosome and plasmid dimers. Cre recombinase from bacteriophage P1, specific to the *loxP* target sequence, catalyzes the circularization of linear phage genomes. Flp recombinase that is encoded by a 2 micron plasmid from *Saccharomyces cerevisiae* and is specific for the *FRT* target sequence, helps in maintaining the plasmid copy number by inverting one part of the 2 micron plasmid relative to the other part during replication [1- 4].

Site-specific tyrosine recombinases are primarily found in the prokaryotes, but they have also been found in the archaea and some eukaryotes. PSI-BLAST, a computer software, identifies more than 1000 related sequences of the tyrosine recombinase family [3,4]. Therefore, the 3D fold of tyrosine site specific recombinases is adapted to recognize a wide variety of target sequences. This phenomenon can be exploited to evolve recombinase variants that recognize target-like sequences that resemble the native target sequences. Such target-like sequences for a recombinase can be found in the human genome on average every 5000 bp (5 kb) [5]. For the purpose of gene therapy, target-like sequences can be used as recombination targets to either integrate or replace a gene region.

Recombinases can be used as tools for genome manipulation. In principle, recombinases can be altered to recognize target-like sequences using directed evolution approaches that include a combination of several techniques: DNA shuffling, PCR mutagenesis and error prone PCR. The quest to generate variants of tyrosine recombinases with desired target specificity began when several mutant and chimeric λ

integrase variants were shown to recombine target sites for HK022 integrase [4,5]. The research was boosted by the advances in two molecular evolution techniques: error-prone PCR and DNA shuffling which allow for the fast generation of recombinase variant libraries [6-8]. The solved structures of the Cre/*loxP* and Flp/*FRT* complexes [7,8] provided another boost: they helped utilize another molecular evolution technique, site-directed mutagenesis of amino acids that can affect DNA binding so that they can bind to the target-like sequences efficiently.

For this project, I will focus on two of the approximately 600,000 recombinase target-like sequences. One of the target-like sequences (*FL-IL10A*) is located in the interleukin gene on chromosome 1 (Figure 1.1). The second target-like sequence (*LL-69058*) is located in the beta-globin gene on chromosome 11 (Figure 1.2). These target-like sequences resemble the *FRT* and *loxP* sequences, respectively. The long term goal of my project is to evolve and test the Flp and Cre recombinase variants specific for these target-like sequences.

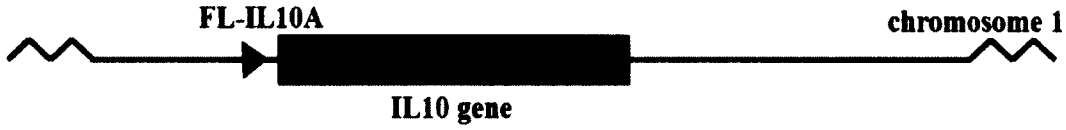


Figure 1.1: Relative location of the FL-IL10A sequence.

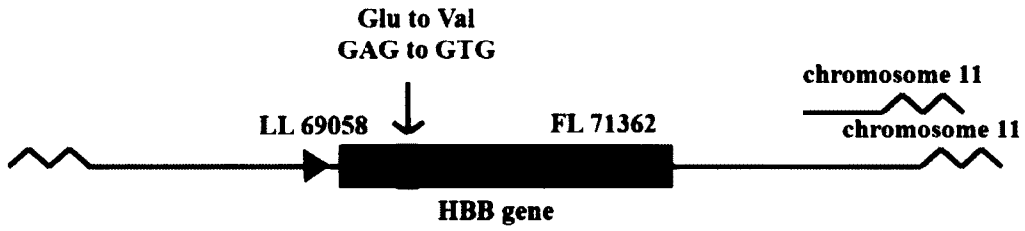


Figure 1.2: Relative locations of LL-69058, FL-71362 and E6V (Glu to Val) mutation.

1.2 Objectives

The objectives for my research project are as follows:

1. Test variants of Flp recombinase that can target *FRT*-like sequence *FL-IL10A* from the human interleukin *(IL10)* gene in mammalian cells (Figure 1.1).

2. Evolve and test variants of Cre recombinase that can target *loxP*-like sequence *LL-69058* from the human beta globin *(HBB)* gene in mammalian cells (Figure 1.2).

1.3 Hypotheses

1. Flp recombinase variants functionally tested in *E.coli* can integrate a vector into *FL-IL10A* in mammalian cells. (Figure 1.1)

2. Cre recombinase variants can be evolved in *E.coli* to recognize *LL-69058* and functionally tested in mammalian cells to replace a gene cassette when paired with evolved Flp recombinase variants specific to *FL-71362*. (Figure 1.2)

1.4 Literature Review

1.4.1 Conservative Site Specific Recombination and Types of Recombinases

Site-specific recombinases are actively used in mammalian cells to engineer genes and chromosomes. These recombinases can delete, insert, or invert a gene cassette flanked by the recombinase target sequences resulting in activating or silencing the gene cassette [11].

Site-specific recombination leads to DNA rearrangement at specific sequences called the recombination target (RT) sequences [12]. "Conservative" site-specific recombination means that there is no loss or addition of nucleotides at the time of DNA rearrangement [13]. Recombinases were discovered in bacteria, temperate phages, and

yeast where they exist to perform various events like resolution, inversion and integration. Conservative site specific recombination was discovered by Allan Campbell in 1968 during his studies of bacteriophage lambda and its host bacterium *E.coli* [14].

Site-specific recombinases function on the principle of strand breakage and rejoining and the recombination is divided into several steps. In general, the recombinase catalyzes cleavage, strand exchange, and rejoining of the DNA within the synaptic complex to obtain recombined products [4]. The recombinase binds to the two recombinase target sequences forming a synaptic complex. The core target sequence consists of two inverted repeat sequences flanking a spacer or crossover region [15]. Depending on the recombination mechanism, the requirement for accessory proteins, recognition of target sequence and strand exchange at the target sequences, recombinases are divided into two families: serine (Resolvases/Invertases) and tyrosine (Integrases) recombinases [2,9,13] (Table 1.1).

Table 1.1: Examples of recombinases from serine and tyrosine family and their biological functions [2,9,12-14].

Tyrosine Recombinase		**Biological Function**
	λ-Integrases	**Integration and excision**
	XerC/D	**Excision and resolution**
	Flp	**Inversion and resolution**
	Cre	**Excision**
	Int I	**Integration and excision**
	XisA/XisC	**Excision**
	FimB, FimE	**Inversion**
	Int of Tn916/Tn 1545	**Integration and excision**
Serine Recombinase		
Resolvase		
Invertase		
Integrase		

Some recombinases belonging to the Invertase and Integrase family have accessory binding sites for expression of enhancer elements away from crossover regions (synapse) [17]. The orientation of target sequences and the location of the enhancer sequences for some recombinases are shown in Figure 1.3.

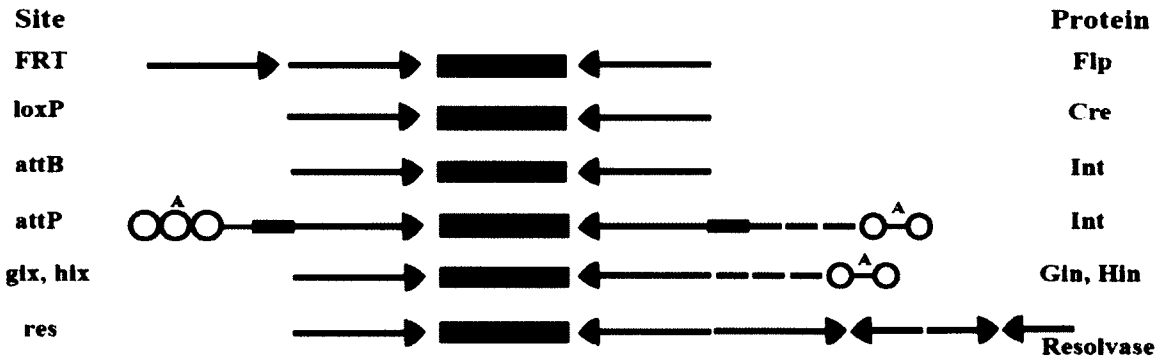


Figure 1.3: Target sequences for conservative site-specific recombinase families. The brown rectangular box represents the binding target sequence and the inverted arrows represent the orientation of the target sequence. The circle (A) represent the accessory site away from the target sequence express the enhancer elements [14].

1.4.1.1 Serine Recombinase Family. Serine residues play a major role as a conserved amino acid that attacks the DNA and covalently links to the DNA during strand exchange. The Serine recombinase family is sub-divided into groups based on performance of different recombination events like resolution, inversion, and integration. The major groups of this family are Resolvases (Tn3, γδ), Invertases (Hin, Gin, Cin), and Integrases (ϕC31) (Table 1.1). The Tn3 resolvase, mobile genetic elements called transposons found in prokaryotes, require two 28 bp *res* sequence in direct repeat and six resolvase dimers to form a fully active synapse. The ϕC31 integrases interact with DNA attachment sequences called phage *attP* and bacterial *attB* sequences to give recombinant products. These attachment sites favor the integration of phage genome into the *E.coli* genome forming hybrid attL and attR attachment sites. The *attP* and *attB* sequences are 39 bp and 34 bp in length, respectively. Evolved ϕC31 integrases are often used as a tool to manipulate the human genome [2]. When ϕC31 are linked to zinc finger nucleases, they exhibit rapid genome engineering strategies to recombine at specific loci in the

human cells or non-human origin cells [18]. The invertase recombinases like the Gin that recognize *gix*, Hin that recognize *hix* (Figure 1.4) require invert repeats of sequence to give recombinant products.

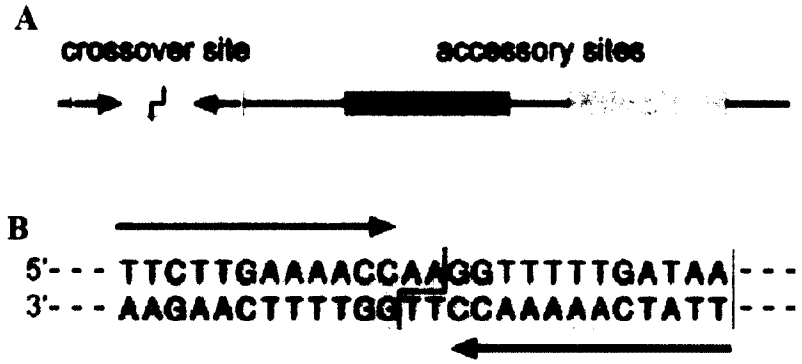


Figure 1.4: Serine recombinase target sequence for invertase family indicating location of accessory sites that express enhancer proteins (A) and target sequence of Hin recombinase (B). The inverted arrows show the orientation of hix target sequence [1].

The Ser residue located in the catalytic domain of serine recombinase cleaves the phosphodiester bond of DNA on all four strands producing a 3'-hyroxyl end and 5'-phosphoserine linkage ends as intermediate products [10,15]. In the process of strand exchange a structural complex called a synapse is formed. The enzyme-DNA complex is rotated and ligated to obtain recombines DNA (Figure 1.5).

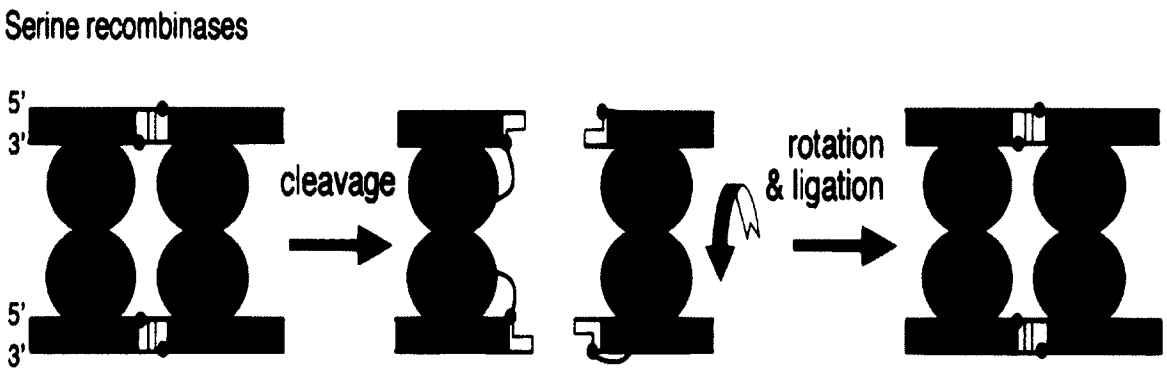


Figure 1.5: Schematic representation of strand exchange mediated by serine recombinase [19].

This conserved region contributes to the active site of the recombinase. The C-rotated and ligated to obtain recombined DNA (Figure 1.5). terminal helix-turn-helix (HTH) domain of serine recombinases contain highly conserved residues. The N-terminal catalytic domain consists of at least 10 highly conserved residues including three cysteine and non-conserved residues rich in leucine/isoleucine/valine/methionine [1]. The conserved residues of invertase family Hin recombinase (which are Fis dependent) are Gln151, Arg154 and Leu155, which directly or indirectly interact with Fis. Any mutation in the residues can inactivate Hin. γδ resolvase has conserved residues at positions Lys54 and Glu56 which participate in protein-protein interaction between cut DNA strands (Figure 1.6) [17].

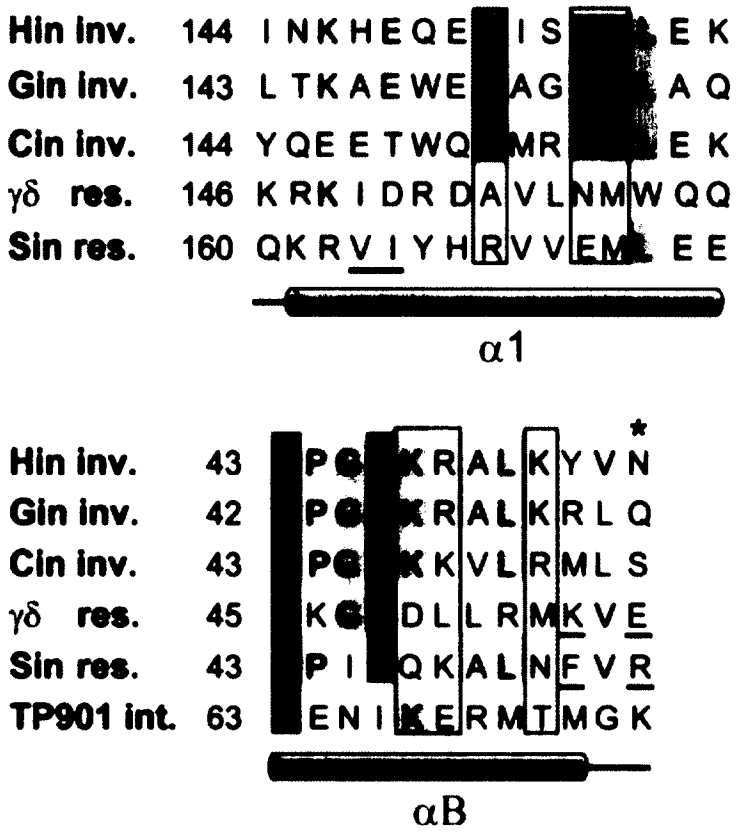


Figure 1.6: Sequence alignment of DNA Binding Domains of helix α1 and helix αB region of invertases and resolvases members belonging to the serine recombinase family [17]. The number at the start of the sequence indicates the position of the first residue and so on. The red rectangular box indicates the important conserved residues.

The synaptic complex for serine recombinase forms only when the DNA is negatively supercoiled. The invertase group of serine recombinase requires accessory proteins for their activity. The Fis/enhancer system acts as accessory proteins and controls the DNA cleavage and exchange [17]. The system is encoded by a sequence region called an accessory sequence located away from the crossover sequence [15,17] (Figure 1.5).

Zinc finger nucleases (ZFN) and TALE DNA-binding domains can be adapted to recognize genomic target sequences. By engineering the ZFN and TALE binding

specificity and reducing or altering the residue in the serine recombinase structure to recognize the genomic target sequences it can help increase the application of evolved recombinases.

1.4.1.2 Tyrosine Recombinase Family. Tyrosine recombinase families are called so because the tyrosine residue plays a major role as a conserved amino acid that attacks the DNA and covalently links to the DNA during strand exchange. The tyrosine residue attacks the phosphodiester bond on DNA producing 3' phosphotyrosyl intermediate. Some members of the tyrosine family include Flp, Cre, λ-Integrase, Xerc/D and Int I (Table 1.1). They can perform recombination events such as excision, integration, resolution and replacement.

Other examples of tyrosine recombinase family members are Int recombinase which recognizes the *attP* and *attB* site, Cre recombinase that recognizes the *loxP* site, and Flp recombinase that recognizes the *FRT* site [11,20]. Figure 1.10 indicates the target sequences *loxP* and *FRT* for Cre and Flp recombinases. The target sequences for recombinases are composed of 6-8 bp spacer regions flanked on either side by 13-14 bp inverted repeats [3,11,20–22]. Change in the bases of the target sequence can render the recombinase incapable of recognizing it [11,20,22].

Tyrosine recombinase binds to the target sequence (bringing sites together) resulting in a highly structured complex called a synapse. In the complex, a pair of strand exchanges occurs which proceed to an intermediate called a Holliday junction (Figure 1.7).

