This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.
Improving *mariner* transposons for transgenesis

Maryia Trubitsyna
I dedicate this thesis to my family.
Declaration

This thesis has been composed by myself and all the work herein was done by myself, unless otherwise stated. This work has not been submitted for any other degree or professional qualification.

Maryia Trubitsyna

February, 2014
Acknowledgements

First of all I want to thank my Supervisors, Prof David J. Finnegan and Dr Julia M. Richardson for giving me the opportunity to do my PhD research project at the University of Edinburgh. Your experience, knowledge, support and kindness has allowed me to face the challenges that I have encountered during these very important four years of my life. I am so grateful for the skills you shared with me, your patience, understanding and your generosity. Thank you, Julia, for taking me to the unforgettable conference in California, where I got so much inspiration and confidence in my work.

Many thanks to Eve Hartswood and Jim Brodie whose warmth and welcoming ways helped me settle into a new country. I particularly appreciate all the local knowledge and Scientific support you offered me. Special thanks to Jacqueline Dornan, for being a great person, fantastic scientist and a warm friend.

I would like to thank Garry Blakely for my starting plasmids and strains, without which I wouldn’t have discovered the *in vivo* transposition. Ken Sawin and his lab for teaching me how to work with yeast and Gracjan Michlewski and his lab for introducing me to the world of mammalian cell culture.

Special gratitude to my mom, dad, my grannies and my brother, who have helped make me the person I am today. Thanks for supporting me at all stages of my life, for believing in me and always being there for me at any time of day or night.

My sincere acknowledgements to the Darwin Trust of Edinburgh for an amazing life changing opportunity to do a PhD, which has opened the doors of World Science to me.

I would like to thank my closest friend and companion, who has made every day of my life meaningful and given me a true feeling of happiness.

It has been a remarkable journey.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Å</td>
<td>angstrom</td>
</tr>
<tr>
<td>a.a.</td>
<td>amino acids</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>CAI</td>
<td>codon adaptation index</td>
</tr>
<tr>
<td>cDNA</td>
<td>coding DNA</td>
</tr>
<tr>
<td>CFU</td>
<td>colony forming units</td>
</tr>
<tr>
<td>cm</td>
<td>centimetre</td>
</tr>
<tr>
<td>CV</td>
<td>column volume</td>
</tr>
<tr>
<td>dH&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>distilled water</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNC</td>
<td>dominant negative complementation</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>E</td>
<td>energy</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetertaacetic acid</td>
</tr>
<tr>
<td>em</td>
<td>emission</td>
</tr>
<tr>
<td>ex</td>
<td>excitation</td>
</tr>
<tr>
<td>FP</td>
<td>Frog Prince transposon</td>
</tr>
<tr>
<td>g</td>
<td>gravity</td>
</tr>
<tr>
<td>GTP</td>
<td>guanosine triphosphate</td>
</tr>
<tr>
<td>HPLC</td>
<td>high pressure liquid chromatography</td>
</tr>
<tr>
<td>HTH</td>
<td>helix-turn-helix</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>IR</td>
<td>inverted repeat</td>
</tr>
<tr>
<td>IRL</td>
<td>inverted repeat left</td>
</tr>
<tr>
<td>IRR</td>
<td>inverted repeat right</td>
</tr>
<tr>
<td>k</td>
<td>Boltzmann constant</td>
</tr>
<tr>
<td>kb</td>
<td>kilo base pairs</td>
</tr>
<tr>
<td>kDa</td>
<td>kilo Daltons</td>
</tr>
<tr>
<td>LARDs</td>
<td>large retrotransposon derivatives</td>
</tr>
<tr>
<td>LTR</td>
<td>long terminal inverted repeat</td>
</tr>
<tr>
<td>M</td>
<td>molecular mass marker</td>
</tr>
<tr>
<td>M</td>
<td>molar</td>
</tr>
<tr>
<td>mAU</td>
<td>milli-absorbance unit</td>
</tr>
<tr>
<td>mg</td>
<td>milligram</td>
</tr>
<tr>
<td>min</td>
<td>minute</td>
</tr>
<tr>
<td>miRNA</td>
<td>micro RNA</td>
</tr>
<tr>
<td>ml</td>
<td>millilitre</td>
</tr>
<tr>
<td>mm</td>
<td>millimetre</td>
</tr>
<tr>
<td>M</td>
<td>molarity</td>
</tr>
<tr>
<td>Mr</td>
<td>molecular mass</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>mS</td>
<td>millisiemens (unit of electric conductance)</td>
</tr>
<tr>
<td>MWCO</td>
<td>molecular weight cut off</td>
</tr>
<tr>
<td>nm</td>
<td>nanometre</td>
</tr>
<tr>
<td>NTS</td>
<td>non-transferred strand</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PB</td>
<td>piggyBac transposon</td>
</tr>
<tr>
<td>PEC</td>
<td>paired end complex</td>
</tr>
<tr>
<td>PEI</td>
<td>polyethylenimine</td>
</tr>
<tr>
<td>PES</td>
<td>polyethersulfone (membrane)</td>
</tr>
<tr>
<td>pi-RNA</td>
<td>piwi interacting RNA</td>
</tr>
<tr>
<td>RC</td>
<td>rolling circle</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>RRS</td>
<td>Rep recognition sequence</td>
</tr>
<tr>
<td>RT/En</td>
<td>reverse transcriptase/endonuclease</td>
</tr>
<tr>
<td>SB</td>
<td>Sleeping Beauty transposon</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl sulphate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>T</td>
<td>temperature</td>
</tr>
<tr>
<td>TE</td>
<td>transposable element</td>
</tr>
<tr>
<td>TIR</td>
<td>terminal inverted repeat</td>
</tr>
<tr>
<td>Tm</td>
<td>melting temperature</td>
</tr>
<tr>
<td>TRIMs</td>
<td>terminal-repeat retrotransposons in miniature</td>
</tr>
<tr>
<td>Trn</td>
<td>transposase</td>
</tr>
<tr>
<td>Trp</td>
<td>transposon</td>
</tr>
<tr>
<td>TS</td>
<td>transferred strand</td>
</tr>
<tr>
<td>TSD</td>
<td>target site duplication</td>
</tr>
<tr>
<td>TSS</td>
<td>transcriptional target site</td>
</tr>
<tr>
<td>Txn</td>
<td>transcription</td>
</tr>
<tr>
<td>UAS</td>
<td>upstream activating sequence</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>v/v</td>
<td>volume per volume</td>
</tr>
<tr>
<td>viRNA</td>
<td>virus-derived mi-RNA</td>
</tr>
<tr>
<td>W</td>
<td>watt</td>
</tr>
<tr>
<td>w/v</td>
<td>weight per volume</td>
</tr>
<tr>
<td>μg</td>
<td>microgram</td>
</tr>
<tr>
<td>μl</td>
<td>microlitre</td>
</tr>
<tr>
<td>μM</td>
<td>micromolar</td>
</tr>
</tbody>
</table>
Amino acids

A  Ala  alanine
C  Cys  cysteine
D  Asp  aspartic acid
E  Glu  glutamic acid
F  Phe  phenylalanine
G  Gly  glycine
H  His  histidine
I  Ile  isoleucine
K  Lys  lysine
L  Leu  leucine
M  Met  methionine
N  Asn  asparagine
P  Pro  proline
Q  Gln  glutamine
R  Arg  arginine
S  Ser  serine
T  Thr  threonine
V  Val  valine
W  Trp  tryptophan
Y  Tyr  tyrosine

Bases

A  adenine
C  cytosine
G  guanine
T  thymine
R  A or G
Y  C or T
M  A or C
K  G or T
S  C or G
W  A or T
H  A or C or T
B  C or G or T
V  A or C or G
D  A or G or T
N  A or C or G or T
Abstract

Transgenesis is a process of introducing foreign genetic material into the genomes of living organisms. One of the tools for transgenesis are the transposable elements (TEs), which include transposons. Transposons are naturally occurring sequences of DNA which are recognised, excised and inserted into a new location by a single enzyme – transposase. Here we show studies of the biophysical properties and activities of two highly related eukaryotic TEs of the mariner family: Mos1 from Drosophila mauritiana and Mboumar-9 (Mbo9) from Messor bouviery.

Using biochemical and molecular methods we examined the properties of transposases in vitro and in vivo. Recombinant transposases were expressed in E.coli and purified using HPLC. Each protein’s activity was assayed for cleavage, integration and the whole transposition reaction. We used a modelling approach to predict the structure of the complex of Mbo9 transposase bound to the specific terminal sequences of the transposon, the paired end complex (PEC), based on the published crystal structure of Mos1 PEC.

We have found that both transposases are elongated dimers in solution and that the first helix-turn-helix domain is involved in the protein dimerization. Moreover we show that mariner transposases cut one of the imperfect inverted repeats more efficiently than the other. The terminal nucleotide of the inverted repeat is important for integration of the transposon into a new target DNA, while having no effect at the stage of cleavage.

Previously, neither Mos1 nor Mbo9 had been shown to have significant activity in mammalian cells. We have developed a new assay that allows chromosomal integration of the desired DNA sequence in vivo in bacterial, yeast and mammalian cells without the use of helper plasmids or mRNA injection. We found the optimal combination of inverted repeats for each of the transposons and have enhanced the transposition efficiency of Mbo9 by changing the sequence of its inverted repeat DNA. This study is a foundation for improving mariner TEs for transgenesis.
# Contents

Declaration ........................................................................................................................... ii  
Acknowledgements ............................................................................................................... iii  
Abbreviations ....................................................................................................................... iv  
Amino acids ............................................................................................................................ vi  
Nucleotides ........................................................................................................................... vi  
Abstract ................................................................................................................................ vii  

**Chapter 1: Introduction ..................................................................................................... 1**  
1.1 Classes of transposable elements .................................................................................. 2  
1.2 Mechanisms of transposition ......................................................................................... 5  
1.2.1 DDE-transposases ....................................................................................................... 6  
1.2.2 Target primed retrotransposases ............................................................................... 6  
1.2.3 Tyrosine (Y) transposases ......................................................................................... 9  
1.2.4 Serine (S) transposases ............................................................................................. 9  
1.2.5 Rolling-circle (RC) or Y2-transposases .................................................................... 10  
1.3 Transposable elements as tools for genetic manipulations ............................................. 10  
1.3.1 Transposable elements showing promise for transgenesis applications ................... 11  
1.3.2 Target site selection .................................................................................................. 22  
1.3.3 Possibilities of precise integration targeting .............................................................. 23  
1.3.4 Ways of delivering transposases to the cells .............................................................. 24  
1.4 Regulation of transposition ........................................................................................... 25  
1.4.1 Overproduction inhibition ......................................................................................... 25  
1.4.2 Dominant-negative complementation ....................................................................... 26  
1.4.3 pi-RNA ..................................................................................................................... 26  
1.4.4 CpG methylation ....................................................................................................... 27  
1.5 Tc1/mariner elements transposition .............................................................................. 27  
1.5.1 Mechanism of Mos1 transposition ............................................................................. 27  
1.6 Difference in the inverted repeats ................................................................................. 33  
1.7 Aims of the project ......................................................................................................... 36  

**Chapter 2: Purification and biophysical characterisation of the transposases.. 38**  
2.1 Introduction ..................................................................................................................... 39  
2.2 Results .............................................................................................................................. 39
Chapter 1: Introduction
1.1 Classes of transposable elements

Transposable elements constitute a significant fraction of the genomes of living organisms. For example, 45-69% of the human genome is transposable element DNA (de Koning et al, 2011) and in rice and corn they make up ~85% of the total genome (Fedoroff, 2012).

Transposable elements can, or could in the past, migrate within the host genomes. They can integrate in the regulatory regions of genes, disrupt ORFs of the expressed genes, enhance expression of silenced genes and cause deletions, duplications and genome rearrangements. Transposable elements are widespread throughout bacteria to humans and transmissions can occur between species generating a universal genome modification tool.

Transposable elements are traditionally divided into two classes based on their structural composition (Finnegan, 1989). Class I or RNA mediated transposable elements (retrotransposons) can have Long Terminal Inverted repeats (LTR retrotransposons) or lack them (non-LTR elements). LTR elements encode structural proteins (gag gene) and protease, reverse transcriptase and integrase domains (pol gene). The presence of these two genes is required for the formation of the virus-like particles. Some elements have an additional env-like gene, which encodes active or deficient proteins required for formation of the viral particles. One such element gypsy was shown to be infectious (Song et al, 1994). Elements like gypsy demonstrate the possible emergence of the transposable elements during evolution from retroviruses, or vice versa. Another point of view is that env-like proteins could serve as chaperones during replication (Havecker et al, 2004), but this hypothesis is not yet confirmed.

The two types of LTR retroelements: Ty1/copia and Ty3/gypsy differ by the organisation of coding domains in the pol gene (Kim et al, 1998). Bel/Pao LTR retroelements present in metazoan genomes are very close to Ty3/gypsy. In addition they have an env gene of 4 to 10 kb coding for putative env-like proteins (Malik et al, 2000). Retroviridae specifically circulate in the genomes of vertebrates; they are similar to Ty3/gypsy and Bel/Pao LTR retroelements, but are more complex and
### Transposable Elements

<table>
<thead>
<tr>
<th></th>
<th>Class I (RNA mediated)</th>
<th>Class II (DNA mediated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTR retroelements</td>
<td>Ty3/gypsy Ty1/copia Bel/Pao Retroviridae</td>
<td>R2 LINEs Tc1/mariner hAT Mu piggyBac P element</td>
</tr>
<tr>
<td>non-LTR retroelements</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>autonomous</th>
<th>LARDs TRIMs</th>
<th>SINEs</th>
<th>MITEs</th>
</tr>
</thead>
</table>

### Table 1-1 Traditional classification of transposable elements.

Transposable elements are divided into two classes – RNA mediated (Class I) and DNA mediated (Class II). Class I transposons have an RNA intermediate stage during transposition and can have long terminal repeats (LTR) at the end of the transponson or lack them (non-LTR). Autonomous elements encode all the proteins needed for transposition to occur. Non-autonomous elements do not encode all the proteins involved in transposition, they require a functional transposase expressed from an autonomous element to be transposed.
incorporate additional genes necessary for their life cycle and transmission between the hosts (Llorens et al, 2011).