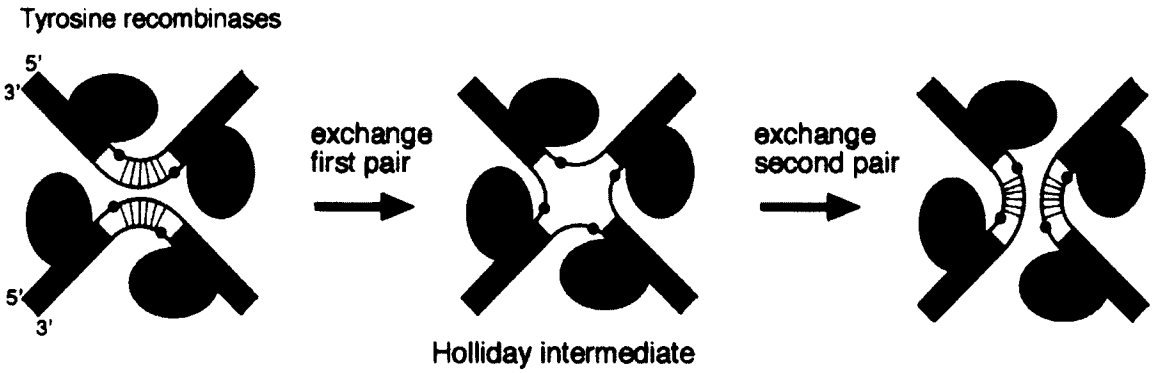
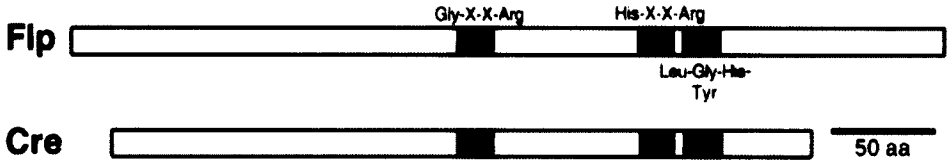


Figure 1.7: Schematic representation of strand exchange mediated by tyrosine recombinase [19].

The Holliday junction is resolved when the second pair strand exchange occurs forming recombined products [10,23,24]. Tyrosine recombinase on cleaving the DNA strand, produces a 5'-hydroxyl end and 3'-phosphotyrosine ends [10,24,25]. For tyrosinerecombinase, there should always be sequence homology at the spacer region/ crossover region for recombination to occur. The catalytic domain of tyrosine recombinase has four conserved residues located at the C-terminal of the protein (Arg-His-X-X-Arg-Tyr) and form the RHR traid (Arg191, His305 and Arg308) in Flp [10,12,21,25,26], (Figure 1.8). The RHR triad is responsible for the orientation of the phosphodiester bonds in the recombination steps [12,26].



	Arg		His		Arg		Lys		Trp		Tyr		
Flp	191	114	305	3	308	-85	223	107	330	13	343	80	*423 aa/49kDa*
Cre	173	116	289	3	292	-91	201	114	315	9	324	19	*343 aa/38kDa*

Figure 1.8: Conserved residues of Flp and Cre recombinase in the C-terminal domain. The tyrosine residing at position 343 in Flp and 324 in Cre in the catalytic site play a vital role in DNA cleavage [27].

Depending on the location and relative orientation of their targets, tyrosine recombinase variants can mediate insertions, deletions, replacements and translocations of DNA fragments. Moreover, site-specific recombinases were shown to perform genetic rearrangements in all cell types tested, from bacteria to plant to human [28,29].

Integration and excision are reversible processes where target sequences are located on different DNA strands. In Figure 1.9.A, the circular molecule can integrate into the genome and can be excised from the genome. Resolution yields a circular DNA plasmid into two small circular molecules. Two target sequences are located on the plasmid in the same orientation (Figure 1.9.B). In the inversion reaction, the two oppositely oriented target sequences flip the region between them (Figure1.9.C). In MCE, or replacement, the gene cassette is exchanged between two genomes or chromosomes in (Figure 1.9.D)

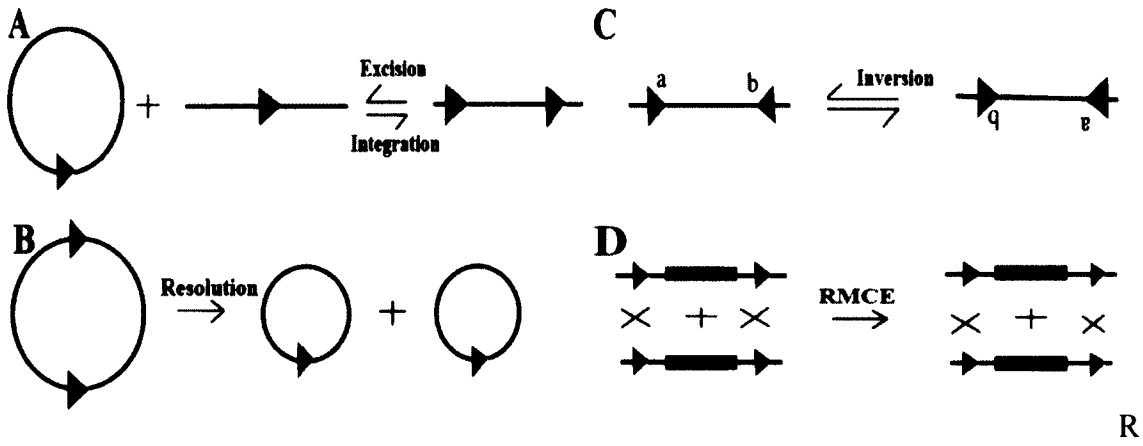


Figure 1.9: Different types of recombination reactions catalyzed by recombinases. A.) Excision and insertion of DNA in the genome at site-specific target sequence. B.) A circular DNA molecules resolving in two small circular DNA molecules by recombining at the target sites. C.) Inversion of DNA occurs when the target sequences are present in opposite directions. D.) RMCE (Recombinase Mediated Cassette Exchange) causes replacement of gene cassette which is flanked on either side with target sequences on two DNA strands.

1.4.1.2.A <u>Cre Recombinase</u>. Cre (cyclization recombinase) recombinases from the bacteriophage P1 were used for the first time in mammalian cells by Sauer in 1989 [30]. Cre recombinases perform two functions, circularization of linear DNA via the crossing over loxP, and resolution of dimers [20]. This recombinase does not require any accessory proteins and can perform recombination independently [23,31]. Cre recombinases bind specifically and efficiently at specific target sequences called loxP sites (Figure 1.10).

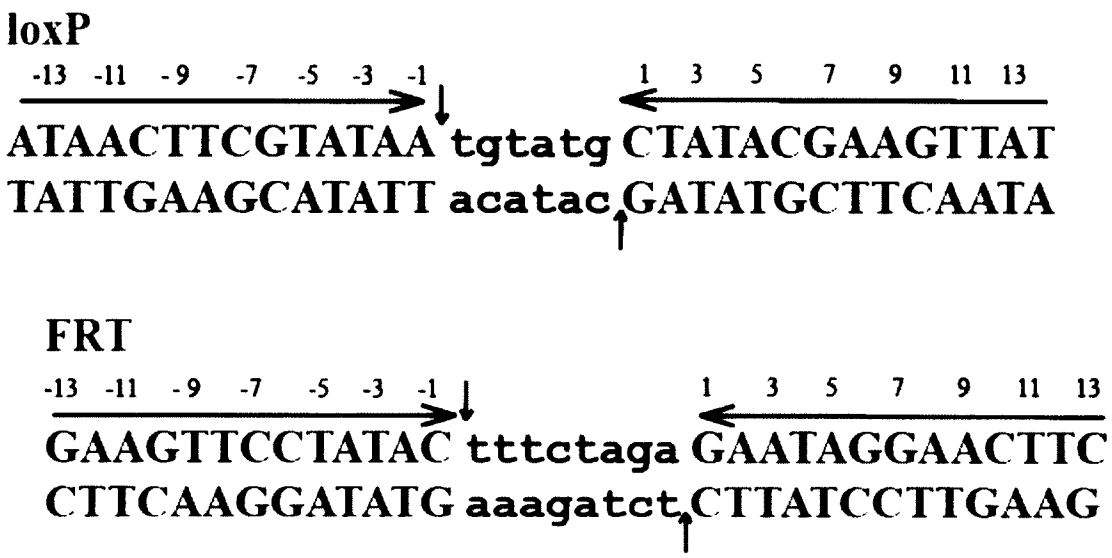


Figure 1.10: loxP and FRT target sequence with 6bp and 8bp spacer flanked with 13-14 bp invert repeats respectively. The inverted arrows indicate the inverted repeats with spacer region located between them. The small arrows point at the region of crossing over when recombinase binds to the inverted repeats. The numbers above the big arrows indicate the location of the bases that vary from 13-14 bp.

1.4.1.2.B <u>Flp Recombinase</u>. Flp recombinase is encoded in the 2μm plasmid of the yeast *Saccharomyces cerevisiae* and recognizes a specific target sequence called FRT (**F**lp **R**ecombination **T**arget) (Figure 1.9). Evans *et al,* in 1990, discovered that the tyrosine residue at position 343 in the Flp is responsible for its catalytic function but does not help in substrate recognition [32]. In yeast, Flp recombinases are responsible for preserving a high plasmid copy number [20].

Cre and Flp tyrosine recombinases have been widely used as tools for genome engineering. When these recombinases are evolved using directed evolution techniques to recognize the target sequences, it opens many new areas of genome manipulation to cure various diseases. The activity of the recombinases can be enhanced further by fusing

them to DNA binding domains of zinc-finger nucleases and TALE. CRISPR technology has also been utilized to guide the delivery of particular genes to desired destinations in the genome.

1.4.2 Evolution of Cre and Flp Tyrosine Recombinase Using Directed Evolution Techniques

In nature, living organisms undergo strict condition constraints that force them to evolve at the DNA level by inducing mutations in their structure. These mutations can be beneficial or harmful based on the survival of the organism. It takes several generations for the beneficial mutations to express in their genome. This phenomenon of introducing mutations can be applied in the laboratory and be used to enhance the property of a particular protein, which can result in improving the yield, or increasing the efficiency of an enzyme or particular pathway. DNA mutagenesis and DNA shuffling are the mostly widely used laboratory techniques for evolving a particular protein.

DNA mutagenesis can be performed in two different ways, depending on the knowledge of the protein structure. They are classified as random evolution technique and targeted evolution technique.

1.4.2.1 Random Evolution

1.4.2.1.1 Error-prone PCR. Error prone PCR or PCR mutagenesis was first discovered by Leung, *et al* in 1989 [33]. The protocol was modified by Cadwell and Joyce in 1992 [34]. Random mutations were introduced in the gene fragment by altering regular PCR conditions (Figure 1.11).

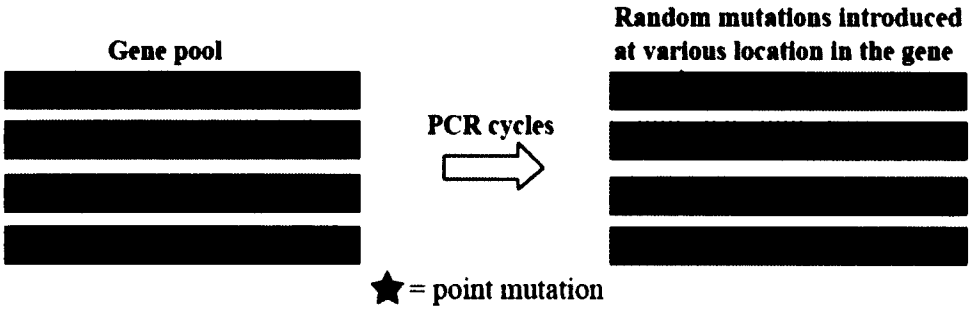


Figure 1.11: Diagrammatic representation of Cyclic Mutagenesis or Error Prone PCR, causing introduction of random mutations.

The parameters altered in the reaction include: (1) increasing the concentration of Taq polymerase; (2) increasing extension time; (3) increasing concentration of $MgCl^{2+}$; (4) addition of 0.5mM $MnCl^{2+}$ and (5) increasing 1mM concentration of GTP, CTP, TTP and 0.2mM ATP. Parameters of steps 4 and 5 have higher chances of introducing random mutation. The goal behind altering the PCR reaction parameter was to reduce the fidelity of amplifying the gene fragment [33].

1.4.2.1.2 DNA Shuffling. DNA shuffling, an evolution technique, randomly introduces mutations (insertion, deletion, or change in the codon) in the protein structure. Stemmer was the first to discover the technique of DNA shuffling, where he began with shuffling the LacZ gene to observe changes in the β-lactamase pathway [8]. The procedure, as shown in Figure 1.12, involves shuffling of two or more homologus genes with different mutations or allele genes. The genes are digested with DNAse, resulting into fragmentation of the genes. Overlapping ends are produced upon digestion as DNAse makes random cuts in the gene. Small fragments of the digested gene are observed on the gel. The gene fragments are assembled in a PCR reaction to reform the original gene.

The overlapping ends act as primers to anneal the fragments together. During this PCR, in the absence of a primer, random mutations are introduced into the gene fragment. The last step involves the use of specific forward and reverse primers to amplify the whole gene. The PCR mix is a library of different recombined genes having different mutations. The mutations are usually additive enhancing the property of the protein or enzyme. Screening of the gene pool for their activity can help in narrowing down to the best possible variant which has much better activity than the original genes used for shuffling [8].

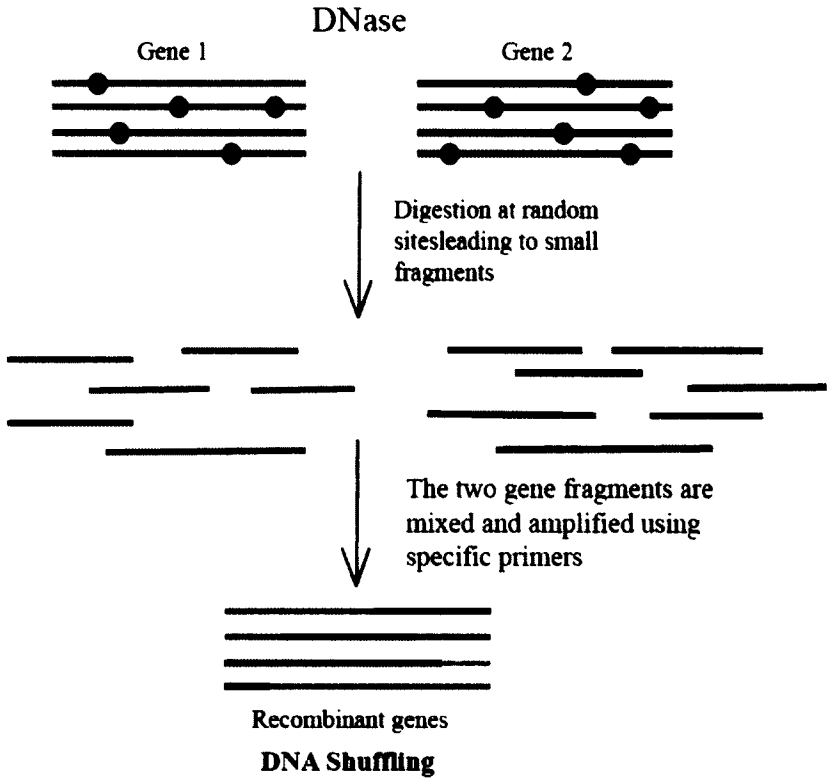


Figure 1.12: General overview of DNA shuffling

Zhao and Arnold, in 1997, came up with various modifications for DNA shuffling that focused on controlling the number of mutations introduced in the shuffling process. The rate of mutagenesis was controlled by manipulating the concentration of DNA

polymerase, Mg^{2+} and Mn^{2+} in the PCR reactions. This approach was thought to have a higher impact on the evolution of proteins and understanding the evolutionary related genes [9].

1.4.2.2 Targeted Evolution - Site-Directed Mutagenesis. The main goal of site directed mutagenesis is to introduce a single mutation at a particular site in the gene using mutagenic primers under regular PCR conditions. The goal is to obtain a single mutation, either a point mutation, deletion or insertion, with a minimum number of PCR cycles. Mutagenesis using mismatched oligonucleotides is currently being used in research to construct mutants. Other methods include mutagenesis using single stranded vectors, mutagenesis using double stranded vectors and mutagenesis by total chemical synthesis, as shown in Figure 1.13 [35].

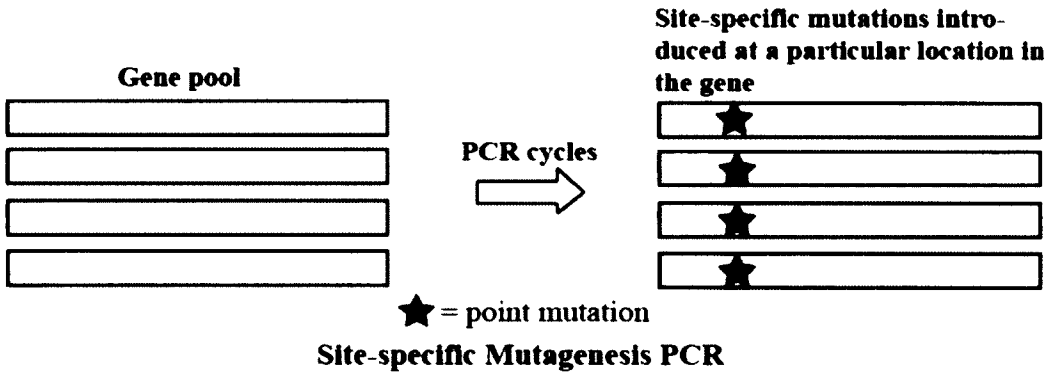


Figure 1.13: Diagrammatic representation of mutation introduced due to site directed mutagenesis.

The basic operations performed in the method with mutagenic oligonucleotides involves using primers that have the mutagenic site either on one or both primers. The primers should anneal at the 5' end of the mutagenic sites and should be able to amplify the whole plasmid. Regular PCR conditions are set to allow the primers to anneal at the

required site with a minimum number of cycles. Once the PCR product is obtained, it is treated with DpnI to get rid of the parent strand and transformed through the *E. coli.* Plasmids obtained are confirmed through sequencing for the desired mutation [36].

1.4.3 Enhancing the Efficiency of Evolved Cre and Flp Recombinase Variants

To enhance the efficiency of evolved Cre and Flp recombinase variants, the recombinase can be fused to DNA binding domains of the Zinc-finger (ZF) module and TALE (transcription activator-like effector) module. These modules were designed to have specificity in binding to the DNA because of their DNA binding domain. Similarly, CRISPR protein is another tool used in genome engineering for expanding the targeting efficiency of the recombinases.

The ZFR (**Z**inc **F**inger **R**ecombinases) perform the recombination autonomously and do not activate the DNA response pathway [18]. But the short-comings in using zinc-finger proteins are that they are limited only to the 64 possible triplet codon combinations [37]. Furthermore, constructing zinc-finger proteins specific to certain DNA sequences is time consuming and labor intensive [42]. They perform integration with low specificity and can bind to off-target sequences and thus increase the risk of having higher rates of failure [38]. Due to these drawbacks, ZFR was replaced by more accurate technology like TALE and CRISPR.

CRISPR was used as an immune defense against invading DNA from bacteriophage viruses in many diverse bacteria and archaea groups [39,40]. This is the most recent addition in genome modification and genome engineering, where Cas9 activity is based on the RNA-guided DNA endonuclease [40-42] . CRISPRs are used as a defense mechanism in bacteria and archaea against exogenous DNA. They are a fusion of

CRISPR RNA (crRNA) and CRISPR associated (Cas) protein that together direct the degradation of double strand DNA [40]. The double stranded breaks in the DNA due to Cas9 protein activate the non-homologus end joining (NHEJ) or the homology- directed (HDR) DNA repair pathway [42,43].

The nuclease in the Zinc-finger Nucleases (ZFN) and TALE nucleases can be substituted with site-specific recombinase [43], keeping the DNA binding domain and its function as it is, such that the recombinase can bind to the target sequence, and the DNA binding domain can be designed to bind to the DNA sequences upstream and downstream of the target site. DNA binding domains fused with recombinase can help in enhancing the efficiency of the recombinase by enhancing the interaction with their respective target recombination sequence. Discussed below are, in detail, the TALE DNA binding domains used to manipulate the genomic DNA. My project focuses on the use and optimization of TALE DNA binding domain linked to Cre recombinase to recognize the LL-69058 DNA sequences upstream of the beta globin gene.

1.4.3.1 TALE DNA-binding Domain. Transcription activator like effector (TALE) proteins provide an alternative for zinc finger recombinases. Advantages of using TALE proteins over ZFRs include greater DNA binding specificity, ease of engineering TALE DNA binding domains, and low toxicity and high cleavage activity when combined with recombinase [18,19]. TALEs are naturally occurring DNA-binding domains derived from the plant bacterial pathogen *Xanthomonas* spp. and *Ralstonia solanacearum* [18,20,21]. The TALE proteins were used to modulate gene transcription in host plants to facilitate bacterial colonization where the proteins are delivered to the nucleus and then to a specific promoter sequence to activate gene expression [35,37].

Generally, the DNA binding domain consists of 33-35 amino acid repeat modules that can be rearranged to bind to a new target DNA sequence [33,35,36]. Each DNA base (A, C, T and G) is encoded by repeat-variable di-residues (RVD's) called NI, HD, NG and NN respectively. The 'NI' RVD codes for 'A' DNA base, 'HD' for 'C', 'NG' for 'T' and 'NN' for 'G' DNA base. In Figure 1.14 below, the di-residues can be used to engineer the TALE DNA binding domain for any given sequence [38,39]. The TALE binding domain is similar to the ZF binding domains, but the specificity to bind to each nucleotide in the sequence in higher and it can be engineered to bind to any target sequence, makes the TALE's preferred over ZF's [47].

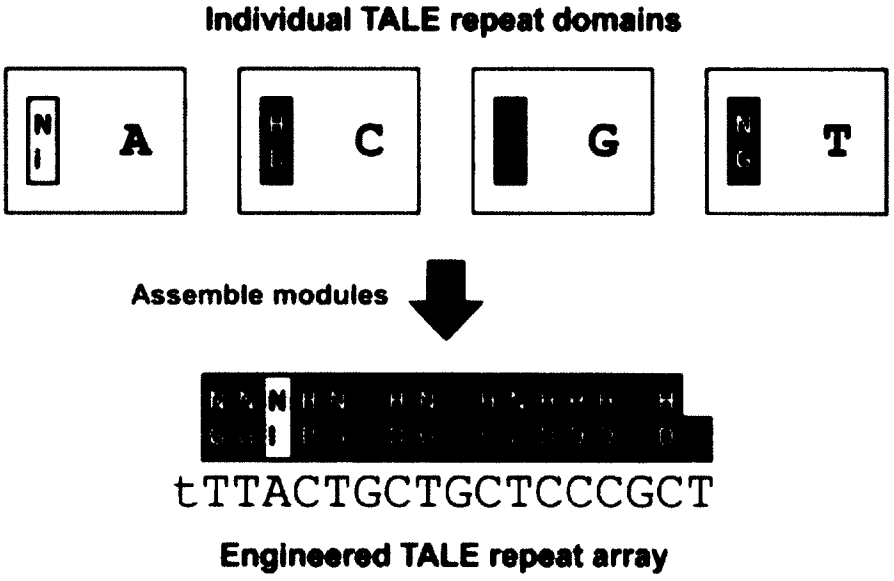


Figure 1.14: Coding pattern for each nucleotide base and engineering TALE DNA binding domain for a particular DNA sequence with the di-residue repeats.

In my studies, TAL (TALE) plays an important role in binding upstream and downstream of the LL-69058 target sequence, allowing the recombinase to bind specifically to the recombination target sequence. The protein in the TAL's is

recombinase, Flp or Cre, which acts as the catalytic domain, and the TAL module, which acts as the DNA binding domain. As shown in Figure 1.15, the TAL DNA binding domain was engineered to bind to the sequence upstream of the LL-69058 target sequence, making it feasible for the evolved Cre recombinase to identify and bind with higher efficiency to its respective target site. Similarly, TAL can be engineered for another DNA sequence, where recombination activity is to be performed, thus improving the probability of having site-specific recombination rather than random recombination in the genome with other partially similar target sequences.

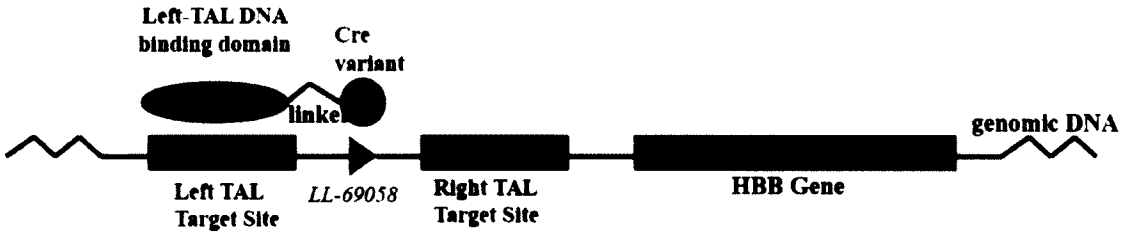


Figure 1.15: TAL DNA-binding domain designed to recognized the left TAL target site present upstream of the LL-69058 target sequence in the human beta globin (HBB) gene. The evolved Cre variants are attached to the TAL DNA binding domain proteins via a linker region.

1.5 Using Evolved Cre and Flp Recombinase to Recognize Genomic LL and FL Target Sequence

The human genome has native recombinase target-like sites about every 5kb [5]. These genomic target-like sequences (LL and FL) are not similar to the loxP and FRT sequences in certain locations (Figure 1.16), but they can be utilized for recombination reactions by introducing or deleting certain DNA bases from the recombinase gene. The modification of these recombinases can be achieved using evolution techniques like DNA shuffling, PCR mutagenesis and error prone PCR. The modified or evolved Cre and Flp recombinases can be used to test if they can recognize genomic LL and FL sequences.

The applications of using site-specific tyrosine recombinases to modify the human genome are very wide spread and are utilized in gene therapy, gene targeting, and curing certain diseases [49]. When the recombinases are fused with TAL DNA-binding domain it helps in further increasing the efficiency of recombinases to bind at desired target sequences. Initially, when recombinase enzymes were discovered, gene knock outs were performed in mice after flanking the gene of interest with loxP target sequence. Using TAL-recombinase fusion will be a useful approach as it would act as a good gene therapy technique for eradication of many diseases. My project focuses on the potential cure for genetic diseases. Native recombination target sequences, loxP-like (LL) and FRT-like (FL) sequences, as discussed above, are utilized to perform integration and replacement recombination events. I believe that by using TAL DNA-binding domain fused to tyrosine recombinase, the efficiency of recombinase will be enhanced and my goal of attempting to find a potential remedy for replacing the mutated beta-globin gene with a healthy gene cassette can be accomplished.

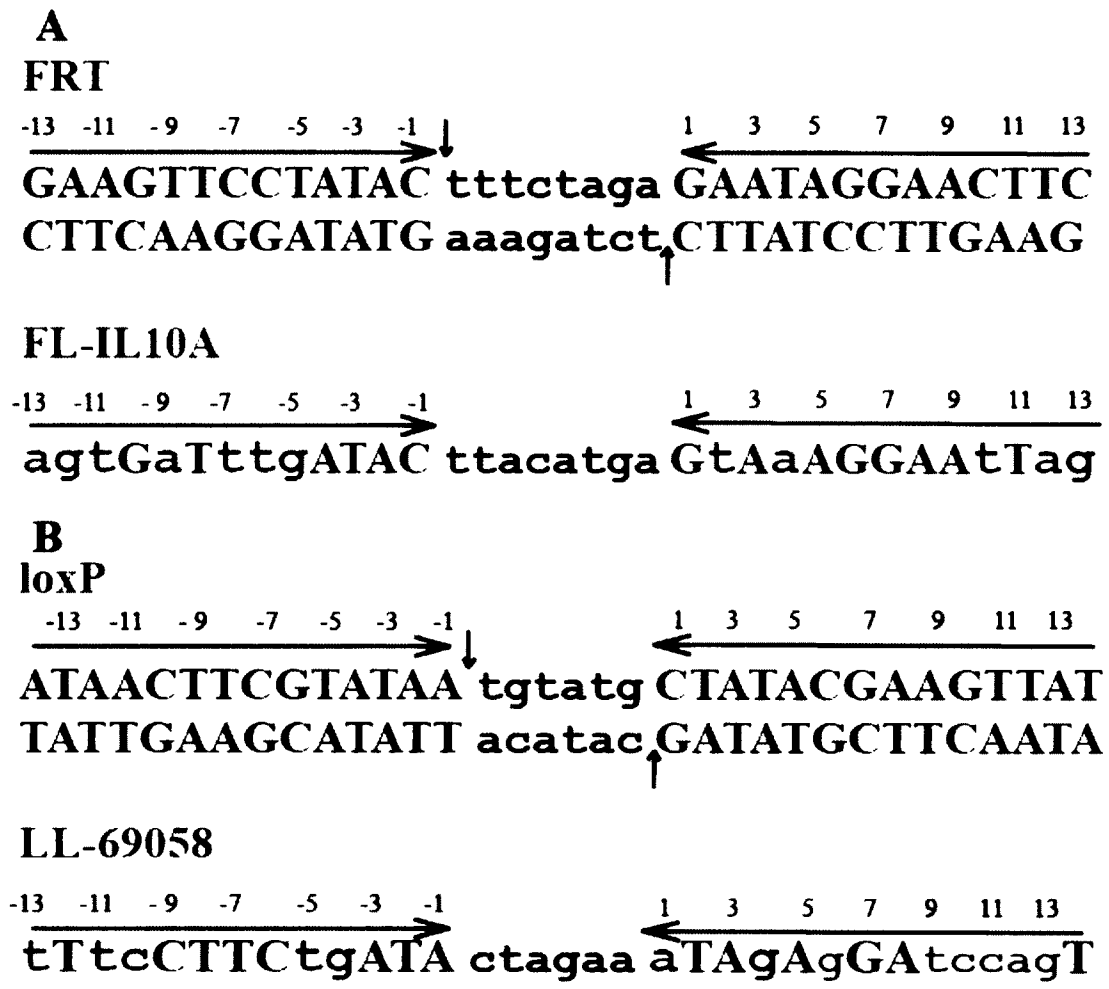


Figure 1.16: Comparison of FRT and FL-IL10A target sequence for evolved Flp recombinase (A) and loxP and LL-69058 target sequence for evolved Cre recombinase (B). The green color in the FL-IL10A and LL-69058 target sequences indicate similar bases to FRT and loxP target sequences, while red color indicates differences in the bases compared to FRT and loxP target sequence.

CHAPTER 2

METHODS

2.1 Objective 1 - Testing *FL-IL10A* Integration Using Evolved Flp Recombinase

2.1.1 Cell Lines and Transfection

Human embryonic kidney HEK 293 cells (ATCC, CRL-1573) were used as model mammalian cells and native genomic platform cells. The cells were propagated in DMEM medium. Cell transfections were performed using Polyfect (Qiagen).

2.1.2 Vectors

The reporter vectors for the integration experiments were based on the pcTD vector that is described in Anderson *et al* [50]. The platform reporter pCMV-IL10A-DsRed-neo and the incoming reporter pIL10A-puro-2A-EGFP, which were used in the integration experiments in HEK 293 cells, were generated as follows. The DsRed-neo^R cassette from pIRES2-DsRed-Express (Clontech) was PCR amplified and subcloned into pcTD downstream from the CMV promoter to obtain pCMV-IL10A-DsRed-neo. *FL-IL10A* upstream of the DsRed-neo^R cassette was introduced during the amplification of the cassette. The puro^R gene from pCAGGS-Flpe-puro (Addgene plasmid 20733) and the EGFP gene from pIRES2-EGFP (Clontech) were PCR-assembled into one gene through the connecting 2A peptide and cloned into pcTD downstream from the CMV promoter to obtain pCMV-IL10A-puro-2A-EGFP. *FRT/FL-IL10A* upstream of the puro-2A-EGFP

gene was introduced during its amplification. The CMV promoter in this plasmid was then deleted to obtain pIL10A-puro-2A-EGFP.

For the experiments with the intact HEK 293 cells, the pCMV-IL10A-puro-2A-EGFP reporter was modified by cloning the PGK promoter from pDIRE[25] (Addgene plasmid 26745) upstream of the promoterless hygro^R gene. To express Flp variants in mammalian cells, the respective genes were subcloned into pOG44 (Invitrogen) in place of the Flp-F70L variant gene as described in Anderson *et al* [50] (Figure 2.1).

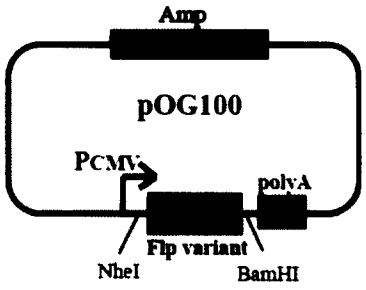


Figure 2.1: Expression vector pOG44. The Flp variant is cloned between NheI and BamHI restriction sites.

2.1.3 Construction of Platform Reporter Cell Lines

HEK 293-IL10A-DsRed-neo cell line was constructed by transfecting HEK 293 cells with pCMV-IL10A-DsRed-neo, propagating them in medium supplemented with neomycin (800 mg/l) and selecting for the red colonies. Several such colonies were expanded and tested for singly integrated platform reporter vector by Southern blotting; the positive colonies were used in the integration experiments.

2.1.4 Integration Experiments

The analysis of the integration of the reporters into the respective genomic platform reporters was performed in the same general way for all three reporter pairs. The integration reaction was done in 6-well plates. The incoming reporters (0.5 μg) were co-

transfected into the respective platform reporter cells with the vectors that express Flp variants analyzed. Forty eight hours post transfection, 1/6 of the cells were transferred into 100 mm plates, the cells were allowed to become confluent, and the number of the green colonies was counted. Several integration positive colonies from these plates were expanded and analyzed.

The analysis of the integration of modified incoming reporter with the hygromycin gene was tested in the native genomic platform. The transfection was performed in a 24-well plate. Forty eight hours post transfection, 1/4 of the cells were transferred into a 6-well plate. The cells were allowed to settle in the plate for 24 hours and then the medium was changed to hygromycin containing medium (500 μg/ml). The hygromycin medium was changed every two days in ten consecutive days until colonies were formed. The green and non-colored colonies were counted and the positive green colonies were expanded and analyzed.

2.1.5 Southern Blotting

Approximately 15 μg of genomic DNA isolated from HEK 293 cells were digested by SacI overnight followed by incubation with RNase A at 37° C for 1 hr. The digested DNA was then heat treated at 65°C for 5 min followed by cooling on ice for 5 min. The DNA fragments were separated on 0.7% TAE agarose (molecular biology grade, BioRad). The gel was then treated with the depurination solution (0.25M HCl) for 15 min followed by the alkaline denaturing solution (two times for 30 min). The DNA fragments isolated from the gel were transferred to the positively charged nylon membrane (Amersham Hybond-N^+, GE Healthcare) in the neutral transfer buffer (20X SSC, G-Biosciences). The membrane was then baked at 80°C for 40 min followed by UV

treatment (5,000 microjoules). Next, the membrane was pre-hybridized in the ULTRAhyb hybridization buffer (Ambion) at 55°C for 30 min and then hybridized with the biotinylated EGFP probe (~ 1 μg) at 55°C overnight. The DNA bands were visualized using the Phototope-Star Detection Kit (New England Biolabs) according to the manufacturer's protocol.

2.1.6 Other Methods

Plasmid DNA was isolated using GeneJET Plasmid Miniprep Kit (Thermo). Amplification of the DNA fragments used for cloning was performed using Pfu-Ultra polymerase (Agilent Technologies). PCR analysis of the mammalian genomic DNA was performed using Taq polymerase (New England Biolabs). Genomic DNA from cultured mammalian cells was isolated using GeneJET Genomic DNA Purification Kit (Thermo). General genetic engineering experiments were performed as described in Sambrook and Russell [51].

2.2 Objective 2 - Evolution and Testing of Cre Recombinase to Recognize *LL-69058* Target Sequence

2.2.1 Bacterial Experiments

The *E.coli* strain NEB 10-beta [araD139Δ(ara-leu)7697 fhuA lacX74 galK (ϕ80Δ(lacZ)M15) mcrA galU recA1 endA1 nupG rpsL (StrR) Δ(mrr-hsdRMS-mcrBC)] from New England Biolabs was used in all bacterial experiments. The deletion assays in *E. coli* were performed essentially as described in Voziyanov *et al* [10].