Transposable elements which have accumulated mutations in the genes encoding the proteins essential for transposition are called non-autonomous elements. Non-autonomous LTR-retrotransposons belong to two classes: LARDs (LArge Retrotransposons Derivatives) and TRIMs (Terminal-Repeat Retrotransposons in Miniature). LARDs are 5.5 to 8.5 kb long, found in barley and rice and probably evolved from Ty3/gypsy LTR-retroelements (Kalander et al, 2004). TRIMs are up to 560 bp long, found in many plant species (Witte et al, 2001). No intermediate forms between TRIMs and LARDs have been found to date.

Non-LTR retrotransposons can be divided into two classes: R1/R2 elements and LINEs (Long INterspersed repetitive Elements). R1 and R2 elements occupy 30-60% of 28S rRNA genes of insects and integrate into the specific sites a short distance apart (Eickbush, 2002; Jakubczak et al, 1991). Retrotransposition of R2 occurs in the germ lines of both sexes, while R1 is mostly active in the male germ line (Perez-Gonzalez et al, 2003). These elements have a single ORF with a reverse transcriptase domain and an endonuclease domain at the C-terminus (Bibillo & Eickbush, 2002). LINEs have one ORF similar to the gag gene and a second ORF encoding a reverse transcriptase and an apurinic-apyrimidinic endonuclease domain (Jamburuthugoda et al, 2008). About 17% of the human genome contains LINEs. SINEs (Short INterspersed repetitive Elements) do not encode any functional protein (non-autonomous) and rely on the activity of the reverse transcriptase of other elements for their transposition (Wallace et al, 2008). SINEs make up around 11% of the human genome; the most widespread SINE is the Alu element (Cordaux & Batzer, 2009).

Class II or DNA transposable elements (transposons) do not have an RNA intermediate stage. Most of them transpose by a cut-and-paste mechanism. Only one protein is required for transposition of most elements – transposase. Transposase recognises specific sequences at the ends of a transposable element, binds to them and excises the element from its location. Then the complex of transposon and
Improving *mariner* transposons for transgenesis

Transposase captures a target DNA forming a target capture complex. A base at the target site is attacked by the 3-OH at the end of the excised transposon and after the transposon is integrated into a new location, the activities of the host DNA polymerase and ligase are required to fill in the gaps.

Members of the *Tc1/mariner* superfamily of transposons encode a transposase with the characteristic DDE/DDD catalytic residues and are flanked with terminal inverted repeats of various lengths. They insert with a TA target site duplication and excise leaving footprints. This superfamily includes such transposons as Mos1, Mboumar-9, Sleeping Beauty, Minos and Himar1.

The hAT superfamily of transposons includes Ac, Tol2, hobo, Hermes, Hector, TcBuster1, SPIN. They have a conserved 50 amino acids dimerization domain at the C-terminus of transposase and integrate with 8 bp duplications. The transposons of this family are the most abundant DNA transposons, found in the many organisms from plants to humans (Arensburger et al, 2011).

More than 10 superfamilies of the DNA transposons have been described to date (Bao et al, 2009). The most recent members (Harbinger, Transib, Mirage, Helitron) were identified using the *in silico* methods: computational analysis, multiple alignment and phylogenetic analysis (Kapitonov & Jurka, 2004).

The non-autonomous DNA transposons have accumulated inactivating mutations in either the transposase ORF or/and the inverted repeats and can not be recognised by the transposase. Some of the elements lost the transposase gene and consist of two inverted repeats only and are called Miniature Inverted repeat Transposable Elements (MITEs). The non-autonomous transposons outnumber the active transposons in the most genomes. They do not express an active transposase and are a silent genome baggage unless moved by a transposase expressed from an autonomous element (Slotkin & Martienssen, 2007).

**1.2 The mechanisms of transposition**

There are five mechanisms described that can incorporate all the known retrotransposons and transposons to date. By the mechanism of transposition,
retro/transposases are divided into five types (Figure 1-1) (Curcio & Derbyshire, 2003):

1. DDE-transposases;
2. Target primed retrotransposases;
3. Tyrosine (Y) transposases;
4. Serine (S) transposases;
5. Rolling-circle (RC) or Y2-transposases.

### 1.2.1 DDE-transposases

DDE-transposases (Figure 1-1 a and b) contain two Asp (D) and one Glu (E) residues or three Asp residues in the catalytic domain which coordinate two divalent ions (Mg$^{2+}$ or Mn$^{2+}$) required for the transposase activity. These elements include $P$ element, members of the Tc1/mariner family, Mu, Tn5, Tn10, IS10 and IS50. Within this group there are various pathways for transposition. The general scheme for the elements of the Tc1/mariner family is shown in Figure 1-2. Mu and Tn3 transposases nick the transposon at its donor location at the 3’ end. After strand transfer the complementary strands of the transposon DNA are synthesised at both new and old locations (Rice & Mizuuchi, 1995). IS3 elements are nicked at one 3’ end only, which attacks the same transposon strand, forming a closed circle. A new transposon strand is synthesised on the circularised template and after two 3’ end nicks and strand transfer the transposon is inserted into a new location (Sekine et al, 1999).

DDE-retrotransposons include Ty1/copia, Ty3/gypsy and Bel (Figure 1-1 b). The product of the pol gene of LTR-retrotransposon has three activities: aspartic protease, reverse transcriptase and DDE-transposase (known as integrase). A copy of the LTR-retrotransposon is transcribed, then the reverse transcriptase synthesizes a DNA strand on the RNA template of the transposon. Integrase catalyzes joining of the terminal CA nucleotides with the target DNA. Finally, the gaps are filled in by the host enzymes (Fujiwara & Mizuuchi, 1988; Kirchner & Sandmeyer, 1996).

### 1.2.2 Target primed retrotransposases

The second mechanism of transposition includes target primed (TP)
Improving mariner transposons for transgenesis

**Figure 1-1 Transposition pathways (Curcio & Derbyshire, 2003).**

Transposases use different pathways for moving transposons into the new locations. Transposon can be cut out or copied out of the donor location and can be pasted in or copied into a new location.

- Transposable element is shown in blue, flanking DNA in green, target DNA in orange, RNA copy in purple, DNA copy in pink. Txn – transcription, RT/En – reverse transcriptase/endonuclease activities.
  - a) DDE transposases cut out and paste in the same copy of transposons;
  - b) DDE retrotransposases copy transposon into an RNA intermediate (copy out). The full cDNA copy is generated prior to pasting into a new location;
  - c) Target primed retrotransposases with reverse transcriptase and endonuclease activities copy out an RNA intermediate, which is integrated into a new location while being transcribed by reverse transcriptase (copy in);
  - d) and e) Tyrosine (Y) transposases have a circular intermediate form which is pasted in by recombination;
  - f) Serine (S) transposases cut out transposon by four concerted cleavages. A circular intermediate is formed, which is pasted in a new location by recombination;
  - g) Rolling circles (or Y2) transposases integrate one DNA strand into a new location either through an intermediate circular stage or without it. The integrated strand serves as a template for synthesis of the second transposon DNA strand in a new location.
Figure 1-2 Scheme of Tc1/mariner transposons cleavages and repair during transposition.

Transposase is not shown. After recognition of the inverted repeats (IRL and IRR) by transposase (a) the first strand cleavage (of the non-transferred strand) occurs by hydrolysis (grey arrows) two or three nucleotides within the transposon ends (b). The second strand (transferred) is cleaved at the junction of the transposon and the flanking DNA by the second hydrolysis reaction (c). The formation of the paired end complex (PEC) is required for the second strand cleavage where two transposon ends are brought together by transposase dimer. After the full excision of the transposon (d), complex of the transposase with transposon (PEC) captures target DNA forming a strand transfer complex (e).

The double stranded break of the donor DNA is repaired creating two or three nucleotide footprint (one of the possible footprints is shown), depending on where the first strand cleavage occurred. The transferred strand of the transposon is joined to the thymidine of the target site by one-step trans-esterification reaction (lavender arrows) (f). The single stranded gaps at a new location are repaired by the host enzymes, forming TA target site duplication (g). Trn – transposon, IRL and IRR – inverted repeats left and right, only three terminal nucleotides of the inverted repeats are indicated. Elements on the scheme are not to scale.
Improving *mariner* transposons for transgenesis

retrotransposases (Figure 1-1 c). These retrotransposases have reverse transcriptase and endonuclease (RT/En) activities. They include mammalian LINEs – non-LTR retrotransposons. Endonuclease, encoded by these elements, makes a nick in the target DNA creating a 3’ end, which serves as a primer for cDNA synthesis on the retrotransposon RNA template. Then endonuclease nicks the top strand of the target DNA and the new cDNA anneals to the target DNA by microhomology at the 3’ end, then the reverse transcriptase synthesizes the second strand of the transposon. The host enzymes fill in the gaps (Feng et al, 1996).

1.2.3 Tyrosine (Y) transposases
Tyrosine transposases (Y-transposases or conjugative transposases) are members of the tyrosine recombinase family, but do not require homology between the transposon ends and the target DNA sequence. The mechanism by which these transposases can accommodate the mismatches between the inverted repeats and the target DNA in their active sites is not yet revealed.

The transposon is excised in a circular form or circular cDNA is synthesized by reverse transcription on an RNA template, which is a transcribed copy of the transposon (Figure 1-1 d and e). During a series of cleavages and exchanges tyrosine serves as a nucleophile and makes covalent bonds with the DNA backbone phosphate. These transposases excise transposon precisely from the donor location and integrate without target site duplications. This mechanism of transposition was shown for Kangaroo, DIRS1 and Tn916 transposons (Caparon & Scott, 1989; Duncan et al, 2002; Goodwin & Poulter, 2001).

1.2.4 Serine (S) transposases
Serine transposases (S-transposases) unlike tyrosine transposases make four concerted cleavages at the each end of the transposon. After four trans-esterification reactions phosphoserine bonds are formed between the 5’ phosphate at the cleavage sites and the hydroxyl group of the conserved serine residue (Smith & Thorpe, 2002). Then 3’-OH nucleophilic attack of the opposite strand on phosphoserine linkages results in a circular intermediate transposon with a two base pair coupling sequence. Serine transposases integrate DNA into a conserved GA dinucleotide
sequence, which is responsible for the polarity of the transposon integration (Wang et al, 2006). The most studied S-transposons are Tn5397, Tn4451 and IS607 (Kersulyte et al, 2000; Lyras & Rood, 2000; Wang et al, 2006).

1.2.5 Rolling-circle (RC) or Y2-transposases

Y2 transposases or rolling circle transposases use the mechanism of transposition known for replication of bacterial plasmids and bacteriophages. These transposons integrate selectively into GAAC or CAAG sequences and do not produce target site duplication upon insertion (Mendiola et al, 1994). These transposases have two conserved tyrosine residues separated by three residues. One tyrosine is required for forming a 5’ phosphotyrosine linkage at the start of the transposon excision/replication and the second tyrosine terminates the DNA synthesis by forming the second 5’ phosphotyrosine bond with the transposon donor DNA. IS91 (Mahillon & Chandler, 1998) and Helitrons (Kapitonov & Jurka, 2001) belong to the Y2-transposases.

1.3 Transposable elements as tools for genetic manipulations

Since transposons can be excised and inserted into a new location carrying a DNA sequence of interest they can be useful for gene therapy – treatment of genetic disorders in patients by switching off or bringing in active genes. Nowadays the most complete and characterised vectors for gene therapy are viral systems. DNA transposable elements were developed as additional tools. Since a fatality after the use of adenoviral gene therapy in 1999 the initial enthusiasm for the viral based vectors was diminished (Raper et al, 2003). The major complications of gene therapy treatment are due to the insertional mutagenesis caused by the retroviral vectors or a severe inflammatory response to a viral protein. Undesirable random integration can lead to activation of the proto-oncogenes or inactivation of tumour suppressor genes. One of four patients treated in 2002 with retroviral vector for X-linked severe combined immunodeficiency developed lymphoblastic leukaemia as one proviral integration occurred within LMO-2 locus, which is a leukaemia oncogene (Hacein-
Bey-Abina et al, 2003). Moreover, the viral particles have a limit in the DNA length that can be transferred – 5 to 10 kb.

Transposable elements can be an alternative to viral vectors in transgenesis. Transposon based vectors can stably integrate fragments of genomic DNA up to 200 kb (Li et al, 2013b). Transposons are easier to engineer and produce, but the challenge is the efficient delivery to the cells as well as precise targeting. The most promising transposons up to date are listed in Table 1-2.

1.3.1 Transposable elements showing promise for transgenesis applications

1.3.1.1 P element

P element was isolated and characterised in 1983 (O'Hare & Rubin, 1983), but the presence of these elements was recognized earlier through the hybrid dysgenesis phenomenon. This causes high rates of mutation in germ line cells of Drosophila after crossing a male strain with autonomous P elements with a female strain lacking them (M strain). Transposition can occur as eggs laid by M strain females do not contain a transcription repressor of the transposase gene in the cytoplasm (Kidwell et al, 1977). P element is 2.9 kb long (87 kDa) and is flanked with 31 bp perfect inverted repeats (Figure 1-3a), it integrates with 8 bp target site duplication (O'Hare & Rubin, 1983). P element prefers to insert within 100 kb of the donor site (50 fold higher rate) than outside of this region (Tower et al, 1993). Excision of the P element is initiated by cleavage at the 3’ end of the transposon DNA at the junction of the inverted repeat and the flanking DNA, while the 5’ end is cleaved 17 bp inside the inverted repeat sequence (Beall & Rio, 1997). After excision, P element leaves 4-7 bp footprints (rarely 14-17 bp) (Gloor et al, 2000).