2.2.2 Cell Lines and Transfection

Chinese Hamster Ovary (CHO-TD1) cells were used as model mammalian cells. The CHO-TD1 cell line was constructed as described by Anderson *et al* [50] and are

derivatives from CHO Flp-In cells (Invitrogen). Human embryonic kidney HEK 293 (ATCC, CRL-1573) cells were used as wild type mammalian cells. The CHO-TD1 and HEK 293 cells were propagated in F-12K and EMEM/DMEM medium (Cellgro) respectively. Cell transfections were performed using DNA-In (MTI globoSTEM).

2.2.3 Oligonucleotides - Site-directed Mutagenesis

Mutagenic primers were designed to introduce particular mutations in the iCre recombinase gene. The list of various primers used and their positions are listed in the Table 2.1. The entire plasmid unwinds and was amplified with the forward and reverse primers in the presence of high fidelity polymerase Pfu (Aligent). The linear PCR product was then run on 1.5% agarose gel and treated with DpnI (New England Lab) ensuring degradation of the parent plasmid. The PCR mixture was then transformed through the competent *E.coli* cells and the circular form of the plasmid was regained on plasmid isolation. The plasmids were sequenced to confirm the desired mutation.

Table 2.1: Forward and reverse primer sequences for introducing various site-directed mutations in the Cre recombinase gene.

Targeted Site-directed mutation position for iCre recombinase	Forward Primer	Reverse Primer
Methionine 24	5'-GGATGCCACCTCTGATGAAGTCATGAAG AACCTGATGGACATGTTCAGG-3'	5'-CCTGAACATGTCCATCAGGTTCTTCAT GACTTCATCAGAGGTGGCATCC-3'
Valine 24	5'-GGATGCCACCTCTGATGAAGTCGTGAAG AACCTGATGGACATGTTCAGG-3'	5'-CCTGAACATGTCCATCAGGTTCTTCACG ACTTCATCAGAGGTGGCATCC-3'
Methionine 32	5'GAAGAACCTGATGGACATGTTCATGGACAGGCAGGCCTTCTCTGAACAC-3'	5'-GTGTTCAGAGAAGGCCTGCCTGTCCATGAACATGTCCATCAGGTTCTTC-3'
4044	5'GACAGGCAGGCCTTCTCTGAANNN ACCTGGAAGNNNCTCCTGTCTGTGTG CAGATCCTG-3'	5'-CAGGATCTGCACACAGACAGGAGNNNCT TCCAGGTNNNTTCAGAGAAGGCCTG CCTGTC-3'
24143	5'-CCCCAACAACTACCTGTTCTGCCGGGTC AGA AAGAATGGTGTGGCTGCCCC-3'	5'-GGGGCAGCCACACCATTCTTTCTGACCC GGCAGAACAGGTAGTTGTTGGGG-3'
6972	5'-GGAAATGGTTCCCTGCTGAACCT NNN GATGTG NNN GACTACCTCCTGTACCTG-3'	5'-CAGGTACAGGAGGTAGTCNNNCACATC NNNAGGTTCAGCAGGGAACCATTTCC-3'
262	5'-CCCAACTGTCCACCCGGGCCCTG NNN GGGATCTTTGAGGCCACCCACC-3'	5'-GGTGGGTGGCCTCAAAGATCCCNNNCA GGGCCCGGGTGGACAGTTGGG-3'
A1 (25, 29, 32, 33, 35)	5'-GATGAAGTCAGGAGGAACCTGATGAGGA TGTTCGAACTCAGGAGGGCCTTCTCT GAACAC-3'	5'-GTGTTCAGAGAAGGCCCTCCTGAGTTCGA ACATCCTCATCAGGTTCCTCCTGACT TCATC-3'
A2 (337)	5'-GGACTCTGAGACTGGGGCCATGGTG GAG CTGCTCGAGGATGGGGAC-3'	5'-GTCCCCATCCTCGAGCAGCTCCACCA TGG CCCCAGTCTCAGAGTCC-3'

2.2.4 PCR Mutagenesis (Error prone PCR) and DNA Shuffling

PCR mutagenesis of the wild-type recombinase genes and DNA shuffling of the variant recombinase genes were performed as described in Bolusani et al [52]. In brief, mutagenic PCR was performed using Taq polymerase in the Mg-free buffer supplemented with 0.25 mM MnCl2, 0.5 mM MgCl2, 200 μM dNTP, 100 nM of each primer, and ~100 ng of DNA template. After amplification, the PCR product was digested with SacI and SphI (iCre recombinase variant gene libraries) and was ligated to pBAD33 [52] digested with the respective enzymes. To perform DNA shuffling, genes of interest were amplified using Taq polymerase under non-mutagenic conditions. The PCR products were mixed in an equimolar ratio and fragmented with DNase I for 5 min on ice in the buffer containing 50 mM Tris–HCl (pH 7.4) and 10 mM MnCl2. The resultant DNA fragments were reassembled in the two-step PCR reaction: first, using Pfu-Ultra polymerase without the addition of specific primers and then using Taq polymerase with specific primers that anneal outside the coding region of the genes. The TAL-Flp libraries were constructed by my lab colleague, and I tested the functionality of the Flp libraries for the *FL-71362* target sequence in the bacterial and mammalian cells.

2.2.5 Bacterial Screening Assay

The libraries obtained from site-directed or PCR mutagenesis and DNA shuffling were required to be screened through the bacterial system to detect potential activity by the recombinases. The Flp and Cre libraries were cloned in the pBAD33 expression vector between SacI and SphI restriction sites. These libraries were tested against a reporter vector p24 which consists of the hybrid FRT/FL-71362 and loxP/LL-69058 sites flanking the LacZ gene (Figure 2.2).

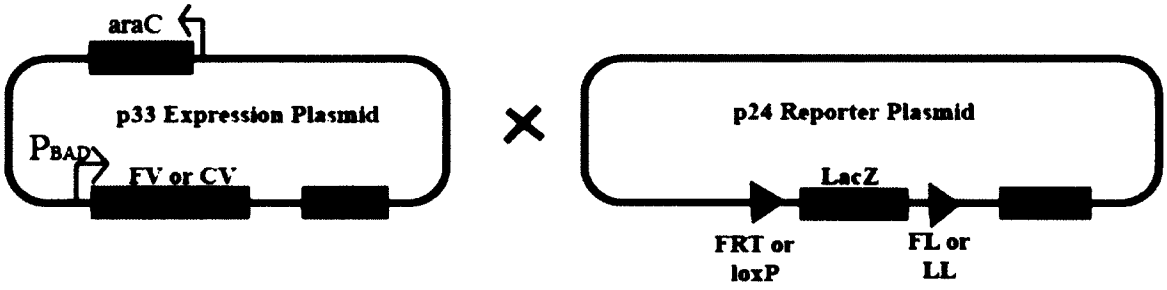


Figure 2.2: Expression p33 and reporter p24 plasmids for bacterial screening assay.

Active libraries recognize the sites and bind to the target site in p24 resulting in the excision of the LacZ gene. Both the expression and the reporter vectors were transformed in the *E.coli* bacterial strain. These strains were initially made competent using the p24 reporter plasmid and then used to transform the expression plasmid for monitoring activity and screening.

2.2.6 Vectors

The reporter vectors for the replacement experiments were based on the pcTD vector that is described in Anderson et al [50]. The platform reporter p1372 (EF1α-LL-69058-neo-STOP-FL-71362-DsRed) and the incoming reporter p1345 (69loxP/LL-EGFP-CMV-71FRT/FL), were used in the replacement experiments in CHO TD-In cells. The platform p1372 reporter was created by lab colleague. The incoming p1345 reporter was generated by subcloning the EGFP gene from pIRES2-EGFP (Clontech) downstream of *loxP/LL-69058* sequence followed by CMV promoter upstream of *FRT/FL-71362* sequence, into pcTD.

The platform p1372 reporter and the incoming p1345 reporter, which were used in the replacement experiments in CHO TD-In cells, were constructed by modifying the respective reporters that were used to analyze the dual RMCE reaction by the Flp/Cre recombinase pair. For this, the neo^R-STOP cassette of the platform reporter was flanked

by the *LL-69058 - FL-71362* sequence while the EGFP-CMV cassette of the incoming reporter was flanked by the *loxP/LL-69058 - FRT/FL-71362* sequence.

For the experiments in the intact HEK 293 cells, the p979 (69loxP/LL-CMV-EGFP -71FRT/FL-EF1α-puro-2A-DsRed) incoming reporter was modified by cloning the inverted PGK promoter from pDIRE (Addgene plasmid 26745) upstream of the promoterless hygro^R and ployA (from p6350) gene.

2.2.7 Construction of Platform Reporter CHO Cell Line

To construct the CHO TD-(EF1α-LL-69058-neo-STOP-FL-71362-DsRed) cell line, CHO TD-In cells were co-transfected, in 6-well plates, with the p1372(EF1α-LL-69058-neo-STOP-FL-71362-DsRed) reporter (0.2 μg) and the pOG-TD1-40 vector (2 μg), which express the TD1-40 variant of TD recombinase. Forty eight hours post-transfection, 1/6 of the cells were transferred into a 100 mm plate containing F12-K medium supplemented with hygromycin (550 mg/l). About 10 days later, several hygromycin resistant, colorless colonies were transferred into a 96-well plate and their sensitivity to zeocin was tested; zeocin sensitive colonies were expanded and used in the replacement experiments.

2.2.8 Construction of TAL Recombinases

2.2.8.1 Golden Gate TAL Construction. The DNA-binding domain of TAL is built by the assembly of RVD's for a particular sequence. The sequence upstream and downstream of the *LL-69058* target sequence was used as the DNA-binding domain for constructing left and right TAL respectively. The protocol for constructing the DNA-binding domain involves the following steps illustrated in Figure 2.3 and 2.4.

>direct sequence

AGAGAAAGCATTTAAGAGAATAAAGCAATGGAAATAAGAAATTTGTAAAT

TTCCTTCTGATAACTAGAAA**TAGAGGATCCAGT**TTCTTTT**GGTTAACCTA**

AATTTTATTTCATTTTATTGTTTTATTTTATTTTATTTTATTTTATTTTGTGTA

ATCGTAGTTTCAGAGTGT

>inverse sequence

ACTGGATCCTCTATTTCTAG**TTATCAGAAGGAAA**TTT**ACAAATTT**

CTTATTTCCATTGCTTTATTCTCTTAAATGCTTTCTCTATTATTGCTAAATA

AATAGAGATCTCTCACTTTTTCTACCTGTCTCAACCCTCATCAGGTACTTG

Figure 2.3: The direct and inverse sequences of the human beta-globin gene near the LL-69058. The target sequence is shown in green. The DNA binding domain for TAL downstream direct sequence (right TAL) and upstream inverse sequence (left TAL) are shown in blue.

RVD Assembly For Right TAL DNA-binding Domain

A. TAL DNA-binding domain (Right) - 15 nucleotide sequence

GGTTAACCTAAATTT

G	G	T	T	A	A	C	C	T	A	A	A	T	T	T
NN	NN	NG	NG	NI	NI	HD	HD	NG	NI	NI	NI	NG	NG	NG
1	2	3	4	5	6	7	8	9	10	1	2	3	4	LR

pFUSA vector pFUSB4 vector

B. TAL DNA-binding domain (Left) - 15 nucleotide sequence

ACAAATTTCTTATTT

A	C	A	A	A	T	T	T	C	T	T	A	T	T	T
NI	HD	NI	NI	NI	NG	NG	NG	HD	NG	NG	NI	NG	NG	NG
1	2	3	4	5	6	7	8	9	10	1	2	3	4	LR

pFUSA vector pFUSB4 vector

Figure 2.4 The construction of TAL DNA binding domain using single RVD's to clone into pFUSA and pFUSB vectors for the sequence located (downstream) right (A) and (upstream) left (B) of the LL-69058 target sequence.

A sequence consisting of 18,15 and 12 nucleotides located upstream and downstream of the LL-69058 target sequence was selected to construct the TAL DNA-binding domain. Figures 2.3 and 2.4 indicate the upstream and downstream

selected sequence and the construction of the binding domain for 15 nucleotides with single RVD's.

I. The first step was to assemble 10 nucleotides of the sequence with their respective RVD's (Figure 2.4). The assembly of first ten RVD had to be cloned in the pFUSA vectors. The remaining nucleotides, 11-14, were assembled and cloned in another vector called pFUSB4. The last residue (LR) of the sequence was kept as a single RVD and was not used in the first round of assembly. A 20µl PCR mixture consisting of pFUSA vector backbone, the first 10 nucleotide RVD, ligase, BsaI, ligase buffer, DTP and the remaining volume made up with water was prepared. A second 20 µl PCR mixture consisting of pFUSB4 vector backbone, remaining 4 nucleotide RVD's, ligase, BsaI, ligase buffer, DTP and the remaining volume made up with water was prepared. The PCR mix was incubated in a PCR machine for 20 cycles at 37^0C for 15 minutes,16^0C for 10 minutes and then incubated at 50^0C - 15 minutes and 16^0C - 15 minutes and then heated again at 80^0C - 20 minutes. After PCR, the mixture was passed through *E.coli* via transformation and plated on spectinomycin X-gal media plates. The pFUS vectors consist of LacZ appear blue if the RVD's do not assemble in them, otherwise on RVD assembly the colonies appear white.

II. The white colonies were picked, the plasmid was isolated and sequenced so as to use the correctly assembled RVD's in pFUSA and pFUSB4 to further clone into the final pOG-TALE core vectors.

III. The pFUSA, pFUSB4 plasmid and the last residue (LR) were then assembled in the pOG-TAL core. The TAL core was amplified from the pTAL3 vector using different N and C terminal primers. The N terminal primers were designed to bind at the

28, 63 and 75 nucleotide position and the C terminal primers are designed to bind at the 117, 131 and 152 position in the pTAL3 plasmid.

The PCR mixture for final assembly consisted of pOG-TALE core vector, pFUSA, pFUSB4 vector, last repeat (LR), ligase, Esp3I, ligase buffer, DTP and the remaining volume made up with water. The PCR mixture was incubated for 20 cycles at 37^0C - 5 minutes and 16^0C - 10 minutes and then heated to 37^0C - 15 minutes, 16^0C minutes and heated again at 80^0C - 15 minutes. After PCR, the mixture was passed through *E.coli* via transformation and plated on Ampicilin X-gal plates. The pOG-TAL core has LacZ gene and if the pFUS vectors and last repeat were inserted in the pOG-TAL core, the colonies appeared white in color.

IV. The white colonies were picked and the plasmids were isolated. The plasmids were sent for sequencing and the correct plasmids were ready for being used to attach the Cre recombinase gene.

V. The Cre recombinase was amplified from the various variants mentioned in the tables above using Cre specific primers with additional sequence at the N terminal. The additional sequence is called the linker region which gives flexibility to the recombinase in order to find LL-69058 target sequence. The PCR product was digested with respective enzymes and cloned in the correctly assembled pOG-TAL core vectors.

2.2.8.2 Optimization of Linker Region Between the TAL Domain and Recombinase. The linker region that connects the TAL binding domain and recombinase were adjusted using different lengths of linkers called half, one and full (one and half) spacer sequence. The sequence of the linker regions are listed in Table 2.2. The linkers were added to the iCre variant using primers. The iCre gene along with different linker lengths were cloned in pOG vector between

NheI and BsrGI restriction sites and in pBAD33 vector between SacI and SphI restriction sites.

Table 2.2: Linker region sequences for half, one and half and full length. The underlined sequence indicate the common linker on which the one and half and full linker are built (Bold).

Half linker	5'-GTCCCCATCCTCGAGCAGCCTCACC-3'
One linker	**5'GCCGCTTCCGCCTGAGCC**GTCCCCATCCTCGAGCAGCCTCACC-3'
Full linker	**5'TGATGTGCCTGAGCCGCCGCTTCCGCCTGATCCTCCGCTTCC** G TCCCCATCCTCGAGCAGCCTCACC-3'

2.2.9 Dual Recombinase-Mediated Cassette Exchange (RMCE) Experiment in CHO Model Cells

The cassette exchange reaction between the incoming reporter p1345 (69loxP/LL-EGFP-CMV-71FRT/FL) and the platform reporter p1372 (EF1α-LL-69058-neo-STOP-FL-71362-DsRed) (integrated into CHO TD-In) was performed similar to the integration experiments. The incoming reporter (0.36 μg) and the expression vectors pOG-TAL-Cre (0.003 μg) and pOG-TAL-Flp (0.02 μg) were co-transfected into CHO TD-1(EF1α-LL-69058-neo-STOP-FL-71362-DsRed) model cell line. Forty eight hours post transfection, 1/16 of the cells were transferred into 6-well plates, the cells were allowed to become confluent, and the number of the green, red, and green/red colonies was counted. Several green/red colonies from these plates were expanded and analyzed.

2.2.10 Optimization of Right and Left pOG-TAL-recombinases for Efficient Binding in CHO Model Cells

The pOG-TAL-Cre variants were paired based on different lengths of the TALE binding domain. Various concentrations of the left pOG-TAL-Cre recombinase were paired with the right pOG-TAL-Cre recombinase to analyze maximum attainment of green/red colonies in the CHO model cells.

2.2.11 Integration of pOG-TAL-Flp and pOG-TAL-Cre in Wild Type HEK 293

The analysis of the integration of the incoming reporters 979 (69loxP/LL-CMV-EGFP-inv PGK-hygro-pA-71FRT/FL-EF1α-DsRed) into the respective genomic platform reporters was performed in the same general way for all three reporter pairs (Left/right TALE recombinases and incoming 979 reporter). The integration reaction was done in 24-well plates. The incoming reporters (0.45 μg) were co-transfected into the genomic platform with the vectors that express pOG-TAL-Flp (0.2 μg) variants (0.02 μg in the replacement experiments in CHO model cells) and pOG-TAL-Cre (0.3 μg) (0.003 μg in the replacement experiment in CHO model cells) in separate wells. Forty eight hours post transfection, 1/4 of the cells were transferred into 6-well plates, the cells were allowed to settle in the well for 24 hours and then the medium was changed to hygromycin (500 μg/ml). The hygromycin medium was changed every 2 days for 10 consecutive days until colonies were formed. The number of the green, red, green/red and non colored colonies was counted. Several integration positive colonies for pOG-TAL-Flp (just red not green) and pOG-TAL-Cre (just green not red) from these plates were expanded and analyzed.