Transposition of P element in vitro requires GTP as a co-factor to promote protein-DNA assembly (Kaufman & Rio, 1992; Tang et al, 2005). P element requires host factors for its transposition and is specific for Drosophila (O'Brochta D & Handler, 1988). P elements are widely used for mutagenesis in flies (Hummel & Klambt, 2008).
### Table 1-2 Transposon showing promise for transgenesis.

The table lists transposases which are currently used or are promising candidates for transgenesis applications. TSD – target site duplication.

<table>
<thead>
<tr>
<th>Element</th>
<th>From</th>
<th>Family</th>
<th>TSD</th>
<th>Active in</th>
<th>First ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>P element</td>
<td><em>Drosophila melanogaster</em></td>
<td>P</td>
<td>8 bp</td>
<td><em>Drosophila sp.</em></td>
<td>O’Hare &amp; Rubin, 1983</td>
</tr>
<tr>
<td>piggyBac (PB)</td>
<td><em>Trichoplusia ni</em></td>
<td>piggyBac</td>
<td>TAAA</td>
<td>insect, mouse, human</td>
<td>Elick et al, 1996</td>
</tr>
<tr>
<td>Tol2</td>
<td><em>Oryzias latipes</em></td>
<td>hAT</td>
<td>8 bp</td>
<td>zebrafish, frog, chicken, mouse, human cell line</td>
<td>Koga et al, 1997</td>
</tr>
<tr>
<td>Sleeping Beauty (SB)</td>
<td>fish, reconstructed</td>
<td>Tc1/mariner</td>
<td>TA</td>
<td><em>E. coli</em>, fish, frog, mouse, rat, pig, human cell lines</td>
<td>Ivics et al, 1997</td>
</tr>
<tr>
<td>Frog Prince (FP)</td>
<td><em>Rana piniens</em>, reconstructed</td>
<td>Tc1/mariner</td>
<td>TA</td>
<td>fish, hamster and human cell lines</td>
<td>Miskey et al, 2003</td>
</tr>
<tr>
<td>Hsmar1</td>
<td>human, reconstructed</td>
<td>Tc1/mariner</td>
<td>TA</td>
<td>bacteria, human cell lines, zebrafish</td>
<td>Morgan, 1995</td>
</tr>
<tr>
<td>Passport</td>
<td><em>Pleuronectes platessa</em></td>
<td>Tc1/mariner</td>
<td>TA</td>
<td>human and pig cell lines</td>
<td>Clark et al, 2007</td>
</tr>
<tr>
<td>Tn5</td>
<td>bacteria</td>
<td>IS4</td>
<td>9 bp</td>
<td>bacteria, yeast, <em>T. brucei</em></td>
<td>Berg et al, 1975</td>
</tr>
</tbody>
</table>
Improving mariner transposons for transgenesis

**Figure 1-3 Schemes of transposases and their inverted repeats.**
Transposase length in amino acids (a.a.) and the inverted repeats in base pairs (bp) are to scale. Violet – DNA-binding domain, orange – catalytic domain, orange vertical lines – catalytic DDE residues, dark blue – GTP-binding domain, green – nuclear localization domain, dark grey – outer repeat, light grey – inner repeat. Red and blue asterisks indicate which repeats differ by the number of bp indicated between the grey arrows. Horizontal arrows indicate difference within one inverted repeat.
Improving *mariner* transposons for transgenesis

**Figure 1-4 Schemes of transposases and their inverted repeats.**

Transposase length in amino acids (a.a.) and the inverted repeats in base pairs (bp) are to scale. Violet – DNA-binding domain, orange – catalytic domain, orange vertical lines – catalytic DDE/DDD residues, blue – clamp loop, green – nuclear localization domain, dark grey – outer repeat, light grey – inner repeat. Red and blue asterisks indicate which repeats differ by the number of bp indicated between the grey arrows. Horizontal arrows indicate difference within one inverted repeat. Vertical arrows indicate difference between the left and right inverted repeats.
1.3.1.2 PiggyBac

PiggyBac (PB) is a transposon of 2.4 kb with 13 bp perfect outer inverted repeats and 19 bp inner inverted repeats, which are asymmetrically placed with respect to the ends of the transposon (Figure 1-3 b) (Cary et al, 1989). The first sequence described contained a single nucleotide mutation, which interrupted the single open reading frame of the 64 kDa transposase (Elick et al, 1996). The maximum length of DNA sequence that piggyBac can transpose is at least 207 kb, which is the highest reported to date (Li et al, 2013b).

The preferred integration sites are within the expressed genes and associated with an open chromatin structure; 98% of all integrations are at TTAA sites (Li et al, 2013a). PiggyBac generates 1-4 copies per cell clone in HeLa cells and the excisions from the TTAA sites are precise without leaving any footprint (Fraser et al, 1996; Grabundzija et al, 2010).

PiggyBac transposase has been used to transform fibroblasts and melanoma cells to induce pluripotent stem cells by co-expression of Oct4, Sox2, Klf4 and c-Myc factors (Woltjen et al, 2009; Yin & Bi, 2013; Yusa et al, 2009). Transgenic chickens, zebrafish, mouse, rats and bovine embryos have also been generated using the piggyBac system (Katter et al, 2013; Kim et al, 2011; Liu et al, 2013; Lobo et al, 2006).

The advantages of using piggyBac transposase are: excision does not leave a footprint, it is one of the most active transposases tested so far in mammalian cell lines and is able to transpose the biggest DNA fragments (Grabundzija et al, 2013; Li et al, 2013b; Wu et al, 2006b).

1.3.1.3 Tol2

The Tol2 element was discovered in the genome of the medaka fish Oryzias latipes and belongs to hAT superfamily of transposons (Koga et al, 1996). Elements of this family are found in plants, animals and fungi. They contain a C-terminal dimerization domain and integrate with 8 bp target site duplication (Essers et al, 2000; Rubin et al, 2001). The active Tol2 element is 4.7 kb long and contains three
Improving *mariner* transposons for transgenesis

introns. Tol2 transposase is 649 amino acids long (Figure 1-3 c) and binds to the 200 and 150 bp ends of the transposon, containing 12 bp terminal inverted repeats (Kawakami & Shima, 1999). Elements with shorter ends could not be transposed (Urasaki et al, 2006).

Tol2 inserts preferentially into transcription start sites and fewer integration events were found in silenced heterochromatin regions. It does not have a TA target site preference unlike elements of the *Tcl1/mariner* family, so its integrations are random in terms of the sequence specificity compared to the elements currently used for transgenesis. Integrated Tol2 elements are usually flanked by 8 bp target site duplications (Grabundzija et al, 2010).

Generally only 1-3 insertions per genome are observed for Tol2 transposition even at the highest transposase concentration in HeLa cells (Grabundzija et al, 2010). The transposition rates of Tol2 are lower than those PB transposase possibly due to the presence of a nuclear export signal at the centre of the primary transposase structure (Iida et al, 2004).

Tol2 has been shown to transpose in mouse embryonic stem cells, *in vivo* in the mouse liver, in zebrafish, frog, chicken and human cell lines (Balciunas et al, 2006; Kawakami & Noda, 2004; Kawakami et al, 2004; Momose et al, 1999; Offield et al, 2000; Wu et al, 2006a).

### 1.3.1.4 Sleeping Beauty

Sleeping Beauty (SB) is a transposon reconstructed from inactive ancient transposase copies in fish (Ivics et al, 1997). A putative transposase amino acid sequence was predicted from twelve *Tcl1/mariner* elements from the eight species of fish. Inactive elements from the atlantic salmon *Salmo salar* and the rainbow trout *Oncorhynchus mykiss* were joined together. The intact coding sequence including the nuclear localisation signal, DNA-binding and the integration activities were restored by systematic introduction of 24 mutations throughout the whole length of the transposase. The final protein is 340 amino acids long (SB10) and contains three characteristic catalytic residues DD(34)E (Figure 1-3 d). Integrations occur to
generate a TA target site duplication. After DNA shuffling procedure to improve the transposition activity about 2,000 mutated transposases were obtained, among them a hyper active version of the SB was generated with around 120 fold increased activity (SB100X) compared to SB10 (Mates et al, 2009).

The terminal inverted repeats (TIR) of Sleeping Beauty are 200 and 250 bp long, each of the TIR contains inner and outer imperfect direct repeats of 28 and 26 bp, which differ by 6 positions (Cui et al, 2002). Both inner and outer repeats are required for efficient transposase binding, but DNA-binding domain of transposase binds tighter to the inner sequence of the TIR. Four changes have been introduced in the inverted repeats based on the consensus sequence derived from the defective transposons in fish (addition of the second TA in the flanking DNA of the IRR, two single nucleotide changes and one deletion in the region between the direct repeats), which increased the transposition rates almost 4 fold (Cui et al, 2002).

Sleeping Beauty is the first transposon shown to have activity in vertebrates. It is used as a tool for genetic modifications in fish, frogs, mice, rats and somatic tissues of humans (Ivics et al, 2009). It has been successfully used to induce pluripotent stem cells from mouse and pig fibroblasts by introducing Oct4, Sox2, Klf4 and c-Myc genes (Grabundzija et al, 2013; Kues et al, 2013). The advantages of the Sleeping Beauty system are that there are no related sequences in mammalian genomes. It is reported that the most active mutant (SB100X) by the number of colonies is around 4% more active than piggyback (PB) transposon in HeLa cells. By the total number of copies integrated SB100X is 3-6 fold more active than PB and 12-16 times than Tol2 (Kues et al, 2013). Sleeping Beauty generates from 2 to 40 integrations (at the high transposase concentration) per cell clone in HeLa cells. Though at low concentration of transposasases none of the clones had more than two insertions per genome (Grabundzija et al, 2010).

SB is undergoing the first human clinical trials in the United States. The trials involve transfer of the genetically modified T cells in patients with B-lymphoid malignancies. This is planned to be achieved by excision of chimeric antigen receptor from a co-transfected plasmid for transduction of patient peripheral blood.
Improving mariner transposons for transgenesis

(Williams, 2008). Chimeric antigen receptor consists of a single-chain variable fragment of monoclonal antibody fused to endo- and transmembrane domains and is used to induce a monoclonal specificity onto T cells (Pule et al, 2003).

1.3.1.5 Frog Prince

Frog Prince (FP) is a reconstructed transposon from the Northern Leopard Frog (Rana pipies). Only two mismatches had to be corrected in order to match the consensus sequence and to obtain a functional transposase from one of the 8000 copies of this element in the frog genome. Frog Prince transposon has 214 bp inverted repeats (Figure 1-3 e). Each inverted repeat contains two direct repeats, which are 21 bp long and differ in one position. FP belongs to the Tc1/mariner superfamily of transposable elements and integrates with duplication of TA nucleotides. When expressed from a helper plasmid FP transposase causes a 17-fold increase in the number of resistant colonies in HeLa cells compared to the control. After excision FP leaves three nucleotides footprints – CTG or CAG, suggesting that the cleavage of the FP transposon happens with a 3 bp stagger. Frog Prince transposon is active in human, hamster, frog and fish cell lines (Miskey et al, 2003). No transposition activity was observed for FP in the ascidian Ciona intestinalis, when the source of transposase was injected mRNA (Sasakura et al, 2010).

1.3.1.6 Hsmar1

Another reconstructed transposon is Hsmar1, which was found in human as a part of Setmar protein. A Setmar (or Metnase) protein (671 amino acids) is an N-terminal methyltransferase SET domain fused in frame to Hsmar1 transposase (345 amino acids). Setmar is expressed in most human tissues and is involved in the DNA repair: it methylates histone H3 at lysine 4 and 36, which might facilitate access of the DNA repair factor to the damaged DNA. The protein methylase domain is fused to the DNA-recognition domain of the human transposable element Hsmar1 (Lee et al, 2005). All copies of Hsmar1 in the human genome are inactive.

Hsmar1 transposon was reconstructed based on the consensus sequence of the inactive copies of this element. Twenty-one codons were changed in the Hsmar1 ortholog from chimpanzee to obtain an active transposase (Miskey et al, 2007).
Hsmar1 transposon is flanked with 30 bp perfect inverted repeats (Figure 1-3 f). When Hsmar1 is co-transfected with a helper expression plasmid it causes increase in the number of resistant colonies by 23 times in HeLa cells compared to the control (Miskey et al, 2007). Hsmar1 integrates with duplications of TA nucleotides and leaves a 3 bp footprint at the donor location (Liu et al, 2007).

1.3.1.7 Minos
Minos element was found in the fly Drosophila hydei as a 1775 bp element containing a 60 bp intron (Franz & Savakis, 1991). The coding sequence of the Minos transposase is 1023 bp (341 amino acids) with the highest similarity to the Tc1 element from C.elegans. The transposon is flanked by 255 bp perfect inverted repeats (Figure 1-4 g). Each of the inverted repeats contains inner and outer 18 nucleotides directed repeats. Transposase binds to the outer repeats for excision. The role of the inner repeats is not yet known, but both direct repeats are essential for the whole transposition reaction (Pavlopoulos et al, 2007). During transposon excision the first cleavage happens 4 nucleotides within the inverted repeat sequence (Arca et al, 1997).

When mRNA of Minos transposase was co-injected with transposon donor DNA in Drosophila embryos transformation efficiency was 20% higher than with a helper plasmid. From 96 insertions analysed, Minos integrated into introns five times more often. As it was spliced out, integrations did not interfere with the genes functions. Minos does not prefer any specific sequences for integration and it creates a TA target site duplication (Metaxakis et al, 2005).

After excision Minos element leaves 6 bp footprints in Drosophila species, while in mouse model footprints varied from 0 to 10, which might be due to the action of the host repair enzymes to restore the sequence at the donor location (Arca et al, 1997; Zagoraioi et al, 2001).

Minos has been shown to be active in the silkworm Bombyx mori, in the tunicate Ciona intestinalis and in cultured mouse and human cells (Drabek et al, 2003; Sasakura et al, 2003; Shimizu et al, 2000; Zagoraioi et al, 2001). Minos activity has
not been compared directly to other elements, but in HeLa cells the number of resistant clones when helper plasmid is co-transfected is 10 times higher than in the control, 1 to 10 insertions per clone were identified (Klinakis et al, 2000).

### 1.3.1.8 Passport

Passport (Figure 1-4 h) is a naturally active transposon from the plaice *Pleuronectes platessa*, which has 74-81% amino acid identity to the distinct elements in Atlantic salmon and frogs. It is the most active in HeLa cells, with at least 40-fold increase in transposition efficiency compared to the control, when the helper plasmid is used as the source of transposase. The highest transposition activity of the native Passport is two times less than SB11 transposase (the next generation Sleeping Beauty, but before SB100X) in HeLa cells (Clark et al, 2009).

Passport transposase showed a low increase (5 fold) in the transposition efficiency in the porcine endometrial gland epithelium cells compared to the control, when tested alongside with Tol2 (21 fold increase), PB (28 fold), SB11 (10 fold) (Clark et al, 2007).

### 1.3.1.9 Mos1

Mos1 (*mariner*) transposon was first noticed by the occurrence of unstable mosaicism in *Drosophila mauritiana* eye colour (Jacobson & Hartl, 1985). The presence of red colour patches on the peach colour eyes was due to the integration of a transposable element of 1286 bp long with 28 bp imperfect inverted repeats (Figure 1-4 i). This element appeared to be non-autonomous and did not cause eye mosaicism when introduced to other *Drosophila* strain (Capy et al, 1990; Garza et al, 1991). The autonomous Mos1 element giving rise to the red eye mosaicism was found among other copies of *mariner* in *D.mauritiana*, it contained 11 nucleotides substitutions compared to the original copy (Medhora et al, 1991).

Mos1 integrates *in vitro* with duplications of TA nucleotides exclusively when Mg$^{2+}$ is used as a divalent ion. With Mn$^{2+}$, the transposition rate is decreased ~15 fold and target site selection is less stringent (Tosi & Beverley, 2000). Other elements like
Tn10 and Hsmar1 have increased transposition rate with Mn$^{2+}$ ions (Allingham & Haniford, 2002; Liu et al, 2007).

Mos1 has been used to generate stable transgenic lines in *D. melanogaster* (Lidholm et al, 1993), in the mosquito *A. aegypti* (Coates et al, 2000), in the silkworm *Bombyx mori* (Wang et al, 2000), in *C. elegans* (Bessereau et al, 2001), in protozoa *Leishmania major* (Gueiros-Filho & Beverley, 1997), in zebrafish *D. rerio* (Fadool et al, 1998) and chicken (Sherman et al, 1998).

### 1.3.1.10 Mboumar-9

The Mboumar-9 (Mbo9) transposon was discovered in the satellite DNA of the ant *Messor bouvieri*. A single ORF encodes the transposase of 345 amino acids with 68% amino acid identity to Mos1 transposase. Mboumar-9 transposon ends are imperfect 32 bp inverted repeats (Figure 1-4 j) (Munoz-Lopez et al, 2008).

Mboumar-9 transposase fused to the maltose binding protein has been shown to be capable of transposon cleavage and integration into target plasmid DNA *in vitro*. All the integrations in the presence of Mg$^{2+}$ occurred with TA target sites duplication, while with Mn$^{2+}$ other sites were also duplicated. Manganese reduced the *in vitro* transposition rate of Mbo9 ten fold. No target sequence preference other than the TA has been observed for Mbo9 integrations (Munoz-Lopez et al, 2008). Mboumar-9 has not been tested for activity except in *E.coli*.

### 1.3.1.11 Tn5

Tn5 is a bacterial transposon flanked by 1.4 kb inverted repeats, which are themselves IS50 transposable elements (Berg et al, 1975). There are 19 sequential critical base pairs annotated within these 1.4 kb, which are recognised by Tn5 transposase (Figure 1-4 k). Tn5 is a bacterial transposon, which carries three antibiotic resistances (kanamycin, streptomycin and bleomycin) and encodes two proteins (Mazodier et al, 1985). One is an active full-length form of transposase and the second, called inhibitor (Inh) lacks the first 55 amino acids. The C-terminal of the Tn5 transposase contains a dimerization domain. The inhibitor dimerizes with a full-length protein creating defective dimers, which cannot recognize the inverted repeats of Tn5 and perform transposition. Expression of the Inh protein is increased by Dam
methylation of DNA and the highest transposition rates are observed in dam strains (Yin & Reznikoff, 1987). Tn5 transposon prefers a high G/C content in target DNA and integrates with a duplication of 9 nucleotides (Green et al, 2012; Lodge et al, 1988).

Purified Tn5 transposase can be electroporated as a DNA-protein complex into a number of bacterial strains (Goryshin et al, 2000). Tn5 is a tool for sequencing applications as well as making chromosomal integrations in vitro and in vivo by electro transfection (Reznikoff & Goryshin, 2000). It has been successfully used in 52 strains of gram negative bacteria, in 18 strains of gram positive bacteria, in archea, *S.cerevisiae* and *T.brucei* (www.epibio.com).

### 1.3.2 Target site selection

Transposable elements differ by how they select a new location for the excised transposons. Whereas elements like Himar1, Mos1, Minos, Frog Prince and Mboumar-9 transposons integrate in TA targets, no sequence specificity around this dinucleotide sequence has been found for these elements (Crenes et al, 2009; Le Breton et al, 2006; Metaxakis et al, 2005; Miskey et al, 2003; Munoz-Lopez et al, 2008). Some elements like Tn10 integrate randomly within genomes, but require specific 6 nucleotides target site GCTNAGC (Halling & Kleckner, 1982).

The piggyBac transposon inserts preferentially into transcriptional units, CpG islands (much higher for primary human T-cells, than for HEK293 and HeLa cells), and transcriptional start sites (TSS). It recognises TTAA target site, but does not have any preferences for AT rich regions near the recognition site. It has been reported that piggyBac has decreased levels of integration into TSS of proto-oncogenes (Galvan et al, 2009).

Detailed studies of the target site selection of Sleeping Beauty transposase did not reveal a unified mechanism of how the transposon finds its new location. Most likely AT-rich regions and high bendability of DNA serve as attractive sites for Sleeping Beauty transposase (Liu et al, 2005).
Bacterial transposons Tn3 and Tn4652 have more pronounced preference for AT-rich regions, they integrate into TA(T/A)TA or T(A/T)(T/A)(T/A)(A/T) sites respectively (Davies & Hutchison, 1995; Kivistik et al, 2007). Other transposons like Tn5 are more specific in target site selection. They do not only integrate into specific 9 nucleotides (GNTYWRANC), but also generate clusters upon integration separated by approximately 5 nucleotides, which means that the region around the selected 9 nucleotides should be suitable for more than one integration event (Goryshin et al, 1998). Notably these specific DNA target sequences in most cases are palindromic (Linheiro & Bergman, 2008). Integration near the donor location is called ‘local hopping’ and is shown for other elements like Sleeping Beauty and piggyBac (Carlson & Largaespada, 2005; Wang et al, 2008).

Tn7 recognises a specific attTn7 site (64 nucleotides) and integrates near it duplicating 5 nucleotides (CCCGC) upstream of the recognition site (DeBoy & Craig, 2000; Waddell & Craig, 1989). On the other hand Tol2 transposon does not show preference for any sequence for its integration, but prefers regions close to transcriptional start sites (Grabundzija et al, 2010).

Transposons like Kangaroo or DIRS1, which are integrated by recombination do not create a target site duplication upon insertion, as there is no staggered cleavage of the DNA occurring prior to integration (Goodwin & Poulter, 2001).

**1.3.3 Possibilities of precise targeting of integration**

Transposase with precise and changeable target specificity would be an ideal candidate for genome therapy, as it would minimise the effect of random integrations. The use of the transposases fused to the sequence specific DNA-binding domains in order to guide integration to the specific sites was first proposed more than 10 years ago (Kaminski et al, 2002), but the first attempts of fusing DNA-binding domains to the Sleeping Beauty transposase resulted in complete or partial loss of the transposase activity (Wilson et al, 2005).

Introduction of the flexible linker (KLGGGAPAVGGGPK) between the transposase and the DNA binding domain might have solved the problem of SB inactivation.
upon fusion. The DNA-binding domain of the adenovirus-associated Rep protein was fused to the N-terminal of SB, PB or Tol2 transposases with the addition of the flexible linker. The Rep protein recognises specific sequences – Rep recognition sequences (RRS) – in the human genome. These consist of GAGC repeats. Both PB and SB retained 80% of activity compared to the unmodified transposases. The highest enrichment (15 fold) of the insertions close to the RRS was observed for the chimeric Sleeping Beauty transposase. Fusion to the N-terminal of Tol2 transposase retained less than 10% of the native protein activity, indicating that Tol2 is sensitive to protein fusions (Ammar et al, 2012).

The C-terminal fusions of the zinc finger DNA-binding domain from mouse transcription factor Zif268 to ISY100 transposase resulted in ~50% of integrations into the TAs adjacent to a single Zif268 binding site with 10-30% of wild type ISY100 activity (Feng et al, 2010).

PiggyBac transposase was fused to the Gal4 DNA binding domain at its N- or C-terminus and retained transposition activity. Gal4 is a yeast transcription activator protein, which recognises and specifically binds to the Upstream Activating Sequence (UAS) (Brand & Perrimon, 1993). PB with N-terminal Gal4 fusion targeted the TTAA near the UAS 10% more efficiently than C-terminal fusion. Integrations near UAS happened 30% more often with the N-terminal Gal4 PB, than with the native PB transposase during a plasmid into plasmid integration in HEK293 cells (Owens et al, 2012).

1.3.4 Ways of delivering transposases to cells
Transposases can be delivered to the cells as purified protein, helper plasmid or mRNA. Purified Himar1, Mos1 and Mboumar-9 transposases have been used to study transposition \textit{in vitro} (Lampe et al, 1996; Munoz-Lopez et al, 2008; Pelicic et al, 2000). Helper plasmid is an expression vector, carrying a functional transposase gene, which is expressed in the recipient cells. Helper plasmid and mRNA are used to study transposition \textit{in vivo} either into the target plasmid DNA, which is co-injected or into the genomic DNA of the cells.
Ex vivo delivery of genes can be achieved by electroporation, nucleofection or magnetofection of the transposon and expression plasmid into the target cells in culture. The expression of the transposase from the helper plasmid declines over several days, especially in dividing cells (VandenDriessche et al, 2009). This way of DNA delivery was used in keratinocytes from six patients with junctional epidermolysis bullosa with Sleeping Beauty system. The selected cells were able to regenerate human skin on immunodeficient mice, which showed stable restoration of the expression of deficient epidermal protein (Ortiz-Urda et al, 2003).

In vivo gene delivery can be achieved by polyethylenimine (PEI) into the lungs or brain tumours (Liu et al, 2006; Ohlfest et al, 2005). PEI is a polymer with cationic properties. It condenses DNA into positively charged molecules, which are brought into cells by endocytosis (Boussif et al, 1995). Intravenous injection and hydrodynamic tail vein injection in mice has been successful for both SB and Tol2 transposases (Balciunas et al, 2006; Bell et al, 2007; Ohlfest et al, 2005). Hydrodynamic injection has been successfully used on pigs for delivering plasmid DNA in liver cells (Yoshino et al, 2006).

1.4 Regulation of transposition
Since transposition causes genome rearrangements, high levels of transposase activity are not beneficial for the fitness of an organism. Transposition levels have to be controlled in order for an organism to survive and reproduce. There are several mechanisms of transposition regulation.

1.4.1 Overproduction inhibition
One of the mechanisms regulating transposition is overproduction inhibition (OPI). It was observed, when Mos1 transposase was put under the heat-shock-protein-70 promoter, expression of higher amounts of transposase by heat-shock resulted in decreased levels of Mos1 excision (by 37%) (Lohe & Hartl, 1996). Possible explanations for this phenomenon are that oligomeric forms of transposase can serve as an inhibitor, or oligomerization at higher concentrations of transposase is promoted in the absence of DNA, or that formed aggregates of transposase are inactive.
Mos1, Sleeping Beauty, Himar1 and Passport transposases have been shown to be affected by overproduction inhibition (Clark et al, 2009; Lampe et al, 1998; Yant et al, 2000), whereas Tol2, piggyBac are less sensitive (Kawakami & Noda, 2004; Wu et al, 2006a).

**1.4.2 Dominant-negative complementation**

Dominant-negative complementation (DNC) was noticed, when male flies with defective Mos1 transposase gene (by ethylmethane sulfonate mutagenesis) were crossed with females with active Mos1 transposase and the excision rates in progeny was reduced. This could possibly be due to the formation of hetero-oligomers of active and mutated transposases, which are inactive, or due to the occupation of Mos1 transposon ends by inactive transposase making them inaccessible for an active protein (Lohe & Hartl, 1996).