CHAPTER 3

RESULTS

3.1 Results for Objective 1: Testing *FL-IL10A* Integration Using Evolved Flp Recombinase[53]

3.1.1 FV7 is Active in Human Cells

Having established that FV7 is active in the integration and replacement reactions in CHO cells, we next tested the ability of FV7 to utilize the *FL-IL10A* sequence as a substrate in human cells in a model setting. For this, we integrated an *FL-IL10A* bearing platform reporter into human embryonic kidney 293 cells, HEK 293. In the integrated platform reporter, *FL-IL10A* is located between the CMV promoter and the DsRed gene (Figure 3.1A). In the incoming reporter, the hybrid *FRT/FL-IL10A* sequence is located upstream of the promoterless puro-2A-EGFP gene. Upon integration of the incoming reporter into the platform reporter, the expression of the puro-2A-EGFP gene is activated so the cells become green and resistant to puromycin.

To test what fraction of the green colonies represent the correct integration of the incoming reporter into the platform reporter, we subjected ten green colonies to PCR and Southern blot analyses. All ten colonies were positive for the integration specific PCR bands, (Figure 3.1D), the identity of which was confirmed by sequencing. The Southern blot analysis showed that in all but one colony the incoming reporter was singly integrated into the platform reporter (Figure 3.1).

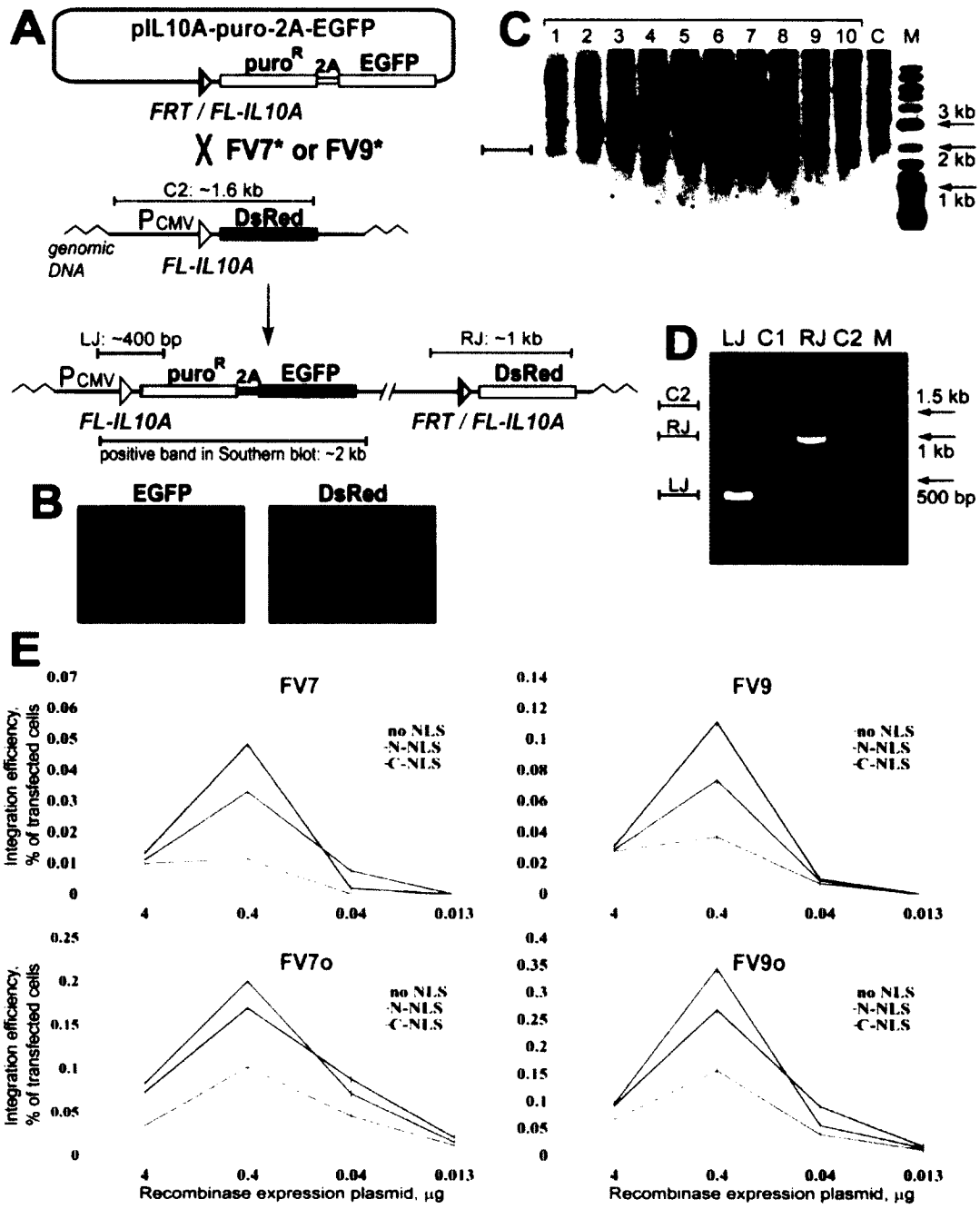


Figure 3.1: Flp variants FV7, FV9 and their derivatives can integrate a reporter into the FL-IL10A site pre-integrated into the human genome. (A) Schematic outline of the integration assay. The horizontal blue and red bars in panels (A) and (D) schematically represent the PCR products at the left junction (LJ) and the right junction (RJ) of the integration product and the control PCR product of the platform reporter (C2). (B) Expanded green, puromycin-resistant cells do not express DsRed. (C) Southern blot analysis of the expanded green, puromycin-resistant colonies. The green bar in panels (A) and (C) shows the location of the

integration-specific DNA fragment. M, biotinylated DNA marker. (D) PCR analyses of a typical green, puromycin-resistant colony. LJ and C1, the PCR analysis of the green cells and the control platform cells using the primers that anneal on the CMV promoter and the puroR gene. RJ and C2, the PCR analysis of the green cells and the control platform cells using the primers that anneal on the vector backbone and the DsRed gene. M, DNA marker. (E) The integration activity of the Flp variants FV7, FV9, FV7o, and FV9o as a function of the location of the NLS sequence and the input of the corresponding recombinase expression vectors at transfection. The results are represented by the data points connected by a line of the respective color. The data points show the mean value of five experiments; the error bars indicate standard deviation.

3.1.2 The Activity of FV7 in Human Cells can be Enhanced by 'Humanizing' the FV7 Gene and by Adding NLS to its Termini

The activity of FV7 in human cells was low, so we explored two approaches to increase it: (1) codon optimization of the FV7 gene to maximize its expression in human cells; and (2) addition of the SV40 nuclear localization sequence, NLS, to its N- or C-termini to facilitate the crossing of FV7 into the nucleus. These modifications were successfully applied to increase the activity of Cre and Flp in mammalian cells [54,55] In the optimization experiments, in addition to FV7, we used a *de novo* evolved Flp variant, FV9, which recombines *FL-IL10A* in the bacterial deletion assay slightly better than FV7. FV9 differs from FV7 at position 166: FV9 has wild-type Lys at this position while FV7 has Glu.

We tested the integration activity of the original and the modified recombinases at the different inputs of the respective expression plasmids. FV7 was able to generate green cells only at the higher inputs: 4 μg and 0.4 μg of the expression plasmids per transfection (Fig. 3.1E, top panel). FV9 was more active than FV7 but it was not able to generate green cells at the lowest input of the expression plasmids tested: 0.013 μg (Figure 3.1E, top panel). The 'humanized' versions of FV7 and FV9 – FV7o and FV9o –

had higher integration activity than FV7 and FV9 and were able to generate green cells even at the lowest input of the expression plasmids (Figure 3.1E, bottom panel). Unlike FV7 and FV9 that had essentially flat activity at the higher inputs of the expression plasmids (4 μg and 0.4 μg), FV7o and FV9o had a pronounced peak of activity at 0.4 μg.

Although the addition of NLS to the N- or C-termini of both non-optimized and optimized recombinase variants increased their integration activity, this modification was not powerful enough to allow the not-optimized FV7 and FV9 to generate green cells at the lowest input of the expression plasmids (Figure 3.1E, top panel). The variants with added NLS followed the FV7o and FV9o pattern and had a pronounced peak of activity at 0.4 μg of the respective expression plasmids added at transfection (Figure 3.1E). For FV7 and FV9, it was the N-terminally placed NLS that helped them to achieve their highest integration activity. In contrast, the 'humanized' FV7o and FV9o benefited more from the C-terminal NLS. It was FV9o with this modification, FV9o-NLS, which showed the highest integration activity of all recombinase variants tested: ~ 0.35% of the transfected cells (Figure 3.1E).

3.1.3 FV9o-NLS Can Target *FL-IL10A* Located in its Native Chromosomal Environment.

Finally, we tested the ability of FV9o-NLS to integrate a reporter into the native *FL-IL10A* sequence in the intact HEK293 cells (Figure 3.2). For this, we modified the targeting vector described in the previous section to include the constitutively expressed hygromycin B phosphotransferase gene under the control of the PGK promoter (Figure 3.2A) to help enrich the cell population that bears the integrated targeting vector.

The integration experiments were performed by co-transfecting the targeting vector and the FV9o-NLS expression vector into HEK293 cells. In the control

experiments performed in parallel, the HEK293 cells were co-transfected with the targeting vector and the empty expression vector. Forty-eight hours post transfection, the cells were expanded in the hygromycin supplemented medium for about a week and the ratio of the green to non-green colonies assessed. Then, the cells were collected and analyzed by PCR and sequencing.

In the control experiments with the empty expression vector, the ratio of the green to non-green colonies was 2.08±0.16. In contrast, the ratio of the green to non-green colonies in the experiments with FV9o-NLS was significantly lower: 0.62±0.23 (Figure 3.2B). The PCR analyses of the genomic DNA isolated from the collected cells showed that the characteristic, integration-specific bands were generated only when the cells were transfected with the FV9o-NLS expression vector but not with the empty one (Figure 3.2C). Sequencing of the integration-specific bands confirmed the targeting of the *FL-IL10A* sequence (Figure 3.2C).

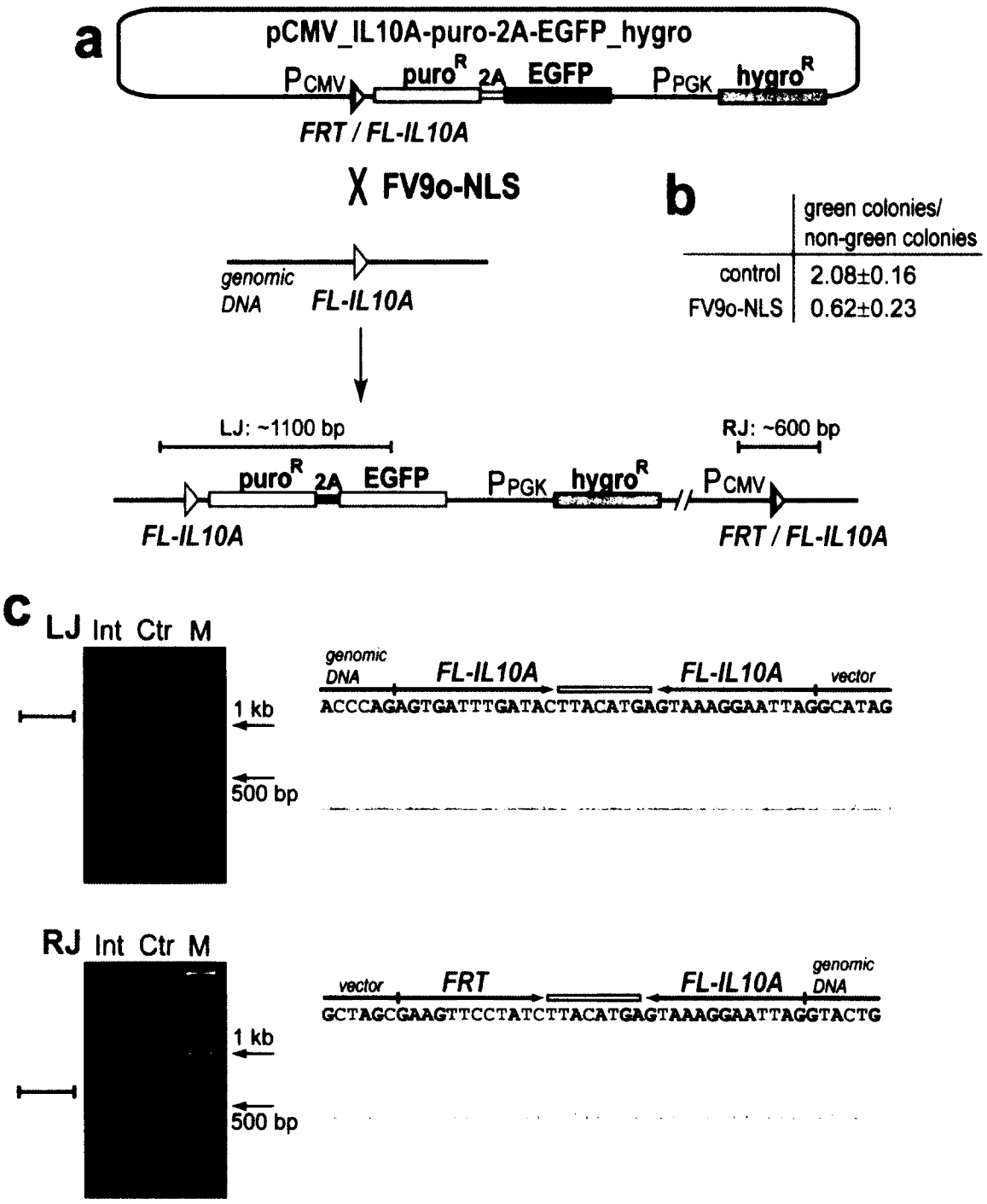


Figure 3.2: FV9o-NLS can integrate a reporter into the native FL-IL10A sequence in the human genome. (A) Schematics of the integration assay. Upon expression of the FV9o-NLS variant, the incoming reporter can integrate into the native FL-IL10A site. Successful integration leads to the loss of the expression of the puro-2A-EGFP gene due to the separation from the promoter. (B) Co-transfection of the reporter and either the empty expression vector or the vector that expresses FV9o-NLS leads to the different ratios of green to non-green colonies. The results show the

mean value of three experiments ± standard deviation. (C) The PCR and sequencing analyses of the pooled transfected and expanded cells. The analyses were performed at the left junction (LJ) and the right junction (RJ) of the integrated reporter vector and the genomic DNA. The blue and green bars in panels (A) and (C) show the location of the integration specific PCR products. Int, PCR analysis of the pooled transfected cells; Ctr, control PCR analysis of the original cells using the same primers as in lane 'Int'; M, DNA marker. The FL-IL10A specific snapshots of the sequencing reads of the integration-positive PCR bands are shown on the right side of panel (C).

The overall goal of the above experiments was to test if the FV7 variant could recognize and integrate gene cassette in the HEK293 cells. FV7 is active in the mammalian cells. Its activity was enhanced when it was fused to C-terminal NLS sequence. Also when FV7 was codon optimized and fused to C-terminal NLS there was 35-fold increase in integration activity.

3.2 Results for Objective 2: Evolution and Testing of Cre Recombinase to Recognize *LL-69058* Target Sequence

3.2.1 TAL-Cre Variant is Active in the Bacterial Deletion System

My first experimental aim was to screen the evolved TAL-Cre variants to monitor efficient recombination activity at the *LL-69058* target sequences in the bacterial deletion assay system. Evolved TAL-Cre variants (containing M24, V24, M32 mutations in Cre gene), Table 3.1 were passed via transformation through p24 reporter containing competent *E.coli* cells. The LacZ gene in the p24 reporter was flanked on either side with either *LL-69058*, *loxP* or hybrid *LL-69058/loxP* target sequences. The TAL-Cre variants were cloned in bacterial pBAD33expression vector. This expression vector containing a pBAD promoter and Arabinose operon have characteristic control on repression, induction and expression of genes cloned into the pBAD vectors [56]. Thus the amount of TAL-Cre gene expression was tightly regulated by the Arabinose operon. The more

the Arabinose in the culture medium, the more active was the Arabinose operon, resulting in higher production of TAL-Cre protein.

The assembled TAL core DNA-binding domain used for fusing the Cre variants was amplified from the pTAL3 vector. Each left and right arm of the TAL DNA-binding domain was assembled with RVD's for 18 nucleotides. The primers amplified the binding core beginning from the 95th position at the N-terminal and from position 117, 131 and 152 at the C-terminal of the pTAL3 vector. Therefore, the Cre variant was cloned in three different assembled TAL core DNA-binding domains named TAL 117+95, TALE 131+95 and TALE 152+95, respectively, for initial bacterial screening. The Cre variants were PCR amplified using half linker region primers in order to fuse the TAL core DNA-binding region and Cre variant.

The different TAL-Cre variants upon being expressed bind to the target sequences flanking the LacZ gene resulting in the deletion of the LacZ gene. If the variants can recognize the target sequence, the colonies on the plate lose their blue color and appear white. Also the p24 reporter on the gel appears to migrate at a lower level compared to the original reporter (Figure 3.3), indicating that the TAL-Cre variant is active in the bacterial system and it deleted the LacZ gene. Various concentrations of arabinose were used to monitor high deletion activity in the bacterial cells.

Figure 3.3: Bacterial screening assay showing an agar plate containing chromogenic substrate X-gal with white and blue colonies (A). Gel image of plasmids isolated from the white and blue colonies (B). The topmost bands is the pBAD33 expression plasmids. The lower band indicates the p24 reporter where if the LacZ is deleted the plasmid migrates lower than the p24 reporter plasmid. Control (Cont) shows presence of two plasmids, containing LacZ (upper band and deleted LacZ (lower band) in the p24 reporter.