**1.4.3 pi-RNA**

pi-RNA or piwi-RNA is a class of small non-coding RNAs, which interact with piwi proteins. Piwi are regulatory proteins required for cell division and maintenance of incomplete differentiation in stem cells. Pi-RNAs are 23-31 nucleotides long and are generated in spermatogenic cells by Piwi-mediated cleavage of single stranded transposable element transcripts. Since pi-RNAs are antisense to transposon sequence they are thought to be involved in transposon silencing (Klattenhoff & Theurkauf, 2008).

Other small RNAs are viRNAs and miRNA. viRNAs are generated from exogenous viral RNA, whereas miRNAs are derivatives from the endogenous Pol II transcripts. viRNA play role in antiviral defence mechanism and miRNA are involved in gene silencing. miRNAs are produced by Dcr1 cleavage and are associated with Ago1 protein, while viRNAs are processed by Dcr2 and are found in complex with Ago2 protein (Obbard & Finnegan, 2008). Both sense and antisense to transposable elements small RNAs associated with Dcr2 and Ago2 have been found in *Drosophila*, which might play roles in the control of the number of transposable elements (Chung et al, 2008).
1.4.4 CpG methylation

CpG methylation of DNA increases transposition by 10-30 times for SB, Minos and FP transposons. These elements are flanked with long inverted repeats, containing direct repeats and more than one transposase-binding sites (Figure 1-3 b, c and d). There was no effect of CpG methylation observed for the elements flanked with short inverted repeats – Tc1, Himar1 and Hsmar1 (54, 27 and 30 bp inverted repeats respectively). The effect of CpG methylation is possibly due to compaction of chromatin structure and enhancing synapsis between the inverted repeats and transposase, as the transposase binding properties were not affected (Jursch et al, 2013).

1.5 Tc1/mariner elements transposition

Transposition of the Tc1/mariner elements does not require host factors and can occur both in different cell lines and different organisms (Lampe et al, 1996; Tosi & Beverley, 2000). Members of the Tc1/mariner superfamily belong to the DDE/D type of transposases. They contain three highly conserved residues in their catalytic domain, which are involved in the coordination of two Mg\(^{2+}\) or Mn\(^{2+}\) ions (Plasterk et al, 1999; Richardson et al, 2006). It has not been established which of these two ions (or both) transposase utilises in vivo. In vitro studies have shown that target site selection is less stringent when Mn\(^{2+}\) is present in the reaction as a co-factor (Lampe et al, 1998; Tosi & Beverley, 2000; Zhang et al, 2001). Integration of these elements generates a target site duplication of two nucleotides, usually TA – which is a distinguishing feature of the transposons of the Tc1/mariner family.

1.5.1 Mechanism of Mos1 transposition

The insight into the mechanism of Mos1 transposition was gained after solving the crystal structure of the catalytic domain of Mos1 transposase (Richardson et al, 2006). The full-length Mos1 in the absence of DNA formed crystals of the catalytic domain only. In the crystal structure two manganese ions were coordinated by three conserved aspartic residues (156, 249 and 284). This finding suggested that Mos1 uses two-ion catalysis mechanism for cleaving DNA strands (Richardson et al, 2006; Yang et al, 2006).
The full-length Mos1 structure was obtained in 2009, when the transposase was co-crystallized with the inverted repeat DNA (Richardson et al, 2009). Mos1 transposase is a protein of 345 a.a. (Figure 1-5), which contains an N-terminal DNA-binding domain (residues 1-112). The DNA-binding domain consists of two helix-turn-helix motifs (residues 8-53 and 74-110) and is joined to the C-terminal catalytic domain through a linker (residues 113-120). Catalytic domain contains a clamp loop (residues 162-189), which is involved in protein-protein interactions (Richardson et al, 2009).

The transposition reaction is initiated by the transposase recognition of the specific sequences, inverted repeats at ends of the transposons. The current hypothesis is that the transposase protein binds to the inverted repeats in a form of a dimer (Auge-Gouillou et al, 2005). The non-transferred strand (the bottom strand) of the transposon is cleaved first during hydrolysis at the 5’ end of the transposon three nucleotides inside the inverted repeat (Dawson & Finnegan, 2003; Plasterk et al, 1999) and may not dependent on the formation of the paired end complex (PEC), when two transposon ends are brought together. No intermediate hairpin stage was observed for the Tc1/mariner elements (Dawson & Finnegan, 2003) (Figure 1-6).

Transposition of Tn5, IS10, IS50 and piggyBac happens through a hairpin formation (Mitra et al, 2008; Reznikoff et al, 2004). The first cleavage is in the transferred strand through hydrolysis. This exposes 3’-OH end of the transposon, which attacks the non-transferred strand of the transposon by trans-esterification reaction forming a hairpin. The second hydrolysis cleaves the hairpin to release the 3’-OH end of the transposon for strand transfer to occur (Bhasin et al, 1999). Tn5 excision occurs precisely at the transposon-flanking DNA junction (Goryshin & Reznikoff, 1998). In the case of piggyback the exposed 3’-OH end of the transposon attacks the non-transferred strand four nucleotides outside of the transposon end, forming a TTAA hairpin. This allows piggyBac transposase to excise without leaving a footprint (Mitra et al, 2008). Hermes – a transposon of hAT family from the housefly Musca domestica – also transposes through a hairpin intermediate. Hermes transposase cleaves the non-transferred strand first. The 3’-OH of the flanking DNA attacks the
Improving *mariner* transposons for transgenesis

**Figure 1-5 Scheme of Mos1 transposase domains (a) and sequences of the inverted repeats (b).**

(a) The scheme of Mos1 transposase ORF. DNA-binding domain (lavender) contains two helix-turn-helix motifs (HTH1 and HTH2), linker (green), catalytic domain (light orange) and clamp loop (blue). Three catalytic aspartic residues (positions 156, 249 and 284) are orange lines.

(b) Sequence of Mos1 inverted repeat left (IRL) and right (IRR). The inverted repeats are 28 bp and differ by four positions: 3, 11, 13 and 28 (marked by asterisk). NTS – non-transferred strand, which is cleaved first; TS – transferred strand, which is cleaved second and is joined to the thymidine of the target site. Lower case – duplicated TA of the target site.

(c) Annealed oligonucleotides of the top and bottom strands of Mos1 IRR. The top strand (transferred) is 3 nucleotides longer (28mer), than the bottom strand (non-transferred, 25mer). Three nucleotides overhang at the 3’ end is an analogue of the cleaved out inverted repeat end as cleavage of the bottom strand occurs three nucleotides inside the transposon sequence.
Figure 1-6 Different mechanisms of the first and second strand cleavages (Richardson et al, 2009).

Tn5 makes a nick on the transferred strand (TS) first, forming a 3’-OH, which attacks the non-transferred strand (NTS) forming a hairpin on the transposon end. By contrast Hermes cleaves the non-transferred strand first and the hairpin is formed on the flanking DNA. Mos1 transposes without a hairpin formation by two successive hydrolysis reactions followed by a trans-esterification strand transfer. Two Mg²⁺ ions, which are coordinated by DDE/DDD catalytic residues, stabilize the substrate and activate nucleophile (H₂O).
Improving *mariner* transposons for transgenesis

non-transferred strand forming a hairpin on the flanking DNA (Figure 1-6) (Zhou et al, 2004).

The second strand cleavage of the Mos1 inverted repeats occurs by the second hydrolysis reaction after two transposon ends are brought together in a PEC (Dawson & Finnegan, 2003). After the transposon leaves its original location a target DNA is captured. The 3’ hydroxyl at the end of transferred strand attacks the 5’ of a TA target sequence as a nucleophile. The transferred transposon strand is joined to the thymine of the target site in a trans-esterification reaction. The single-stranded gaps are then filled by the host repair enzymes, which results in a TA target site duplication (Figure 1-2 e-g). After the excision, transposon leaves three nucleotides footprint (Plasterk et al, 1999).

The crystal structure of Mos1 transposase bound to the two annealed oligonucleotides equivalent to the right pre-cleaved inverted repeat DNA (Figure 1-5 c) is the first structure of the paired end complex of an eukaryotic transposase to have been determined (Figure 1-7) (Richardson et al, 2009). The structure showed that these two ends of DNA are brought together in a trans orientation, in which the DNA binding domain of one monomer recognises the inverted repeat that is placed in the catalytic domain of the other monomer. The transposase dimer is held together by two separate interfaces – two HTH1 motives interact with each other and the clamp loop of one monomer interacts with the linker region of the other monomer. The sequence specific interactions are spread throughout the transposase length. In the HTH1 Lys44 and Arg48 are involved in sequence specific recognition, in the HTH2 these are Gln100 and Gln101. Arg118 (linker), Arg183 and Pro184 (clamp loop) and His293 (catalytic domain) are involved in the sequence specific interaction between the inverted repeat DNA and transposase (more details in the Table 4-1).

The crystal structure of the PEC of Mos1 is a snap-shot of a dynamic transposition process after transposon excision and before the strand transfer has taken place. The orientation of the transposase subunits during the first and the second strand cleavages is still unclear.
Figure 1-7 Orientation of Mos1 transposase monomers in the paired end complex.
(a and b) Two orthogonal views of the crystal structure of Mos1 bound to two inverted repeat (IR) DNA ends in a paired end complex in a trans orientation. DNA binding domains of transposase A hold IRₐ, while its catalytic domain is occupied by the IRₐ. The IR DNA was obtained by annealing of two oligonucleotides (28 and 25 bp) of the IRR of Mos1. This creates staggered ends as if excision has already happened three nucleotides within the transposon end. Monomers of transposase are held together by two interfaces. The first interface is between HTH₁ₐ and HTH₁₉, the second one is between the clamp loop of one monomer and the linker region of the other monomer.
There are at least five theoretically possible orientations of the subunits for the first and the second strand cleavages (Figure 1-8). It is proposed that there is no a single domain or motif in Mos1 transposase which is required for the transposase oligomerization (Auge-Gouillou et al, 2005; Zhang et al, 2001). Himar1 and Sleeping Beauty transposase have been shown to contain a dimerization motif in their N-terminal end (Butler et al, 2006; Izsvak et al, 2002). In contrast, Tn5 utilizes the C-terminal domain for dimerization, but exists as a monomer when not bound to DNA (Davies et al, 1999; Mahnke Braam et al, 1999). P element exists in solution as a preformed tetramer (Tang et al, 2007), while Hermes is active in the form of a hexamer (Hickman et al, 2005).

1.6 Difference in the inverted repeats

Transposable elements are flanked with specific DNA sequences which transposase recognizes and binds to. These sequences are in the inverted orientation (inverted repeats), which is essential for transposition to occur in the right orientation. Excision of Tn5 element flanked with direct repeats is reduced 1000 times compared to the transposon with the inverted repeats (Egner & Berg, 1981). The terminal inverted repeats of some transposons, like Himar1 and Hsmar1 are identical, while others are imperfect. The significance of these differences is not yet fully understood. Since the inverted repeats of some elements are non-identical there might be difference in the transposase binding, cleavage or/and integration of such elements.

Mos1 inverted repeats differ at four positions (Figure 1-5 b) and it has been shown that the DNA-binding domain of transposase binds 10 times stronger to the right inverted repeat than to the left (Zhang et al, 2001). In vitro transposition of the element flanked with two right inverted is $12.5 \times 10^3$ (Auge-Gouillou et al, 2001b) or 50 (Pledger et al, 2004) times more efficient than the native left-right form. Similar results were obtained for Tn5 transposon, for which the minimal DNA sequence required for transposition was found to be 19 bp that differ by 7 positions between the left and right inverted repeats. Tn5 transposase binds to two ends with different affinity and a mosaic end has been constructed, having three nucleotides from one of the inverted repeats and four from the another, which showed 100 fold
Figure 1-8 Possible orientations of the transposase domains during the first and second strand cleavages.

Theoretical orientations of the two transposase monomers (HTH1 and HTH2 are small circles, catalytic domain is a big circle) during the first strand cleavage (a) and the second strand cleavage (b) depending on which motives are involved in dimerization. The first strand cleavage does not require synopsis of two inverted repeats (black triangles), that is why only one inverted repeat DNA is shown for the first strand cleavage (a). Second strand cleavage occurs when two inverted repeats are brought together (b). Trans orientation is the crossed orientation of the transposase subunits (No. 5 and 10). The schemes differ in the dimerization domain and cis or trans orientation of the Mos1 monomers. Scheme (10) is the orientation in which Mos1 transposase is in the crystal structure of the paired end complex.
increased activity in vivo in E.coli (Zhou et al, 1998). In contrast, changing any of the four non-matching nucleotides of Mos1 inverted repeat (Figure 1-5 b) decreased the DNA-binding efficiency measured by the gel retardation assay compared to the right inverted repeat (Auge-Gouillou et al, 2001b). The most significant change was observed for the base A13, which in the crystal structure of the paired end complex is not involved in the base specific recognition. Bases 12 to 15 of the inverted repeat form a minor groove to which amino acid residues 65 to 68 of the transposase bind deeply by shape complementarity (Richardson et al, 2009).

The inverted repeats of Sleeping Beauty are 200 and 250 bp, both of them contain outer (28 bp) and inner (26 bp) direct repeats. A transposon flanked with two left repeats is 2.9 times more active than the native form of transposon in vivo in HeLa cells, while a transposon with two right repeats is 8 times less active (Izsvak et al, 2002).

Additionally, not only the sequence of the inverted repeats is important for transposase recognition, but the distance between the inner and outer inverted repeats. piggyBac has perfect inner inverted repeats of 13 bp and 19 bp perfect internal repeats, which are separated by 3 bp in the left repeat and by 31 bp in the right repeat (Figure 1-3 b). Transposase binds to the right repeat more efficiently, than to the left one (Mitra et al, 2008).

There are several possible explanations for the difference in the inverted repeat sequences of transposons and why the naturally found forms are usually not the most active. The first is that the transposition reaction may have to be coordinated: after the first cleavage happens at one end, the second end is captured and cleaved afterwards. The second end has to stay unoccupied to be able to form a paired end complex. If the interactions of transposase with two transposon ends are the same, both ends will be bound to transposase and formation of the paired end complex will be obstructed.

The second explanation is that transposition has to be maintained at a suboptimal level as it generates genome rearrangement and increase mutation rates. Organisms
with high mutation rate in the genomic DNA will be less likely to leave healthy progeny.

The third explanation is that mutations in the inverted repeats accumulated with time and in future might lead to complete inactivation of these elements. Thus the elements with different inverted repeats are partially inactivated and are on their way to complete silence.

1.7 Aims of the project

The aim of this project was to take further our understanding of the mechanism of transposition of the Tc1/mariner transposable elements by comparative study of the two closely related transposases – Mos1 and Mboumar-9. We expressed and purified Mos1 and Mboumar-9 transposases in order to examine their biochemical properties, oligomeric state, cleavage, integration and the whole transposition activities in vitro. Both transposases were found to be active in vitro and exist as a dimer in solution. We found an uncharacterised ability of purified Mos1 and Mboumar-9 transposases, based on which we developed a new in vivo activity assay. We investigated the significance of the inverted repeat differences by creation of transposons with identical and mutated inverted repeats. We found that for both Mos1 and Mboumar-9 transposons one of the inverted repeats is cleaved more efficient than another and the last base of the inverted repeat might play an important role at the stage of transposon integration.
Improving *mariner* transposons for transgenesis
Chapter 2: Purification and biophysical characterisation of the transposases
2.1 Introduction

The Tc1/mariner family of transposable elements is an extremely widespread group of transposons in eukaryotic cells from fungi to human. To date nine elements of this family have been described to be active. Four of these are naturally active: Mos1 from Drosophila mauritiana (Medhora et al, 1991), Famar1 from the earwig Forficula auriculata (Barry et al, 2004), Mboumar-9 from the ant Messor bouvery (Munoz-Lopez et al, 2008) and Passport from the plaice Pleuronectes platessa (Clark et al, 2009). The other five active elements have been reconstructed – Sleeping Beauty from the fish Danio rerio (Ivics et al, 1997), Himar1 from the horn fly Haematibia irritans (Lampe et al, 1996), Frog Prince from the frog Rana pipiens, Minos form the fly Drosophila hydei and Hsmar1 from human, (reviewed by Ivics, Z., 2009).

Famar1, Passport, Sleeping Beauty, Frog Prince and Minos transposases were shown to be active when expressed from the helper plasmid inside the transfected cells. Recombinant Mos1, Mboumar-9, Himar1 and Hsmar1 transposases were able to transpose elements in vitro (Coates et al, 2000; Lampe et al, 1996; Miskey et al, 2007; Munoz-Lopez et al, 2008). Mboumar-9 and Hsmar1 were purified and characterised in a form fused to maltose binding protein at the N-terminus, while Mos1 and Himar1 were studied as untagged proteins. Transposases of the Tc1/mariner family are dependent on the presence of inverted repeats, divalent ions (Mg$^{2+}$ or Mn$^{2+}$) and target DNA for transposition reaction in vitro.

2.2 Results

We have aligned the amino acid sequences of the active mariner elements using the T-Coffee web server and predicted the protein phylogeny with RAxML web server. This software finds the maximum-likelihood phylogeny of either nucleotide or protein sequence alignments. We visualized the phylogenetic tree with FigTree v1.4.0 to predict the most closely related transposases (Figure 2-1). Mos1 and Mboumar-9 transposases showed the highest amino acid similarity to each other from all nine active elements. We used this high sequence similarity to enable swapping of the protein domains between two transposases, to gain an alternative
Figure 2-1 Phylogenetic tree of active Tc1/mariner transposases

The tree shows the diversity of the amino acid sequences of the active transposases of mariner family described to date. The length of the branches indicates the diversity of the amino acid sequence of each transposase from the recently common ancestral node. Mos1 and Mboumar-9 transposases are the most closely related active transposases.
specific recognition of the inverted repeat DNA. The reasons for the domain swap are described in Section 2.2.10.

Surprisingly, the Minos element from Drosophila hydei is positioned far on the phylogenetic tree from the Mos1 element from Drosophila mauritiana. This can be due to the fact that Minos is a reconstructed element or/and has accumulated inactivating mutations after the two Drosophila species have been evolutionary separated.

2.2.1 Optimisation of the Mos1 and Mboumar-9 genes
The open reading frame of Mos1 as well as Mboumar-9 transposase is 1035 bp. The genes found in nature carry 24 codons that are rare in E.coli. To enhance the expression levels of the eukaryotic proteins in bacterial cells, we have codon optimised both sequences (GenScript). This improved the Codon Adaptation Index (CAI) (Sharp & Li, 1987) from 0.63 to 0.96. The ideal CAI for expression in E.coli system is 1.0; CAI>0.8 corresponds to a predicted high level of expression. We also made a number of silent mutations to introduce 17 unique sites for restriction endonucleases (Figure 2-2), which enable of swapping the fragments between the Mos1 and Mboumar-9 DNA sequences to generate chimeric transposases.

2.2.2 Cloning of Mos1 and Mboumar-9 genes
Artificially synthesized (GeneArt) Mos1 and Mboumar-9 genes (1035 bp) carried a Thr216Ala mutation to improve the protein solubility (Richardson et al, 2004). The genes were cloned by NdeI and XhoI restriction sites. To express the 6His-N-terminal forms of the proteins, the genes were cloned into a pET28a vector (Novagen). The untagged proteins were expressed from a pET30a vector (Novagen). Expression vectors pET28a and pET30a carried a kanamycin resistance cassette.

2.2.3 His-tagged Mos1 and Mboumar-9 purification.
E.coli BL21 (DE3) strain transformed with expression plasmids induced protein expression as described on p. 157. We found that incubation of Mboumar-9 transposase expressing cells at 18°C results in a higher yield of soluble protein (data not shown).
Improving mariner transposons for transgenesis

Purification and biophysical characterisation of the transposases

Figure 2-2 Alignment of the nucleotide sequences of the codon optimised genes of Mos1 and Mboumar-9 transposases

Identical nucleotides shown on the violet background. The red rectangles indicate the unique restriction enzyme sites, common for Mos1 and Mboumar-9 sequences, enabling protein parts to be swapped between the transposases.
Cell disruption of the N-terminal 6His-Mos1 and 6His-Mboumar-9 was performed as described before (Richardson et al, 2004). Imidazole was added to the filtered supernatant to the final concentration of 10 mM. The sample was applied onto an IMAC HiTrap FF 1 ml column at the flow rate of 3 ml/min on ÄKTA HPLC (GE Healthcare) system (see the details in the Material and Methods section). The protein peak was eluted at 370 mM imidazole for Mos1 (Figure 2-3) or at 410 mM for Mboumar-9 transposase (Figure 2-4).

The yield after the affinity step was ~10 mg per litre for each of the proteins. Peak fractions have been pooled together, concentrated on Vivaspin 6 5,000 MWCO PES (GE Healthcare) at 4°C down to <500 µl volume. The samples were further purified by gel filtration on Superdex 200 30/100 GL column at the flow rate of 0.5 ml/min at 4°C (Figure 2-6 and Figure 2-7). On the Coomassie stained gel of the 6His-Mboumar-9 transposase we could observe the products of protein degradation after exposure to room temperature, before the samples were applied on the SDS-PAGE (discussed further in section 2.2.6). The final protein yield was not measured as the fractions after the affinity step were used for optimisation of the cation exchange step. Cation exchange chromatography did not result in higher protein purity. The estimated final yield after the affinity purification followed by gel filtration is 5-10 mg of protein from one litre of the induced culture.

The thrombin (1 or 5 Units) cleavage of the 6His tag was performed at 4°C and 16°C over a time course of 0.5-2 hours (Figure 2-7) and over night (not shown). The expected band shift of 1.8 kDa was not observed. Selective binding of 6His-transposases to the Ni²⁺ charged column proves the presence of the 6His tag. The unsuccessful treatment with thrombin protease can be due to the possible folding of the 6His-tag, making the thrombin site inaccessible for protease cleavage.
Improving mariner transposons for transgenesis

Figure 2-3 Affinity chromatography of 6His-Mos1 transposase
(a) Elution profile of 6His-Mos1 transposase from HiTrap IMAC FF 1 ml column.
(b) Elution profile of the peak fractions after the affinity chromatography.
(c) Fraction of soluble proteins before the IMAC column (before), flow through (X1), M – molecular marker, Mos1 – purified Mos1 transposase.
(d) Peak fractions 1-12 as indicated in (c) visualised by Coomassie stained SDS-PAGE.
Improving *mariner* transposons for transgenesis

Figure 2-4 Affinity chromatography of 6His-Mbo9 transposase

(a) Elution profile of 6His-Mbo9 transposase from HiTrap IMAC FF 1 ml column.

(b) Elution profile of the peak fractions after the affinity chromatography.

(c) Fraction of soluble proteins before the IMAC column (before), flow through (X1), M – molecular marker, Mos1 – purified Mos1 transposase.

(d) Peak fractions 1-12 as indicated in (c) visualised by Coomassie stained SDS-PAGE.
Improving mariner transposons for transgenesis

Figure 2-5 Size exclusion chromatography of 6His-Mos1
(a) The elution profile of 6His-Mos1 from Superdex 200 30/100 GL column, 4°C, 0.5 ml/min. The elution volumes of the two peaks maximum are indicated in blue.
(b) Fractions 1-9 as indicated in (a) visualised by Coomassie stained SDS-PAGE, M – molecular mass marker.

Purification and biophysical characterisation of the transposases
Improving *mariner* transposons for transgenesis

**Figure 2-6** Size exclusion chromatography of 6His-Mboumar-9

(a) The elution profile of 6His-Mboumar-9 from Superdex 200 30/100 GL column, 4°C, 0.5 ml/min. The elution volume of the peak maximum is indicated in blue.

(b) Fractions 1-9 as indicated in (a) visualised by Coomassie stained SDS-PAGE, M – molecular mass marker.
N-terminal 6His-Mos1 is resistant to cleavage with thrombin

Figure 2-7: Attempt to cleave the 6His-tag with thrombin protease

6His-Mos1 (43 kDa) was incubated with 1 or 5 Units of thrombin protease for 30 or 120 min at two different temperatures – 4°C or 16°C. The expected 1.8 kDa band shift was not observed.
2.2.3.1 Activity of the His-tagged proteins
Since we did not manage to cleave the 6His-tag from the transposase, we tested if the N-terminal 6His-tag changed the protein activity. To check the precision of the transposon cleavage we performed cleavage of the first and second strands of the inverted repeat DNA by Mos1 transposase using linear substrates (Figure 2-8 a). The 28 bp of the inverted repeats are flanked with TA nucleotides followed by random DNA sequence, creating a 100 bp substrate. Transferred strand (TS) is the transposon strand, which is joined to the target DNA by the transposase, whereas the non-transferred strand (NTS) is joined to the target DNA by host repair enzymes. Precise cleavage of the NTS should result in the release of a labelled product containing 33 nucleotides. Cleavage of the TS was performed on the “pre-cleaved” substrate, where three DNA oligonucleotides were annealed together to mimic the cleavage of the first strand. The products of the reaction were separated on an 8% denaturing polyacrylamide gel. Either the top or the bottom strands of DNA were labelled with the 5’ IRDye® 700 fluorophore enabling visualisation of the resulted products on the LI-COR System. As the transposase has non-specific nuclease activity, additional unexpected bands were observed (Figure 2-8 b). The results show that the cleavage patterns for tagged and untagged proteins are not similar. The expected products are 33 nucleotides and 70 nucleotides for the first and second strand cleavages respectively. The major product of the 6His-Mos1 cleavage is not 33 nucleotides for the first strand cleavage or 70 nucleotides for the second strand cleavage. We presume that, since the 6His-tag is fused to the N-terminus of the transposase, it might interfere with the DNA binding properties of the transposase disabling the precise transposon excision. For the next experiments in this study we used untagged transposases unless otherwise stated.

2.2.4 Untagged Mos1 and Mboumar-9 expression and purification
Expression was performed as described in Section 2.2.3. Purification of Mos1 transposase was carried out essentially as described before (Richardson et al, 2004). The yield of Mos1 transposase was improved from 3 mg to 9 mg of protein per 1 litre of culture when using E.coli C43 (DE3) as an expression strain.
Improving *mariner* transposons for transgenesis

**Figure 2-8 Comparison of the first and second strand cleavage activity of Mos1 vs 6His-Mos1**

(a) Schemes of the first and second strand cleavages of Mos1 IR from the flanking DNA.

(b) 8% denaturing polyacrylamide DNA gel of the products of transposase cleavage. M – DNA length marker, nts – nucleotides, no trp – negative control, no transposase was added to the reaction.

The expected band size for the first strand cleavage is 33 nucleotides, which is the major product after the Mos1 transposase incubation. For 6His-Mos1 transposase the major product is around 75 nucleotides, which indicates unspecific cleavage of the substrate. For the second strand cleavage the expected product is 70 nucleotides, which is observed for Mos1 transposase. 6His-Mos1 generated the major product of around 40 nucleotides, which is not specific for the transposon ends.
Purification of Mboumar-9 transposase was developed and optimized in this study. Cells expressing Mboumar-9 transposase were pelleted for 1 hour at 8,000 g, 4°C. The cell pellet was resuspended in 10% weight/volume ratio in Mboumar-9 cell resuspension buffer (see p. 137) and incubated for 1-2 hours (rocking) at 4°C. Cell suspension was passed through a 0.8 mm needle prior to Cell Disruptor cell disintegration. Cell debris was pelleted for 1 hour at 50,000 g, 4°C. Supernatant was filtered through a 5 µm filter followed by a 0.45 µm filter before loading onto POROS 20HS cation exchange column (PerSeptive Biosystems). Elution of Mboumar-9 transposase was performed with ascending gradient of NaCl from 400 mM to 1000 mM (Figure 2-9).