The bacterial assay with three different p24 reporter systems indicated some activity at 0.001% and 0.1% arabinose concentrations (Table 3.1). Out of the various TAL-Cre variants passed through the bacterial deletion assay, TAL-Cre variant with a mutation at position 24 position in the Cre gene showed the highest deletion activity. The mutation at position 24 causes a change of amino acid from Arginine to Methionine.

It was observed that the bacterial deletion activity was highest with the pBAD33-TAL-117+95-Half linker-Cre-M24 variant. The pBAD33-TAL-117+95-Half linker-Cre-M24 variant showed 0.75% deletion activity (21 white colonies in 2800 colonies) at 0.01% arabinose on genomic LL-69058 target sequence as compared to 0.7% (20 colonies in 2800 colonies) deletion activity at 0.01% arabinose on wild type loxP target

sequence. Therefore, the pBAD33-TAL-117+95-Half linker-Cre-M24 variant could recognize the genomic target sequence as efficiently at it could recognize the wild type target sequence. The Cre variants fused to TAL 131+95 and 152+95 showed no deletion activity in the bacterial assay. In the control experiments, plain Cre-M24 variant (without fusion to TAL core DNA-binding domain) was passed through the bacterial screening assay to test the deletion efficiency. There was no deletion activity observed and the results were unsuccessful.

Table 3.1: The target sequences used in p24 reporter flanking the LacZ gene and the deletion activity observed under 0.001% and 0.01% of arabinose concentration. The red highlight indicates the only activity observed in the bacterial assay.

Target sequence in p24 reporter flanking the LacZ gene	Arabinose concentration	Deletion activity in bacterial cells (2800 colonies/big plate)
Genomic *LL-69058* TTTCCTTCTGATAA **CTAGAA** ATAGAGGATCCAGT	0.001%	0% white colonies
	0.01%	0.75% white colonies, no semi-blue colonies.
***loxP -loxP* (Full)** ATAACTTCGTATAG **CTAGAA** TTATACGAAGTTAT	0.001%	Just 2 white colonies in 2800 colonies. (negligible activity)
	0.01%	0.7% white colonies, 30% semi-blue colonies.
***Genomic-loxP* (Half)** TTTCCTTCTGATAA **CTAGAA** TTATACGAAGTTAT	0.001%	0% white colonies
	0.01%	Just 1 white colony in 2800 colonies. (negligible activity)

Though the deletion activity with pBAD33-TAL-117+95-Half linker-Cre-M24 was low in the bacterial system, it still showed significant activity to test the TAL-Cre variant in the mammalian cells. The active pBAD33-TAL-117+95-Half linkerCre-M24 variant was then transferred into pOG expression vector to test the replacement activity in

mammalian cells. Furthermore, in the next experiments, pOG-TAL-117+95-Half linker-Cre-M24 variant will be addressed as TAL-Cre-M24.

3.2.2 Selection Of Active Cre Variants Obtained Through DNA Shuffling

As the TAL-Cre-M24 variant showed some activity in the bacterial assay, new Cre libraries were built keeping Cre-M24 as the base template. Table 3.2 below shows various random mutations introduced in the Cre-M24 gene during DNA-shuffling and error prone PCR protocols. The Cre libraries were directly cloned in the mammalian expression vectors to test the replacement efficiency in the model CHO cells. The Cre variants were also fused to the TALE-core DNA-binding domain to compare the activity between Cre variants with and without the TAL-core DNA-binding domain. The experiments with Cre variants not fused to the TAL-core-DNA-binding domain were unsuccessful and did not give any green/red colonies in the replacement experiments. But the Cre variants fused to TAL-core DNA-binding domain had few green/red colonies, but on expansion of these colonies the green color faded leaving just red colonies. These results are unexpected and thus the Cre variants cannot be used for the replacement experiments (Table 3.2).

Table 3.2: Random mutations introduced in Cre recombinase gene using error prone PCR or DNA shuffling methods.

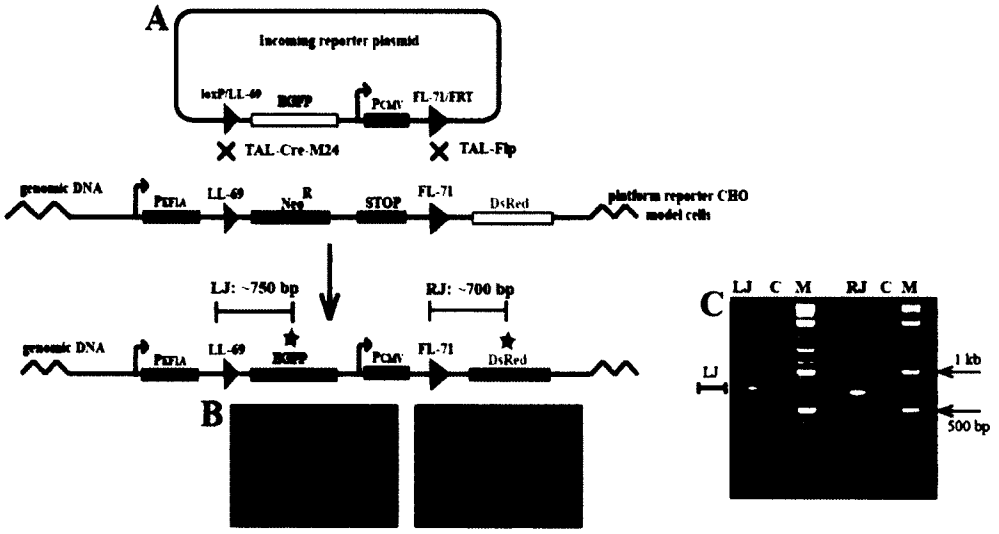
Variant No.	Amino Acid (AA)Change	AA Position No.	Protein Change	Protein Position No.
iCreM24-262-**1**	1. GAA to AAT	1. 808	E to N (Glu to Asn)	262
iCreM24-262-**14**	1. GAA to AAA 2. GAC to AAC 3. GAA to AGC	1. 223 2. 241 3. 880	1. E to K (Glu to Lys) 2. D to N (Asp to Asn) 3. E to S (Glu to Ser)	67 73 262
iCreM24-262-**21**	1. GAA to AGC	1. 880	E to S (Glu to Ser)	262
iCre ScSh **9-1**	1. AGG to ATG 2. AAC to GAC 3. GAT to AAT 4. GTG to GCG 5. GAA to GGA	1. 95 2. 199 3. 589 4. 650 5. 809	R to M (Arg to Met) N to D (Asp to Asp) D to N (Asp to Asn) V to A (Val to Ala) E to G (Glu to Gly)	24 59 189 209 262
iCre **6-2**	Loss of two nucleotides-CT	1036	Loss of Leu codon	337

3.2.3 TAL-Cre M24 Can Perform Efficient Replacement of Gene Cassette in CHO Model Cells

To test if a target-specific TAL-Cre-M24 variant can be combined with another site-specific recombinase to perform dual recombinase-mediated cassette exchange (dual RMCE) in mammalian cells, we paired TAL-Cre-M24 with TAL-Flp in a model setting similar to what we used to analyze the replacement activity of the Flp/HK022 Int recombinase pair [29]. The TAL-Flp variant was constructed and evolved by another member of the lab and I tested and optimized the variant in replacement experiments. For this, the respective cognate target sequences were introduced for Cre and Flp in the

incoming and platform reporters of the replacement system (Figure 3.4) and the platform reporter was then integrated into genome of CHO cells.

A schematic of the assay to test the replacement activity of the TAL-Cre-M24/TAL-Flp recombinase pair is shown in Figure 3.4(A). In the integrated platform reporter, *LL-69058* target sequence is located between the EF1α promoter and the neo^R gene while *FL-71362* target sequence is located upstream of the promoterless DsRed gene. The incoming reporter bears the hybrid *loxP/LL-69058* site upstream of the promoterless EGFP gene whereas *FL-71362/FRT* is positioned downstream of the CMV promoter. If the TAL-CreM24/TAL-Flp recombinase pair is active in the dual RMCE reaction, the neo^R-STOP cassette of the platform reporter will be replaced with the EGFP-CMV cassette of the incoming reporter; this recombination event can be detected by the activation of the expression of both EGFP and DsRed genes in the same cell. The expression of just DsRed will indicate the Flp-dependent integration of the incoming reporter into the platform reporter at the *FL-71362* site while the expression of just EGFP will signify the CreM24-dependent integration of the incoming reporter into the platform reporter at the *LL-69058* site.



D

>Sequencing result of LJ after replacement in CHO model cells

GCTTGGTACTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGGTAAGCTTGGTACCAGAATAAAGCAAT
GGAAATAAGAAATTTGTAAATTTCCTTCTGATAACTAGAAATAGAGGATCCAGTTTCTTTTGGTTAACCTAAA
TTTTATTTCATTCTTCCATCGGCATAGTATATCGGCATAGTATAATACGACAAGGTGAGGAACTAAACCCCTT
TGAAAAACACGATGATAATATGGCCACAACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCC
ATCCTGGTCGAGCTG

>Sequencing result of RJ after replacement in CHO model cells

GGTATGGCATGATAACGCCCGGAAGAGAGTCAATTCAGGGAGATACATTAAGTAACTTAAAAAAAAACTTTAC
ACAGTCTGCCTAGTACATTACTATTTGGAATATATGTGTGCTTATTTGCATATTCATAATCTCCCTACTTTAT
TTTCTGCTAGCACCATGGCCTCCTCCGAGGACGTCATCAAGGAGTTCATG

Figure 3.4: Dual recombinase-mediated cassette exchange, RMCE, mediated by TAL-Cre-M24 and TAL-Flp variants at relative location of the loxP-like sequence LL69 (LL-69058) and FRT-like sequence FL71 (FL-71362) in the CHO model genome. (A) Schematic of the dual RMCE reaction to test the replacement activity of the Cre and Flp variants. The cells, in which the replacement occurred, express both EGFP and DsRed on efficient replacement. (B) The green/red colony formed after the active TAL-Cre-M24 and TAL-Flp variants recognize and replace the gene cassette at the genomic target sequences in model cells. (C) The left and right junction PCR analysis bands confirm the replacement of gene cassette by the correct band size of 750bp at left junction and the band size of 700 bp at the right junction. (D) The sequencing of the PCR bands indicate the genomic LL-69058 sequence at the left junction highlighted in green and the FL-71362 sequence at the right junction highlighted in red. The blue highlighted regions in the left junction sequence are the left and right TALE-core DNA-binding domains for TALE-Cre-M24 variant.

To perform the replacement, we co-transfected the platform reporter cells with the incoming reporter and the vectors that express TAL-Cre-M24 and TAL-Flp. Forty eight hours post-transfection, the cells were expanded and examined for the colonies that express either EGFP or DsRed, or both (Figure 3.4A and B). These colonies and not individual cells that express EGFP and/or DsRed were considered true indicators of recombination events. To confirm that the EGFP and DsRed expression in the green/red colonies reflects the expected reporter replacement, we expanded several such cell groups and analyzed them by PCR (Figure 3.4C) and sequencing. Approximately 1 or 2 green/red colonies were obtained after every transfection from 25 µl of cells transferred in a 6-well plate.

3.2.4 Cre M32 Variant Shows High Recombination Activity when Fused with Left TAL (15 nucleotide), Right TAL (12 nucleotide) and Half Spacer Region

As the number of green/red colonies obtained in the replacement experiment was less with the TAL-Cre-M24/TAL-Flp variants, it was necessary to find an alternative approach to increase the yield of green/red colonies. The TAL-core DNA-binding domains were first altered to observe if there was any significant change in the number of green/red colonies. In the above replacement experiment, the left and right assembled TAL DNA-binding domains were assembled with RVD's for 18 nucleotides. New TAL-core DNA-binding domains were constructed by assembling RVD's for 12 and 15 nucleotides for both left and right arms. The Cre-M24 variant was cloned with a half linker region in the assembled TAL-core DNA-binding domain. Different combinations of left and right arms were optimized by co-transfecting the plasmids in model CHO cells to observe the replacement efficiency.

We observed a 5-fold increase in the replacement activity when the combination of 12 nucleotide RVD (RV12) assembly and 15 nucleotide RVD (LV15) assembly were fused with Cre-M24 variant prior to transfection. On average at least five green/red colonies were observed with every 25 µl transfer of cells into 6-well plate. The transfections were performed by co-transfecting these new TAL RV12 and TAL LV15 Cre-M24 variants with TAL-Flp and the incoming reporter. On expansion and PCR analysis, the PCR produce was of the correct size and the sequencing confirmed the replacement of the gene cassette.

To continue improving efficiency, the linker region fused to the TAL-core DNA-binding domain and Cre-M24 variant was altered. The number of green/red colonies observed with one and full linker spacer region between the TAL-core and recombinase was five times less compared to the original half linker region. This attempt of altering the linker region was unsuccessful and the half linker proved to be the better fit for allowing the Cre variant to recognize the genomic LL-69058 target sequence. Our next aim was to alter the TAL-core length at the N-terminal. The original TAL-Cre-M24 variant had the combination of 117+95. As 117 at the C-terminal was the only position that gave activity in the bacterial assay, no alterations were performed at the C-terminal. New TAL-core plasmids were built with +28 and +63 starting lengths at the N-terminal that were PCR amplified from the pTAL3 vector. The new TAL-core 117+28 and TAL-core 117+63 were assembled with right 12 nucleotide RVD's and left 15 nucleotide RVD's and later fused with Cre-M24 variant with a half spacer region.

The transfections with 117+95, 117+63 and 117+28 TAL-core left and right TAL-core DNA-binding domains fused to Cre-M24 variant were performed in the model

CHO cells along with the incoming reporter and TAL-Flp variant. The replacement efficiency was calculated based on the number of green/red colonies observed and it was noted that 117+95 TAL-core had two times more green/red colonies compared to 117+63 and 117+28 TAL-cores.

Thus the best combination of TAL-core DNA-binding domain fused to Cre-M24 variant was the TAL117+95-half spacer-Cre-M24 with left arm of 15 nucleotides (LV15 and right arm with 12 nucleotide (RV12). In control experiments, Cre-M24 was used in the replacement experiments, to observe if the recombination efficiency without the TAL domain attached. There was no significant activity when plain Cre-M24 variant was added in the experiment. The new version of TAL-Cre-M24 is addressed as TAL-Cre-M24 -LV15/RV12. TAL plays a major role in guiding the Cre-M24 variant to recognize and bind to the genomic target sequence.

3.2.5 Integration in HEK 293 Cells with TAL-Flp Active Variant

The next goal was to test the ability of TAL-Flp variant to integrate 979 (69loxP/LL-CMV-EGFP -71FRT/FL-EF1α-puro-2A-DsRed) reporter into the native *FL-71362* target sequence in the intact HEK293 cells (Figure 3.5). For this, we modified the 979 reporter vector described in the previous section to include the constitutively expressed inverted hygromycin B resistant gene under the control of the PGK promoter (Figure 3.5A) to help enrich the cell population containing the integrated vector.

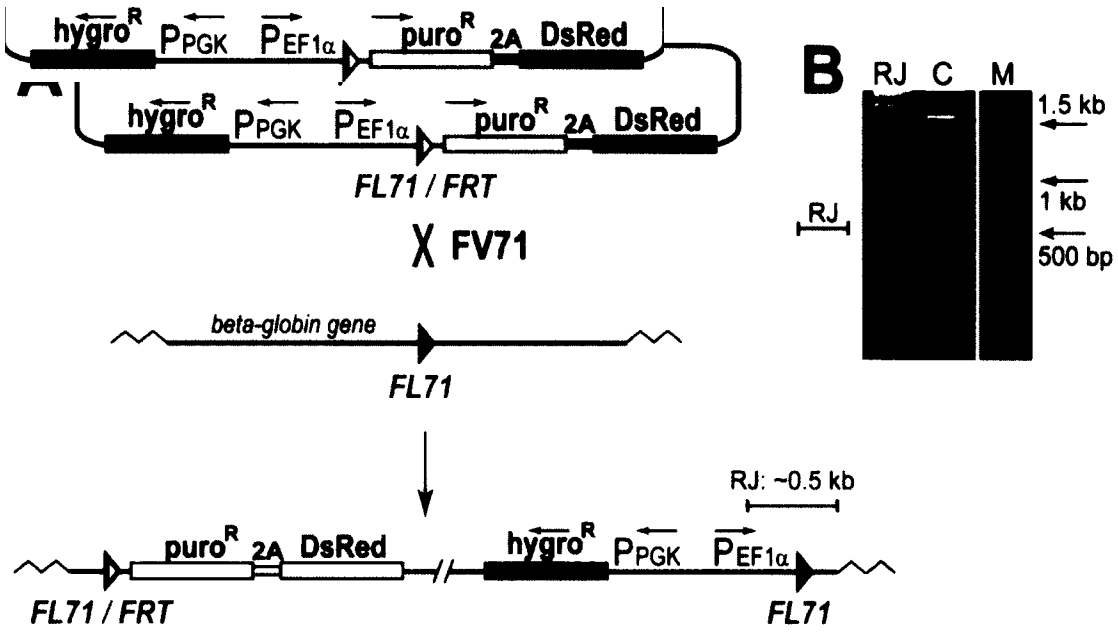


Figure 3.5: Hybrid TAL-Flp variant FV71 specific for the FRT-like sequence FL71 located in the human beta-globin gene can be used to deliver a reporter vector into the gene in intact human cells. (A) Schematic of the integration reaction. (B) The characteristic PCR band is generated in the experimental but not in the control integration reactions. Sequencing of the positive PCR band confirmed its identity.

The integration experiments were performed by co-transfecting the 979 reporter vector and the TAL-Flp expression vector into HEK293 cells. In the control experiments performed in parallel, the HEK293 cells were co-transfected with the 979 reporter vector and the empty pOG expression vector. 48 hours post transfection, the cells were expanded in the hygromycin supplemented medium for about a week and the ratio of the number of green but not red colonies to other colonies (red and green, non colored, and red but not green) was assessed. Then, the cells were collected and analyzed by PCR and sequencing.

In the control experiments with the empty expression vector, the ratio of the green to other colonies was 0.87±1.62. In contrast, the ratio of the green to other colonies in the experiments with TAL-Flp was nearly similar: 0.95±0.23. The PCR analyses of the genomic DNA isolated from the collected cells showed that the characteristic,

integration-specific bands were generated only when the cells were transfected with the TAL-Flp variant expression vector but not with the empty one. The sequencing results of the integration-specific band confirmed the *FL-71362* sequence obtained from the PCR and also confirmed that site-specific integration had occurred.