Mboumar-9 transposase peak eluted at 46 mS/cm conductivity, which corresponds to approximately 620 mM NaCl. Peak fractions were pooled together and concentrated in a Vivaspin 6 5,000 MWCO PES concentrator (GE Healthcare) at 4°C to <500 µl volume. After centrifugation at 13,000 rpm for 10 min to remove aggregates, protein solution was injected onto Superdex 200 30/100 GL size exclusion chromatography column and a peak of 13.71 ml retention volume was collected (Figure 2-10), concentrated and snap frozen in liquid nitrogen, stored at -80°C.

We have compared the stability of Mboumar-9 transposase after storage on ice, in 50% glycerol at -20°C or snap frozen to freshly purified protein. The highest protein folding rate with minimum degradation products was observed, when Mboumar-9 transposase was snap frozen and thawed in cold water (data not shown).

2.2.5 Establishing molecular mass of the proteins in solution.
There are several techniques enabling the molecular mass of particles in solution to be measured. Finding the oligomeric state of the protein in solution (monomer, dimer or tetramer) is a basis for understanding protein-protein interactions in different conditions as well as providing an insight for the mechanism of protein action.

2.2.5.1 Size exclusion chromatography
Size exclusion chromatography allows separation of proteins in a sample according to the hydrodynamic radius of each particle species. Retention volume or elution
Figure 2-9 Cation exchange chromatography of Mbo9 transposase
(a) Elution profile of Mbo9 transposase from POROS 20HS column with ascending gradient of NaCl. The protein peak elutes at conductivity of 46 mS/cm.
(b) Protein fractions before and after (1-3) the cation exchange chromatography as indicated in (a) visualised by Coomassie stained SDS-PAGE, M – molecular mass marker.
Improving *mariner* transposons for transgenesis

Figure 2-10 Size exclusion chromatography of Mboumar-9 transposase
(a) Elution profile of the Mboumar-9 transposase from Superdex 200 30/100 GL column, 4°C.
(b) Protein fractions before and peak fractions after (1-8) the size exclusion chromatography as indicated in (a) visualised by Coomassie stained SDS-PAGE, M – molecular mass marker.
Improving *mariner* transposons for transgenesis

volume is volume of eluent from the point of sample injection to the middle of the protein peak and is a measure of the hydrodynamic radius. Retention volume can be correlated with the molecular mass of the protein if the particles are globular spheres. Proteins of irregular shape have larger hydrodynamic radius and elute as if they were globular particles bigger than their actual mass.

Having a set of standard proteins of a known mass enables plotting of a calibration curve. Using this curve and a retention volume of the protein of unknown mass can give an estimation of the mass of the unknown protein particles in given conditions.

We were using Superdex 200 30/100 GL column, which has a fractionation range of \( M_r \) 10,000 to 600,000 Da for globular proteins. Each gel filtration column consists of a stationary phase – resin, and a mobile phase – void volume. To build a calibration curve we established the void volume \( (V_o) \) of the column (by measuring the elution volume of Blue Dextran 2,000), the geometric column volume \( (V_c) \) is 23.56 ml for Superdex 200 30/100 GL and the elution volume \( (V_e) \) for each of the standard globular proteins (Figure 2-11 a). The elution volume parameter \( (K_{av}) \) (Equation 1) is defined as (Figure 2-11 b).

\[
K_{av} = \frac{V_e - V_o}{V_c - V_o}
\]  

Equation 1

There is a linear relationship between \( K_{av} \) and \( \log_{10} \) of the protein molecular mass (Figure 2-11 b). By plotting a calibration curve (Figure 2-11 c) we can estimate the molecular mass of a protein, knowing its elution volume.

Mos1 transposase elutes at 13.72 ml and Mboumar-9 at 13.86 ml (Figure 2-12 a). These retention volumes correspond to the estimated molecular mass of 97.9 kDa and 91.7 kDa respectively. The estimated protein mass would equal the native protein mass only if the proteins of interest were globular. The predicted molecular mass of the monomer for both Mos1 and Mboumar-9 transposase is 41 kDa. The protein masses predicted from the calibration curve indicate that both proteins are likely to be dimeric in solution \((41 \times 2 = 82 \text{ kDa})\) but more elongated in shape compared to the globular standard proteins.
Improving mariner transposons for transgenesis

Figure 2-11 Calibration curve for standard globular proteins

(a) Elution profile of the mixture of the standard globular proteins – dark blue, and blue dextran – light blue from Superdex 200 30/100 GL, 4°C.
(b) Table showing molecular masses of standard globular proteins and their retention volumes.
(c) Calibration curve for five globular proteins.
Figure 2-12 Size estimation of Mos1 and Mboumar-9 transposases in solution
(a) The overlaid elution profiles of Mos1 and Mboumar-9 transposases from Superdex 200 30/100 GL column at 4°C. Using a calibration curve for the standard globular proteins, the predicted mass of Mos1 transposase is 97.9 kDa and Mboumar-9 transposase 91.7 kDa.
(b) The size distribution by volume of Mos1 and Mboumar-9 transposases measured by Dynamic Light Scattering at 4°C. The estimated molecular mass based on the hydrodynamic radius is 96 kDa for Mos1 and 110 kDa for Mboumar-9 transposase.
2.2.5.2 Dynamic light scattering

Dynamic light scattering is a method of measuring the hydrodynamic radius of particles in solution. Particles smaller than the wavelength of the laser scatter light in all directions (Rayleigh scatter). As the particles in solution undergo Brownian motion the scattering constantly changes with time. This scattering depends on temperature, solution viscosity and the particles’ size. If a sample is homogeneous we can work out the hydrodynamic radius of the particles and estimate its molecular mass (as if they were in a shape of an ideal sphere).

Both Mos1 and Mbourmar-9 transposases were purified by cation exchange and gel filtration chromatography and any aggregates were spun down at 13,000 rpm for 10 min at 4°C prior loading onto Zetasizer APS instrument in the concentration of 0.5-1 mg/ml. All measurements were performed in 3 biological and 3 technical replicas at 4°C. The results presented in Figure 2-12 (b) were averaged from 9 measurements. Both samples were homogenous and the estimated molecular mass from the hydrodynamic radius of the particles was 96 kDa for Mos1 transposase and 110 kDa for Mbourmar-9 transposase. This observation confirms results obtained from size exclusion chromatography, that both proteins are likely to be elongated dimers at a given concentration in solution at 4°C.

2.2.6 SEC-MALLS

Size-exclusion chromatography coupled with multi angle laser light scatter enables measuring of the absolute mass of the molecules in solution. A sample of interest is separated on a gel filtration column, and the Rayleigh scatter is detected at 18 different angles at each time point of the elution chromatogram. The Rayleigh ratio is a function of the molecular mass, which is averaged for the particles eluting at the same time point of the chromatogram. The average molecular masses were determined using ASTRA™ software (Wyatt Technologies). Superdex 200 30/100 GL column was linked with ÄKTA Ettan (GE Healthcare), samples were eluted at 0.5 ml/min flow rate at room temperature.

Proteins were injected in the final volume of 100 µl at the concentration of 1.8 mg/ml for Mos1 and 2.3 mg/ml for Mbourmar-9 transposase. In this experiment
Improving *mariner* transposons for transgenesis

we used 6His-tagged transposases (predicted molecular mass 43 kDa).

Mos1 transposase eluted in a single peak with the average molecular mass of 85.2 kDa (Figure 2-13). This corresponds to the size of the protein dimer – 43 x 2 = 86 kDa. The peak observed was not symmetrical possibly due to the presence of 12% glycerol in a sample.

Mboumar-9 transposase eluted in three major peaks with a minor peak of aggregates. The first peak of 89.3 kDa we consider as a protein dimer. The last peak of 45 kDa could be a protein monomer (43 kDa) or a dimer of two catalytic domains (54 kDa). The middle peak of 65.9 kDa is an intermediate stage, which could be a full length protein dimerized with another catalytic domain. Because of the high back pressure the eluted fractions could not have been collected and therefore analysis on a denaturating gel to visualise the components of the protein fractions could not have been carried out.

It was shown before that Mos1 transposase can undergo proteolysis after Arg118 during crystallization (Richardson et al, 2006). We presume that a similar degradation process has happened to Mboumar-9 transposase at room temperature during the time of elution (40 minutes). The products of degradation of approximate mass of 15 and 25 kDa can be observed in the SDS-PAGE shown in the Figure 2-6 (b) for 6His-tagged Mboumar-9 and Figure 2-9 (b) for untagged protein. From this we conclude that the proteolysis happening regardless the presence of the 6His-tag.

**2.2.7 Thermal Denaturation Assay**

A thermal denaturation assay measures the temperature of protein unfolding, or denaturation. The protein sample is mixed with a fluorescent dye and is placed in a thermocycler to which a detector is linked (BioRad IQ5 ICycler). While the temperature of the sample rises, causing the protein to unfold, the detector registers the increase of fluorescence as the dye binds to the exposed hydrophobic areas of the unfolded protein. The point of the maximum rate of change in fluorescence
Improving mariner transposons for transgenesis

Figure 2-13 SEC-MALLS of Mos1 and Mboumar-9 at room temperature on Superdex 200 30/100 GL column
(a) Elution profile of Mos1 transposase. Protein elutes as a single peak of 85 kDa, corresponding to a protein dimer.
(b) Elution profile of Mboumar-9 transposase. At room temperature the protein degrades to 45 kDa particles from a 89 kDa full-length dimer of transposase with an intermediate stage of 66 kDa.
Improving _mariner_ transposons for transgenesis

Corresponds to the melting temperature \( T_m \) of a protein. Comparison of the \( T_m \) of a given protein in different conditions is a measure of protein stability, folding and binding properties.

We examined the stability of Mos1 and Mboumar-9 transposases in the presence and absence of divalent ions as well as inverted repeats DNA. We used the right inverted repeat of Mos1 and the left inverted repeat of Mboumar-9 for the reasons explained later in Chapter 5. Terminal inverted repeats with a staggered 3’ ends (Figure 2-14), mimicking the cleaved inverted repeat ends were annealed by slow cooling from 95°C to room temperature over night. Proteins were used at a 5 µM concentration, DNA concentration was 7.5 µM and MgCl₂ concentration was 5 mM, where added. After 30 minutes incubation on ice fluorescent dye Sypro Orange (ex/em 485 nm/575 nm, SIGMA) was added to the final concentration of 5x. Any possible aggregations were spun down by 10 minutes centrifugation at 13,000 rpm. Each experiment consisted of 3 x 45 µl reactions in a 96 well plate. Thermal denaturation assay was performed using a gradient of temperatures from 4°C to 95°C, with 1°C step and 30 seconds hold (Bio-Rad).

Analysis of the thermal stability shows that Mos1 as well as Mboumar-9 transposases are stabilised in the presence of Mg²⁺ ions and DNA (Figure 2-15). The melting peak temperature \( T_m \) is shifted by 4.0°C for Mos1 and by 2.7°C for Mboumar-9 upon addition of 5 mM MgCl₂ to the reaction. This fact indicates that the proteins are folded tighter and higher temperature is required for their denaturation. Addition of the “pre-cleaved” inverted repeat DNA ends increases the melting temperature, by an extra 5.0°C for Mos1 and 2.0°C for Mboumar-9 transposase. Our data shows that both proteins can bind the inverted repeat DNA in the absence of Mg²⁺ ions in solution (melting peak shifts by 2.0°C for Mboumar-9 and by 2.7°C for Mos1 transposase compared to the proteins on their own). We also observed a positive peak of protein folding at 22.5°C. Possibly this temperature indicates the ideal condition for protein functioning. Both proteins are originally from cold-blooded animals, flies and ants. This can be an explanation why the highest protein folding rate is at the temperatures close to the environmental.
Figure 2-14 The inverted repeats of Mos1 (a) and Mboumar-9 (b) with 3 nucleotides stagger
The annealed oligonucleotides of the right inverted repeat of Mos1 (a) and the left inverted repeat of Mboumar-9 (b) transposons have 3 nucleotides stagger (shown in a violet or green rectangular), to imitate the product of the cleavage of the transposon inverted repeats by transposase.
Improving *mariner* transposons for transgenesis

Purification and biophysical characterisation of the transposases

Figure 2-15 Thermal stability of Mos1 and Mboumar-9 transposases in the presence of DNA and Mg$^{2+}$ ions

Thermal denaturation assay of Mos1 (a) and Mboumar-9 (b) transposases. Addition of Mg$^{2+}$ shifts the $T_m$ by 2.7°C for Mboumar-9 and 4.0°C for Mos1, which indicates the stabilisation effect of Mg$^{2+}$ ions on the transposases. Further stabilisation can be achieved by adding of the inverted repeat DNA with a staggered end.

(c) A table of the melting temperatures of Mos1 and Mboumar-9 transposases.
2.2.8 Dimerization domain of Mos1

We confirmed that Mos1 transposase is a dimer in solution. Using small-angle X-ray scattering data we showed that Mos1 transposase is an elongated protein with maximum dimension of 185 Å (Cuypers et al, 2012). On the other hand in the crystal structure of the paired end complex the maximum length of the protein dimer is 110 Å (Richardson et al, 2009).

This suggests that in solution transposase is in a different conformation, than seen in the crystal structure of the paired end complex. To find out the dimerization domain of Mos1 transposase we deleted the N-terminal 55 amino acids (containing HTH1), creating Δ55 Mos1 transposase (Figure 2-16 a). If this form is a dimer in solution the dimerization domain is not within the first 55 amino acids. If Δ55 Mos1 protein is a monomer, the dimerization domain might lay within the HTH1.