3.2.6 Integration in HEK 293 Cells with Active TAL-Cre M24- LV15/RV12 Variant

The integration experiment with the TAL-Cre-M24-LV15/RV12 variant was performed three times in the HEK293 cells and the data presented below is from one of the experiments that gave the positive result. The TAL-Cre-M24- LV15/RV12 variant ability to integrate 979 (69loxP/LL-CMV-EGFP -71FRT/FL-EF1α-puro-2A-DsRed) reporter into the native *LL-69058* target sequence was tested in the intact HEK293 cells (Figure 3.6). For this, the modified 979 reporter vector was used which is described in the previous section. The modification was to include the constitutively expressed inverted hygromycin B resistant gene under the control of the PGK promoter (Figure 3.6) to help enrich the cell population containing the integrated vector.

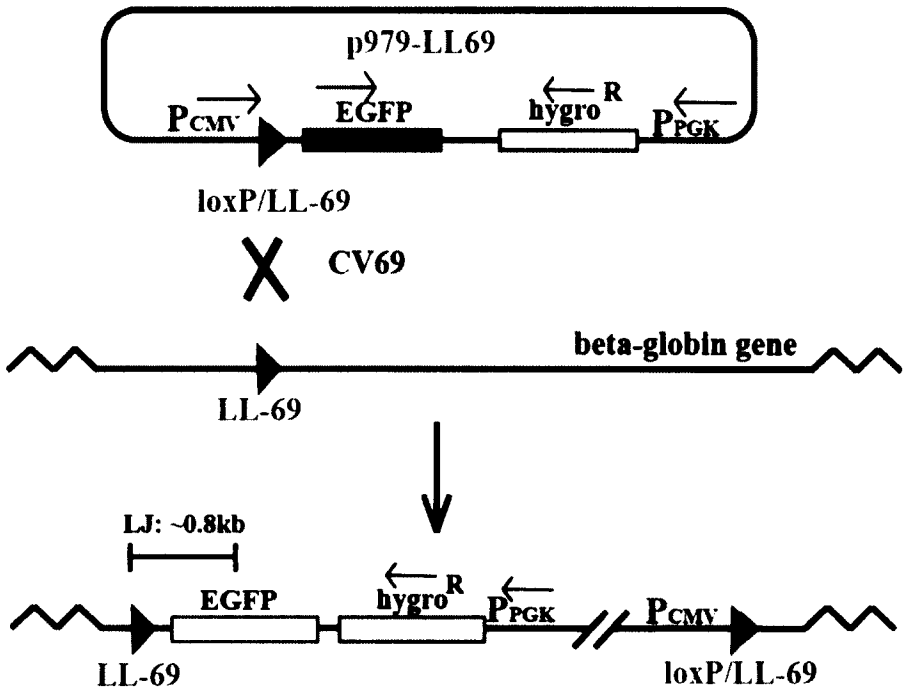


Figure 3.6: Hybrid TAL-Cre-M24 variant specific for the loxP-like sequence LL-69 (abbreviation for LL-69058) located in the human beta-globin gene can be used to deliver a reporter vector into the gene in intact human cells.

The integration experiments were performed by co-transfecting the 979 reporter vector and the TAL-Cre-M24-LV15/RV12 expression vector into HEK293 cells. In the control experiments performed in parallel, the HEK293 cells were co-transfected with the 979 reporter vector and the empty pOG expression vector. Forty eight hours post transfection, the cells were expanded in the hygromycin supplemented medium for about a week and the ratio of the number of red but not green colonies to other colonies (red and green, non-colored, and green but not red) was assessed. Then, the cells were collected and analyzed by PCR and sequencing.

In the control experiments with the empty expression vector, the ratio of the red to all other colonies was 0.142. In contrast, the ratio of the red to all other in the experiments with TAL-Cre-M24-LV15/RV12 was significantly lower: 0.087. The PCR

analyses of the genomic DNA isolated from the collected cells showed that the characteristic, integration-specific bands were generated only when the cells were transfected with the TAL-Cre-M24 variant expression vector but not with the empty one. Sequencing of the integration-specific bands confirmed the targeting of the *LL-69058* target sequence.

The overall goal of the above experiments was to perform replacement reaction in the model CHO and HEK293 mammalian cells with functional Cre variant that was able to recognize the genomic LL-69058 target sequence. Variant TAL-Cre-M24-RV12/LV15 proved to be an active variant that could recognize the genomic site and efficiently replace and integrate gene cassette into the genome.

CHAPTER 4

DISCUSSION

The power of the tyrosine site-specific recombinases to edit genomes is widely realized but only in the approaches that employ the pre-introduced cognate wild-type (or essentially wild-type) recombination targets. The approaches that would utilize genomic target-like sequences in their native chromosomal environment could add a valuable tool to the genome engineering toolbox but are yet to be developed. The work presented in this dissertation shows that the field can be advanced if genome manipulation methods are developed around target-specific recombinase variants that are evolved to be active in bacteria and further modified to optimize their expression level and their delivery into the nucleus in the cells of interest.

The performance of FV7 in human HEK 293 cells did not mirror its performance in hamster CHO cells: the integration efficiency of FV7 in CHO cells was ~0.1% of the transfected cells whereas it was only ~0.01 % in HEK 293 cells. To improve the integration activity of FV7 in human cells, we optimized the codons of the FV7 gene to maximize its expression in human cells and/or added the nuclear localization sequence to the variant to enhance its ability to get into nucleus. The use of the *de novo* evolved Flp variant, FV9, which recombines *FL-IL10A* in *E. coli* slightly better than FV7, was also helpful in increasing the efficiency of *FL-IL10A* targeting.

The effect of the codon optimization on the integration activity of FV7 and FV9 was quite positive allowing the optimized variants, FV7o and FV9o, to have integration activity over almost three-fold higher versus one-fold higher for FV7 and two-fold higher for FV9 (Figure 3.1E). The activity increase was not uniform over the range of the expression plasmid inputs: the peaks of the integration activity were detected at the moderate input of the recombinase expression vectors. The codon optimization effect was more pronounced for FV7: under optimal conditions, FV7o and FV9o were about 10 fold and 5 fold more active than FV7 and FV9, respectively (Figure 3.1E).

The addition of NLS to the original and to the codon-optimized recombinase variants increased their integration activity: FV7 and FV9 showed a higher relative increase than their 'humanized' versions while FV7o and FV9o showed higher absolute integration activity (Figure 3.1E). The integration activity of the NLS modified recombinase variants, as well as the unmodified FV7o and FV9o, exhibited sharp dependence on the input of the respective recombinase expression vectors. The sharp peaks of the integration activity that depend on the recombinase input are not unusual for the tyrosine recombinases and were noted before in the targeting experiments with the purified Cre recombinase [57]. Such sharp peaks may reflect a narrow window of the optimal combination of the efficiency of the reversible integration-excision reactions and the input of a recombinase.

The original and the 'humanized' NLS-modified Flp variants had different activity patterns at the moderate input of the recombinase expression vectors (the one that permits the highest integration activity of the recombinase variants (Figure 3.1E). At this input, FV7 and FV9 with N-terminal NLS showed higher integration activity than their C-

terminal versions. In contrast, FV7o and FV9o variants showed the inverse activity pattern: the variants with C-terminal NLS were more active than their N-terminal counterparts (Figure 3.1E). Since the integration activity of a recombinase in a cell was found to depend not only on the actual ability of this enzyme to recombine a particular DNA sequence but also on the concentration of the recombinase [57] and the timing of its expression [58]. The observed differences in the activity of the NLS-modified variants may be explained by the differences in the expression of the variants and the timing of their appearance in the nucleus. Indeed, the codon optimization and the addition of NLS can affect these parameters. Another factor to consider is a potential effect of NLS and its location on the actual enzymatic activity of a recombinase variant.

Taken together, the codon optimization of a recombinase gene, the addition of NLS, and the use of a more active Flp variant resulted in an approximately 35-fold increase in the targeting efficiency of *FL-IL10A* in human cells in the model setting: FV7 was able to integrate a reporter in ~0.01% of the transfected cells while the 'humanized' version of FV9 with C-terminally placed NLS performed the task in ~0.35% of the transfected cells.

In the model setting of human cells FV7 was capable of targeting *FL-IL10A* accurately: out of ten randomly expanded integration-positive clones, only one clone had vector integrated somewhere else in the genome, in addition to the correctly integrated vector (Figure 3.1C).

Tyrosine recombinase variants that are specific for the genomic target-like sequences are useful only if they are able to utilize these sequences in their native chromosomal environment. Therefore, after improving the integration activity of the Flp variants in human cells in the model setting, we tested whether these variants can recombine the *FL-*

IL10A sequence in its native chromosomal environment in the unmodified human cells. The successful targeting of *FL-IL10A* demonstrated that sequence specific Flp variants are indeed capable of accomplishing the task (Figure 3.2). To our knowledge, similar results for tyrosine recombinases have not been demonstrated before. Indeed, in the earlier report, a Cre variant evolved to recombine a *loxP*-like sequence in the LTRs of HIV-1 was shown to target this sequence in the human cells [59]. However, since a copy of the HIV-1 pseudotype genome used in that study was randomly integrated and stably maintained in the human cells, the resultant HIV-1 provirus should be considered a pre-introduced model recombination reporter, functionally equivalent to any model recombination reporter for this matter. Therefore, although quite remarkable, the experiments described by Sarkar et al. [59] do not demonstrate that tyrosine recombinases variants are able to utilize genomic target-like sequences in their native chromosomal locale.

The performance of TAL-Cre-M24 against just plain Cre-M24 was tested in the bacterial, CHO model cells and HEK293 cells for efficient recombination activity. Though the experimental attempts with plain Cre-M24 were unsuccessful in all the three screening systems, when Cre-M24 was fused to the TAL-core DNA-binding domain recombination sufficient activity was observed. The other Cre variants listed in Table 3.2 had various mutations at the DNA-recombinase interaction points, and were also screened through the bacterial and CHO model cell systems. The random mutations in the Cre variants were such that it would loosen the complexity with which the Cre monomer-monomer interact or Cre-DNA interact leading to a flexible approach for them to recognize the target sequence. These variants were screened with and without attachment

of TAL-core DNA-binding domain, but all attempts to achieve minimal recombination activity failed. There were a few green/red colonies observed with TAL-core DNA-binding domain variant, but the green color always faded in the colonies on expansion. I am not sure about any particular reason behind the fading of green color, but there can be silencing of the EGFP gene when the reporter and expression plasmids are present in the same cell either activating or suppressing certain pathways or there can be random integrations in the genome that cause the EGFP to lose its expression in the cells. TAL-Cre-M24 was the only active variant we attained from all the directed evolution protocols that showed 0.1% replacement activity in the CHO model cells.

The initial construction of TAL-Cre-M24 consisted of 18 nucleotide RVD's assembled on left and right arms of DNA-binding domains. Half spacer region was used to fuse the Cre and the TAL-core DNA-binding domains. There was a minimal recombination activity observed in the bacterial and CHO model cells. On an average just one green/red colony was observed every transfection reaction. Attempts were made to have better replacement activity in CHO model cells by modifying the TAL-core, the DNA-binding domains of TAL-Cre-M24 variant and the linker region present between the Cre and TAL-core DNA-binding domain. Various combinations of the TAL core ranging from 117+95, 117+63 and 117+28 C to N terminal were tested in the model cells. There was a two-fold difference in activity amongst the TAL-core with 117+95 TAL-core being highly efficient. Thus the decision we made of using TAL-core 117+95 in previous experiments was helpful. We also observed that half linker region proved to be a better match between the TAL-core DNA-binding region and Cre-M24 variant as it provided enough flexibility to Cre-M24 to recognize and bind at the genomic LL-69058

target sequence. Half linker regions had two-fold better activity compared with one and full linker spacer regions.

Furthermore, out of the various DNA-binding domain arms, the combination of 15 and 12 RVD's on the left and right arm respectively, proved to be highly efficient when fused with Cre-M24 variant, increasing the recombination activity by five-fold in the model cells. Thus the new version: TALE-117+95-Half linker-Cre-M24, left LV15 and right RV12 combination was 5-fold more efficient than the old version: TAL-117+95-Half linker-Cre-M24, left LV18 and right RV18. The new version TAL-117+95-Half linker-Cre-M24 is referred to as TAL-Cre-M24-LV15/RV12 in all my experiments. Overall it was observed that the major increase in the recombination efficiency was basically due to the change in the DNA-binding domain arms. As the arms were shortened to 15 and 12 nucleotides, they tended to improve by the replacement efficiency by 0.58%. The mutation at the 24th position in the Cre gene helped in releasing the tight bonding between the monomer-monomer protein subunits and when the Cre-M24 variant is cloned with the shorter arms of TAL-domans, it becomes feasible for the recombinase to bind at the *LL-69058* target sequence. As the arms are short and closer to the target sequence (Figure 3.4 D) they bind firmly to the DNA giving full access and flexibility to the Cre recombinase with half linker region. As we have seen in control experiments, plain Cre-M24 just did not show any recombination activity. Thus only when Cre-M24 is bonded to TAL-domain does it serve our purpose.

In the next experiments we tested the TAL-Cre-M24-LV15/RV12 variant in the native genomic environment. As the experiment was performed just once, the data represented here is from one time PCR analysis and sequencing. The TAL-Cre-M24

variant successfully integrated at the genomic *LL-69058* target sequence (Figure 3.6). The results were confirmed through sequencing of the correct size (~800 bp) PCR band. More TAL-Cre-M24 based integration experiments in the HEK293 cells are to be performed in the future to confirm the results. TAL-Flp variant was also tested for its recombination efficiency in the native genomic environment in the HEK293 cells. Several integration experiments were performed and it was observed that TAL-Flp variant recognizes the genomic *FL-71362* as confirmed by PCR band size and sequencing. Therefore, both the evolved TAL-Cre-M24 and TAL-Flp recombinases can recognize the genomic target sequences separately when tested in the human cells.

Future experiments aim at using both the evolved TAL-Cre-M24 and TAL-Flp recombinases simultaneously by co-transfecting them along with the reporter plasmid to recognize the genomic *LL-69058* and *FL-71362* target sequences and perform successful replacement of the gene cassette in the native genomic environment. Various concentrations of the recombinases are to be tested and optimized to have consistent replacement. If the dual RMCE attempt is successful in the genomic environment then we can prove to have a potential cure for the sickle cell disease. As the genomic target sequences *LL-69058* and *FL-71362* flank the beta-globin gene that consists the mutation causing sickle cell disease, replacing the gene cassette between the two target sequences would replace the mutated region with a normal gene region. It could be used as gene therapy for sickle cell disease. This fundamental approach of evolving recombinases to recognize genomic target-like sequences, can also be applied to cure other diseases that tend to possess recombinase target-like sequences in their vicinity in the human genome.

APPENDIX

SUPPLEMENTAL FIGURES AND SEQUENCES

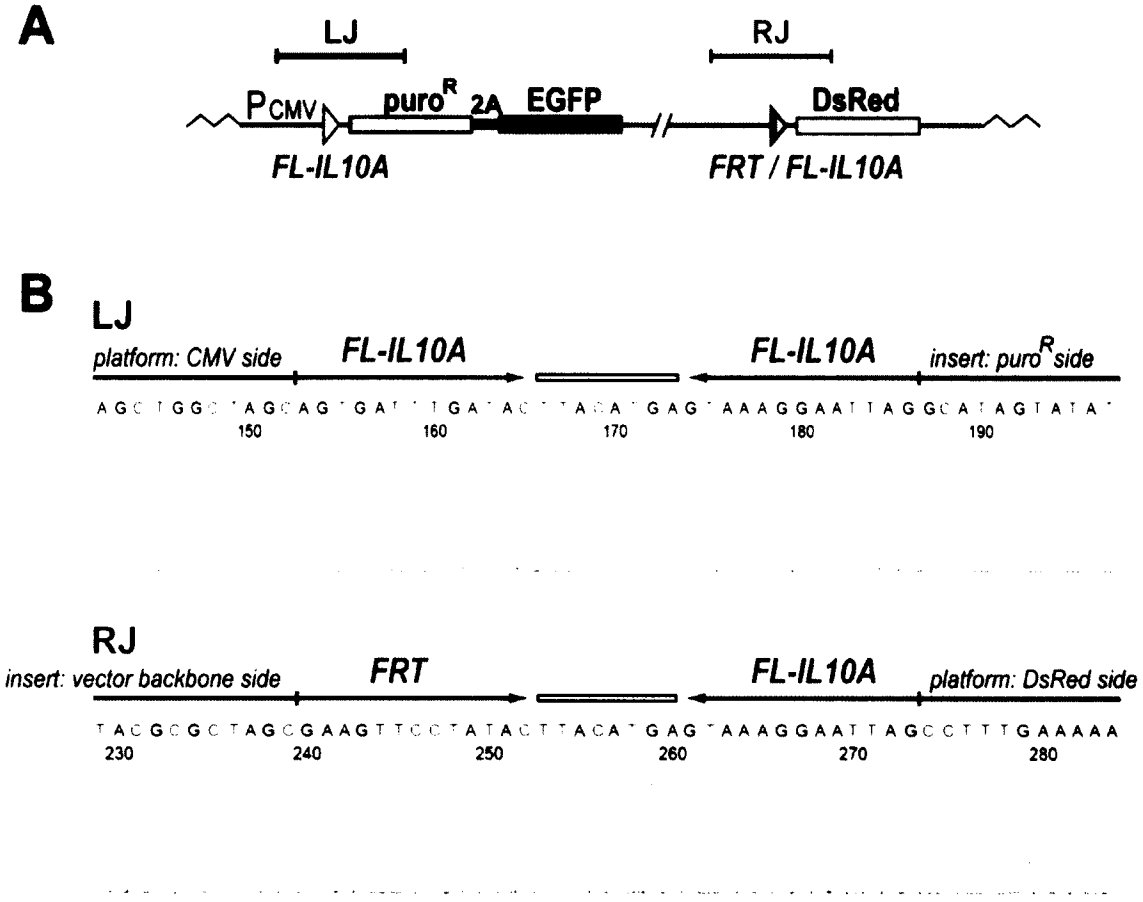


Figure A.1: Sequencing results snapshot of the FV7-mediated integration product of the incoming reporter into FL-IL10A located in the platform reporter pre-integrated into the HEK 293 genome (see Figure 3.1). (A) Schematic of the integration product. The horizontal green and red bars represent the diagnostic PCR products at the left junction (LJ) and the right junction (RJ) of the integration product. (B) The sequencing results snapshots of the respective PCR products are shown below the schematics of the resultant recombination sites and the surrounding sequences.

Protein alignment of the whole sequences for iCreM32-iCre262-1,14,21.

```
************************************************************************
iCreM24
MVPKKKRKVSNLLTVHQNLPALPVDATSDEVMKNLMDMFRDRQAFSEHTWKMLLSVCRSW 60
iCre-262-1
MVPKKKRKVSNLLTVHQNLPALPVDATSDEVMKNLMDMFRDRQAFSEHTWKMLLSVCRSW 60
iCre-262-21
MVPKKKRKVSNLLTVHQNLPALPVDATSDEVMKNLMDMFRDRQAFSEHTWKMLLSVCRSW 60
iCre-262-14
MVPKKKRKVSNLLTVHQNLPALPVDATSDEVMKNLMDMFRDRQAFSEHTWKMLLSVCRSW 60
************************************************************************
      iCreM24
AAWCKLNNRKWFPAEPEDVRDYLLYLQARGLAVKTIQQHLGQLNMLHRRSGLPRPSDSNA 120
      iCre-262-1
AAWCKLNNRKWFPAEPEDVRDYLLYLQARGLAVKTIQQHLGQLNMLHRRSGLPRPSDSNA 120
iCre-262-21
AAWCKLNNRKWFPAEPEDVRDYLLYLQARGLAVKTIQQHLGQLNMLHRRSGLPRPSDSNA 120
iCre-262-14
AAWCKLNNRKWFPAKPEDVRNYLLYLQARGLAVKTIQQHLGQLNMLHRRSGLPRPSDSNA
**************:*****:******************************************
      iCreM24
VSLVMRRIRKENVDAGERAKQALAFERTDFDQVRSLMENSDRCQDIRNLAFLGIAYNTLL 180
      iCreM-262-1
VSLVMRRIRKENVDAGERAKQALAFERTDFDQVRSLMENSDRCQDIRNLAFLGIAYNTLL 180
      iCreM-262-21
VSLVMRRIRKENVDAGERAKQALAFERTDFDQVRSLMENSDRCQDIRNLAFLGIAYNTLL 180
      iCreM-262-14
VSLVMRRIRKENVDAGERAKQALAFERTDFDQVRSLMENSDRCQDIRNLAFLGIAYNTLL
************************************************************************
      iCreM24
RIAEIARIRVKDISRTDGGRMLIHIGRTKTLVSTAGVEKALSLGVTKLVERWISVSGVAD 240
      iCre-262-1
RIAEIARIRVKDISRTDGGRMLIHIGRTKTLVSTAGVEKALSLGVTKLVERWISVSGVAD 240
      iCre-262-21
RIAEIARIRVKDISRTDGGRMLIHIGRTKTLVSTAGVEKALSLGVTKLVERWISVSGVAD 240
      iCre-262-14
RIAEIARIRVKDISRTDGGRMLIHIGRTKTLVSTAGVEKALSLGVTKLVERWISVSGVAD 240
************************************************************************
      iCreM24
DPNNYLFCRVRKNGVAAPSATSQLSTRALEGIFEATHRLIYGAKDDSGQRYLAWSGHSAR 300
      iCre-262-1
DPNNYLFCRVRKNGVAAPSATSQLSTRALNGIFEATHRLIYGAKDDSGQRYLAWSGHSAR 300
      iCre-262-21
DPNNYLFCRVRKNGVAAPSATSQLSTRALSGIFEATHRLIYGAKDDSGQRYLAWSGHSAR 300
      iCre-262-14
DPNNYLFCRVRKNGVAAPSATSQLSTRALSGIFEATHRLIYGAKDDSGQRYLAWSGHSAR 300
****************************.***************************************
      iCreM24
VGAARDMARAGVSIPEIMQAGGWTNVNIVMNYIRNLDSETGAMVRLLEDGD 351
```

```
      iCre-262-1
VGAARDMARAGVSIPEIMQAGGWTNVNIVMNYIRNLDSETGAMVRLLEDGD 351
      iCre-262-21
VGAARDMARAGVSIPEIMQAGGWTNVNIVMNYIRNLDSETGAMVRLLEDGD 351
      iCre-262-14
VGAARDMARAGVSIPEIMQAGGWTNVNIVMNYIRNLDSETGAMVRLLEDGD 351
*************************************************************
```

Figure A.2: Protein sequence alignments of Cre-M24, Cre262-1, Cre262-14 and Cre262-21 using Clustal Omega software. The stars below the sequence indicate match in amino acids while a dot indicate a mutation in the protein sequence. The location of the amino acid can be determined by the numbers present at the end of the sequence.

BIBLIOGRAPHY

[1] W. M. Stark, "The Serine Recombinases," *Microbiol. Spectr.*, vol. 2, no. 6, pp. 1-16, Nov. 2014.

[2] W. R. A. Brown, N. C. O. Lee, Z. Xu, and M. C. M. Smith, "Serine recombinases as tools for genome engineering," *Methods*, vol. 53, no. 4, pp. 372–379, 2011.

[3] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, "Gapped BLAST and PS I-BLAST: a new generation of protein database search programs," *Nucleic Acids Res.*, vol. 25, no. 17, pp. 3389–3402, 1997.

[4] N. D. F. Grindley, K. L. Whiteson, and P. A. Rice, "Mechanisms of site-specific recombination," *Annu. Rev. Biochem.*, vol. 75, no. 1, pp. 567–605, Jun. 2006.

[5] J. L. Shultz, E. Voziyanova, J. H. Konieczka, and Y. Voziyanov, "A genome-wide analysis of FRT-like sequences in the human genome.," *PLoS One*, vol. 6, no. 3, p. e18077, Jan. 2011.

[6] E. Yagil, L. Dorgai, and R. A. Weisberg, "Identifying determinants of recombination specificity: construction and characterization of chimeric bacteriophage integrases.," *J. Mol. Biol.*, vol. 252, no. 2, pp. 163–77, Sep. 1995.

[7] L. Dorgai, E. Yagil, and R. A. Weisberg, "Identifying determinants of recombination specificity: construction and characterization of mutant bacteriophage integrases.," *J. Mol. Biol.*, vol. 252, no. 2, pp. 178–88, Sep. 1995.

[8] W. P. Stemmer, "DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 91, no. 22, pp. 10747–10751, 1994.

[9] H. Zhao and F. H. Arnold, "Optimization of DNA shuffling for high fidelity recombination," *Nucleic Acids Res.*, vol. 25, no. 6, pp. 1307–1308, 1997.

[10] Y. Voziyanov, A. F. Stewart, and M. Jayaram, "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.*, vol. 30, no. 7, pp. 1656–1663, 2002.

[11] B. Sauer and N. Henderson, "Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1," *PNAS*, vol. 85, no. 13, July, pp. 5166–5170, 1988.

[12] Y. Tay, C. Ho, P. Droge, and F. J. Ghadessy, "Selection of bacteriophage lambda integrases with altered recombination specificity by in vitro compartmentalization.," *Nucleic Acids Res.*, vol. 38, no. 4, p. e25, Mar. 2010.

[13] P. D. Sadowski, "Site-specific genetic recombination: hops, flips, and flops.," *FASEB J.* , vol. 7 , no. 9 , pp. 760–767, Jun. 1993.

[14] P. Sadowski, "Site-specific genetic recombination: hops, flips, and flops," *FASEB J*, vol. 7, no. 9, pp. 760–767, Jun. 1993.

[15] N Gopaul and Gregory D Van Duyne, "Structure and mechanism in site-specific recombination."*Curr Opin Struc Biol.*, vol 9, pp. 14-20, 1999.

[16] M. C. M. Smith and H. M. Thorpe, "Diversity in the serine recombinases," *Mol. Biol.*, vol. 44, no. 2, pp. 299–307, 2002.

[17] M. M. McLean, Y. Chang, G. Dhar, J. K. Heiss, and R. C. Johnson, "Multiple interfaces between a serine recombinase and an enhancer control site-specific DNA inversion," *Elife*, vol. 2013, pp. 2, 2013.

[18] T. Gaj, A. C. Mercer, S. J. Sirk, H. L. Smith, and C. F. Barbas, "A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells.," *Nucleic Acids Res.*, vol. 41, no. 6, pp. 3937–46, 2013.

[19] F. J. Olorunniji, S. J. Rosser, and W. M. Stark, "Site-specific recombinases: molecular machines for the Genetic Revolution," *Biochem. J.*, vol. 473, no. 6, pp. 673–684, Mar. 2016.

[20] M.-C. Birling, F. Gofflot, and X. Warot, "Site-specific recombinases for manipulation of the mouse genome," in *Transgenesis Techniques SE - 16*, vol. 561, E. J. Cartwright, Ed. Humana Press, 2009, pp. 245–263.

[21] S. Turan, C. Zehe, J. Kuehle, J. Qiao, and J. Bode, "Recombinase-mediated cassette exchange (RMCE) — A rapidly-expanding toolbox for targeted genomic modifications," *Gene*, vol. 515, no. 1, pp. 1–27, Feb. 2013.

[22] M. R. Capecchi, "Altering the genome homologous recombination by from ES cells to germ line chimera," *Science*, vol. 236, pp.1288-1292, 1987, 1989.

[23] D. N. Gopaul, F. Guo, and G. D. Van Duyne, "Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination.," *EMBO J.*, vol. 17, no. 14, pp. 4175–87, Jul. 1998.

[24] L. Rajeev, K. Malanowska, and J. F. Gardner, "Challenging a paradigm: the role of DNA homology in tyrosine recombinase reactions.," *Microbiol. Mol. Biol. Rev.*, vol. 73, no. 2, pp. 300–309, 2009.

[25] D. Esposito and J. J. Scocca, "The integrase family of tyrosine recombinases: evolution of a conserved active site domain," *Nucleic Acids Res.*, vol. 25, no. 18, pp. 3605–3614, Sep. 1997.

[26] Y. Voziyanov, S. Pathania, and M. Jayaram, "Survey and summary. A general model for site-specific recombination by the integrase family recombinases," *Nucleic Acids Res.*, vol. 27, no. 4, pp. 930–941, 1999.

[27] S. Turan, M. Galla, E. Ernst, J. Qiao, C. Voelkel, B. Schiedlmeier, C. Zehe, and J. Bode, "Recombinase-mediated cassette exchange (RMCE): Traditional concepts and current challenges," *J. Mol. Biol.*, vol. 407, no. 2, pp. 193–221, 2011.

[28] B. Gibb, K. Gupta, K. Ghosh, R. Sharp, J. Chen, and G. D. Van Duyne, "Requirements for catalysis in the Cre recombinase active site.," *Nucleic Acids Res.*, vol. 38, no. 17, pp. 5817–32, Sep. 2010.

[29] E. Voziyanova, N. Malchin, R. P. Anderson, E. Yagil, M. Kolot, and Y. Voziyanov, "Efficient Flp-Int HK022 dual RMCE in mammalian cells.," *Nucleic Acids Res.*, vol. 41, no. 12, p. e125, Jul. 2013.

[30] B. Sauer and N. Henderson, "Cre-stimulated recombination at loxP-containing DNA sequences placed into the mammalian genome.," *Nucleic Acids Res.*, vol. 17, no. 1, pp. 147–161, Jan. 1989.

[31] E. Ennifar, "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.*, vol. 31, no. 18, pp. 5449–5460, Sep. 2003.

[32] R. Evans, L. Parson, K. Bauerll, and B. Teplow, "Identification of the active site tyrosine of Flp recombinase," *J. Biol. Chem.*, vol. 265, no. 30, pp. 18504–18510, 1990.

[33] D. W. Leung, E. Chen, and D. V Goeddel, "A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction," *Technique*, vol. 1, pp. 11–15, 1989.

[34] R. C. Cadwell and G. F. Joyce, "Randomization of genes by PCR mutagenesis," *Genome Res.*, vol. 2, no. 1, pp. 28–33, 1992.

[35] P. Carter, "Site-directed mutagenesis.," *Biochem. J.*, vol. 237, pp. 1–7, 1986.

[36] S. Herlitze and M. Koenen, "A general and rapid mutagenesis method using polymerase chain reaction.," *Gene*, vol. 91, no. 1, pp. 143–147, 1990.

[37] N. E. Sanjana, L. Cong, Y. Zhou, M. M. Cunniff, G. Feng, and F. Zhang, "A transcription activator-like effector toolbox for genome engineering.," *Nat. Protoc.*, vol. 7, no. 1, pp. 171–92, 2012.

[38] T. Gaj, S. J. Sirk, R. D. Tingle, A. C. Mercer, M. C. Wallen, and C. F. Barbas, "Enhancing the specificity of recombinase-mediated genome engineering through dimer interface redesign," *J. Am. Chem. Soc.*, vol. 136, no. 13, pp. 5047–5056, 2014.

[39] P. D. Hsu, E. S. Lander, and F. Zhang, "Development and applications of CRISPR-Cas9 for genome engineering.," *Cell*, vol. 157, no. 6, pp. 1262–1278, 2014.

[40] P. Mali, L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville, and G. M. Church, "RNA-guided human genome engineering via Cas9 Prashant," *Science*, vol. 339, no. 6121, pp. 823–826, 2013.

[41] L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, and F. Zhang, "Multiplex genome engineering using CRISPR/Cas system," *Science*, vol. 339, no. 6121, pp. 819–823, 2013.

[42] X. Zhang, W. H. Koolhaas, and F. Schnorrer, "A versatile two-step CRISPR- and RMCE-based strategy for efficient genome engineering in Drosophila.," *G3 (Bethesda).*, vol. 4, no. 12, pp. 2409–18, 2014.

[43] A. C. Mercer, T. Gaj, R. P. Fuller, and C. F. Barbas, "Chimeric TALE recombinases with programmable DNA sequence specificity.," *Nucleic Acids Res.*, vol. 40, no. 21, pp. 11163–72, Nov. 2012.

[44] C. Mussolino, J. Alzubi, E. J. Fine, R. Morbitzer, T. J. Cradick, T. Lahaye, G. Bao, and T. Cathomen, "TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity," *Nucleic Acids Res.*, vol. 42, no. 10, pp. 6762–6773, 2014.

[45] H. Scholze and J. Boch, "TAL effectors are remote controls for gene activation," *Curr. Opin. Microbiol.*, vol. 14, no. 1, pp. 47–53, 2011.

[46] P. Römer, S. Hahn, T. Jordan, T. Straußß, U. Bonas, and T. Lahaye, "Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene," *Science*, vol. 318, no. 5850, pp. 645–648, Oct. 2007.

[47] J. Boch, H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt, and U. Bonas, "Breaking the code of DNA binding specificity of TAL-type III effectors.," *Science*, vol. 326, no. 5959, pp. 1509–1512, 2009.

[48] M. J. Moscou and A. J. Bogdanove, "A Simple cipher governs DNA recognition by TAL effectors," *Science.*, vol. 326, no. 5959, p. 1501, 2009.

[49] S. M. Tan and P. Dröge, "Comparative analysis of sequence-specific DNA recombination systems in human embryonic stem cells.," *Stem Cells*, vol. 23, no. 7, pp. 868–73, Aug. 2005.

[50] R. P. Anderson, E. Voziyanova, and Y. Voziyanov, "Flp and Cre expressed from Flp-2A-Cre and Flp-IRES-Cre transcription units mediate the highest level of dual recombinase-mediated cassette exchange," *Nucleic Acids Res.*, vol. 40, no. 8, pp. 1–9, 2012.

[51] J. Sambrook, D.W. Russell, *Molecular cloning.: a laboratory manual., 3rd ed.* Cold Spring Harbor, New York, 2001.

[52] S. Bolusani, C.-H. Ma, A. Paek, J. H. Konieczka, M. Jayaram, and Y. Voziyanov, "Evolution of variants of yeast site-specific recombinase Flp that utilize native genomic sequences as recombination target sites.," *Nucleic Acids Res.*, vol. 34, no. 18, pp. 5259–69, Jan. 2006.

[53] R. Shah, F. Li, E. Voziyanova, and Y. Voziyanov, "Target-specific variants of Flp recombinase mediate genome engineering reactions in mammalian cells.," *FEBS J.*, vol. 282, no. 17, pp. 3323–33, Sep. 2015.

[54] C. S. Raymond and P. Soriano, "High-efficiency FLP and ΦC31 site-specific recombination in mammalian cells," *PLoS One*, vol. 2, no. 1, p. e162, Jan. 2007.

[55] D. R. Shimshek, J. Kim, M. R. Hübner, D. J. Spergel, F. Buchholz, E. Casanova, A. F. Stewart, P. H. Seeburg, and R. Sprengel, "Codon-improved Cre recombinase (iCre) expression in the mouse.," *Genesis*, vol. 32, no. 1, pp. 19–26, Jan. 2002.

[56] L. L. M. Guzman, D. Belin, M. J. Carson, J. Beckwith,"Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter," *J. Bacteriol.*, vol. 177, no. 14, pp. 4121–4130, 1995.

[57] W. Baubonis and B. Sauer, "Genomic targeting with purified Cre recombinase.," *Nucleic Acids Res.*, vol. 21, no. 9, pp. 2025–2029, 1993.

[58] C. Logie and a F. Stewart, "Ligand-regulated site-specific recombination.," *Proc Natl Acad Sci USA*, vol. 92, no. 13, pp. 5940–5944, 1995.

[59] I. Sarkar, I. Hauber, J. Hauber, and F. Buchholz, "HIV-1 proviral DNA excision using an evolved recombinase.," *Science*, vol. 316, no. 5833, pp. 1912–5, Jun. 2007.