2.2.8.1 Δ55 Mos1 construction, expression and purification.

The mutant Δ55 Mos1 protein gene was amplified with primers No. 7 and 10 (list of primers p. 140) on the template of codon optimised Mos1 gene. The resulting PCR product of 895 bp was cloned into pET30a expression vector (Novagen) by NdeI and XhoI sites for untagged protein expression. The protein of 34.4 kDa was expressed and purified (Figure 2-16 b) using purification method for Mos1 transposase described before (Richardson et al, 2004).

2.2.8.2 SEC-MALLS of Δ55 Mos1 mutant

To find out the mass of the Δ55 Mos protein in solution, we performed SEC-MALLS as described before (section 2.2.6). The measured mass in solution indicated that Δ55 Mos1 is a monomer in solution (Figure 2-16 c). The predicted protein mass is 34.3 kDa, the observed – 33.7 kDa. This suggests that the protein dimerization domain might lay within the first 55 amino acids.

The reciprocal experiment of establishing the mass of the first 150 amino acids of the Mos1 transposase was performed (Cuypers et al, 2012). It showed that protein containing N-terminal 150 amino acids of Mos1 is present in solution in a form of a
Improving *mariner* transposons for transgenesis

**Figure 2-16 Construction and SEC-MALLS analysis of the Δ55 Mos1 truncated transposase**

(a) A scheme of the Mos1 transposase truncation. The N-terminal 55 amino acids of Mos1 transposase were deleted to create Δ55 Mos1 transposase.

(b) A Coomassie stained SDS-PAGE showing the purified Mos1 (41 kDa) and Δ55 Mos1 (34 kDa).

(c) Elution profile of the Mos1 and Δ55 Mos1 transposases coupled with multi angle laser light scatter to establish the absolute mass of the proteins in solution. Δ55 Mos1 transposase is 34 kDa, which corresponds to a monomer. Mos1 transposases is 85 kDa in solution, which corresponds to the size of the protein dimer.
Improving mariner transposons for transgenesis

dimer with a small proportion of tetramer. Together these results confirm that the
dimerization of Mos1 involves HTH1.

2.2.9 The first and second strand cleavages of Mos1 and Mboumar-9 inverted repeats

Cleavage of the inverted repeats from the flanking DNA during excision of the transposon happens by cleavage of the NTS first three nucleotides within the transposon followed by cleavage of the TS at the junction of the inverted repeat end with the flanking DNA. The DNA strands were labelled with 5’ IRDye® 700 fluorophore, which enabled visualisation of the cleavage products using the LI-COR system.

The first strand cleavage was analysed by labelling the bottom strand at the 3’ end. The precise cleavage should result in the release of a product of 33 nucleotides. We incubated the inverted repeat of Mos1 with increasing concentrations of Mos1 or Mboumar-9 transposases and visualised the products of cleavage after separation on 7.5 M Urea 8% polyacrylamide denaturating gel (Figure 2-17). Cleavage by Mos1 transposase resulted in two bands of 33 and 29 nucleotides as was observed before for Mos1 (Richardson et al, 2009). Mboumar-9 transposase cuts the inverted repeat of Mos1 with major bands of approximate length of 68 and 38 nucleotides (estimated from the gel). Notably, there is a 33 nucleotides band produced by Mboumar-9 cleavage of Mos1 inverted repeat. This could suggest that Mboumar-9 transposase might specifically cut the NTS of Mos1 inverted repeat with low efficiency.

We analysed the reciprocal experiment of cutting the first strand of Mboumar-9 inverted repeat by each of the transposases (Figure 2-18). In this experiment Mboumar-9 also gave bands of 33 and 29 nucleotides. This could mean that Mboumar-9 also cuts within three nucleotides of the transposon end as Mos1 does on the Mos1 inverted repeat. The third band was observed after the Mboumar-9 cleavage of the Mboumar-9 end – of approximately 51 nucleotides long. A similar band does not appear when Mos1 transposase cleaves Mos1 IR. This band could be either a linear product if Mboumar-9 cleaves the sequence within its own inverted repeat or a denatured hairpin if Mboumar-9 uses the hairpin formation mechanism.
Improving *mariner* transposons for transgenesis

**Figure 2-17 First strand cleavage of Mos1 IR**

Specific cleavage of the non-transferred strand of Mos1 inverted repeat within the 3 nucleotides of the inverted repeat releases a product of 33 nucleotides long (a). The products of cleavage are visualised on the 8% polyacrylamide denaturating DNA gel (b). M – DNA length marker, nts – nucleotides, no trp – negative control, no transposase was added to the reaction. The band of the expected size is present only when Mos1 IRR is cleaved by Mos1 transposase. Incubation of the Mos1 IRR with Mboumar-9 transposase results in two major bands with estimated length of 68 and 38 nucleotides.
Figure 2-18 First strand cleavage of Mboumar-9 IR
Specific cleavage of the non-transferred strand of Mboumar-9 inverted repeat within the 3 nucleotides of the inverted repeat releases a product of 33 nucleotides long (a). The products of cleavage are visualised on the 8% polyacrylamide denaturating DNA gel (b). M – DNA length marker, nts – nucleotides, no trp – negative control, no transposase was added to the reaction. Cleavage with the Mboumar-9 transposase results in three major bands of 29, 33 nucleotides (similarly to Mos1 cleaving the NTS of Mos1 IRR) and an additional band of 51 nucleotides. Incubation of the Mboumar-9 IR with Mos1 transposase results in unspecific cleavage pattern with the major product of 77 nucleotides.
This can be checked by labelling of both DNA strands or an internal radioactive label. The second strand cleavage was performed by labelling of the top strand at the 5’ end of the inverted repeat. The precise cleavage should result in release of 70 nucleotides from Mos1 inverted repeat or 60 nucleotides from Mboumar-9 inverted repeat as the flanking DNA of the Mboumar-9 inverted repeat is 10 nucleotides shorter, than of the Mos1 inverted repeat.

Three oligonucleotides were annealed to produce a pre-cleaved inverted repeat end three nucleotides within the transposon end. Cleavage of the Mos1 inverted repeat by Mos1 transposase resulted in the presence of the major band of 70 nucleotides as expected (Figure 2-19). Mboumar-9 transposase cleavage of Mos1 inverted repeat did not result in a specific 70 nucleotides band, but the cut happened efficiently within the inverted repeat end producing a band of an approximate size of 62 nucleotides (Figure 2-19).

Mboumar-9 transposase cut the Mboumar-9 inverted repeat specifically, producing 60 nucleotides band as the main product (Figure 2-20) and a band of 52 nucleotides, which can be generated after the cleavage within the sequence of the inverted repeat. Mos1 transposase cleavage of the Mboumar-9 inverted repeat resulted in a band of estimated 65 nucleotides, which could be produced by cutting within flanking DNA sequence at the 3’ end of the TS.

We have generated a combined map of the Mos1 and Mboumar-9 inverted repeats with their flanking DNA indicating the possible cleavage sites of the most efficient cleavages by each of the transposases. In Figure 2-21 the Mos1 inverted repeat is on the violet background and Mos1 transposases cuts are indicated as violet arrows. Mboumar-9 inverted repeat is on the green background and Mboumar-9 transposase cuts are indicated as green arrows. This map assumes that all the products were generated as linear single stranded DNA oligonucleotides. Using this method we could only visualise the products of the closest to the 5’ end label cut of the DNA produced.
Specific cleavage of the transferred strand of Mos1 inverted repeat releases a product of 70 nucleotides long (a). The products of cleavage are visualised on the 8% polyacrylamide denaturing DNA gel (b). M – DNA length marker, nts – nucleotides, no trp – negative control, no transposase was added to the reaction. The characteristic band of 70 nucleotides is the major product of Mos1 transposase cleavage. Mboumar-9 transposase cuts Mos1 IR efficiently releasing the unexpected product of 62 nucleotides long.
Improving *mariner* transposons for transgenesis

**Figure 2-20 Second strand cleavage of the Mboumar-9 IR**

Specific cleavage of the transferred strand of Mboumar-9 inverted repeat releases a product of 60 nucleotides long (a). The products of cleavage are visualised on the 8% polyacrylamide denaturating DNA gel (b). M – DNA length marker, nts – nucleotides, no trp – negative control, no transposase was added to the reaction. The characteristic band of 60 nucleotides is the major product of Mboumar-9 transposase cleavage. The second band of 53 nucleotides is observed for Mboumar-9 transposase. Mos1 transposase does not produce the 60 nucleotides specific product.
Improving mariner transposons for transgenesis

Purification and biophysical characterisation of the transposases

Figure 2-21 Cleavage sites within the inverted repeats and flanking DNA of Mos1 and Mbounar-9

The right inverted repeat of Mos1 on the violet background (a) and the left inverted repeat of Mbounar-9 on the green background (b) within the flanking DNA. The arrows indicate the locations of transposases cleavage: violet for Mos1 transposase and green for Mbounar-9 transposase. The cleavage positions were estimated based on the size of the cleavage products with the highest intensity from the gels in the Figures 3.16 – 3.19.
2.2.10  Exchange of the HTH2 domain between the transposases

The crystal structure showed that the two monomers of Mos1 transposase are in *trans* orientation, when bound to the nicked inverted repeat DNA (Richardson et al, 2009). This structure captures the state of the transpososome (complex of the transposase and inverted repeat DNA) after two cleavages have occurred, leaving unrevealed the orientation of the first and second strand cleavages in the process of transposition.

We found that both Mos1 and Mboumar-9 transposases are elongated dimers in solution and that the first helix-turn-helix domain is involved in the protein dimerization. We presumed it would be possible to swap only the second helix-turn-helix domain (HTH2) between the two proteins (Figure 2-22). This could enable transposase to dimerize by the identical HTH1 domain, but retain the specific recognition of the different inverted repeat. Knowing that the chimeric transposase recognises Mboumar-9 inverted repeats, we could introduce an inactivating mutation in the catalytic domain (DDD to DDA) to abolish the protein cleavage activity. To find out the orientation of the transposase monomers during the cleavage of the inverted repeat we could mix an inactivated chimeric protein with the active Mos1 transposase. If the Mboumar-9 inverted repeat is cleaved, we could conclude that the cleavage happened is in the *trans* orientation of the transposase monomers (Figure 2-22 b).

2.2.10.1  Construction and purification

A SacI/PstI fragment of 144 bp (amino acids 76 to 124 – HTH2) of the codon optimised Mboumar-9 transposase gene was introduced into Mos1 transposase instead of the corresponding fragment in Mos1, creating a Chim7 transposase. To abolish the protein activity we introduced Asp284Ala mutation of the catalytic aspartic acid residue. This mutation should inactivate the cleavage activity of the Chim7 transposase, so only if the cleavage of the first or second strand happens in the *trans* orientation the specific cleavage products should be observed. The Chim7 and Chim7 DDA protein were expressed and purified using the Mos1 purification protocol described before (Richardson et al, 2004).
Improving *mariner* transposons for transgenesis

**Figure 2-22 Design of the chimeric Mos1 transposase for cleavage experiments**

(a) A scheme of the domain organisation of Mos1 and Mboumar-9 transposases. Two DNA-binding helix-turn-helix domains (HTH1 and HTH2) are followed by C-terminal catalytic domain with three essential aspartic residues (DDD).

(b) A scheme of the inverted repeat DNA cleavage by a heterodimer of the inactive chimeric transposase and the active Mos1 transposase. The chimeric protein (Chim7) recognises Mboumar-9 inverted repeats, but can not perform the cleavage due to the Asp284Ala mutation (DDA) in the catalytic domain. If the Mboumar-9 inverted repeats are cleaved specifically, the cleavage is happening in *trans* orientation of two transposase monomers.
2.2.10.2 Cleavage activity of the chimeric protein

The substrates for cleavage were used as described before in Section 2.2.9. After one hour incubation of three different concentrations of transposase with the first strand cleavage substrate, products of the reaction were separated on a 7.5 M Urea 8% polyacrylamide gel (as described on p. 147).

Analysis of the gel in Figure 2-23 showed that cleavage of the Mboum9 inverted repeat by Mboum9 transposase resulted in three major bands. A band of the expected specific cleavage of 33 nucleotides, and two bands around 29 nucleotides (which was observed for Mos1 transposase earlier in Figure 2-17) and 51 nucleotides. Incubation of Mboum9 inverted repeat with Mos1 transposase resulted in a high number of bands possibly due to unspecific digest. The chimeric protein (Chim7), with a HTH2 domain of Mboum9 had the same specificity as Mos1 transposase. We did not observe any bands specific to Mboum9 transposase cleavage performed by Chim7. As predicted the Asp284Ala mutation (Chim7 DDA) abolished the protein activity in vitro and no cleavage products were observed.

As the first strand cleavage of the Mboum9 inverted repeat by the Chim7 transposase occurred in the Mos1 transposase pattern, we wanted to check if swapping the HTH2 domain to Mboum9 still retained the specificity of Mos1 transposase in a chimeric protein. To do so we used a nicked inverted repeat of Mos1 transposase as by our assumption cleavage of the second strand has to be more specific, than of the first strand (Dawson & Finnegan, 2003). The expected specific cleavage product of 70 nucleotides was clearly the major product of Mos1 cleavage (Figure 2-24). Incubation of the Mos1 inverted repeat substrate with chimeric transposase did not result in a 70 nucleotides band. This would indicate that the chimeric protein lost its specificity towards the Mos1 inverted repeat.

If our assumption is right and cleaving of the second strand on a nicked substrate is easier for transposase, maybe the chimeric transposase would have more specific activity for the second strand cleavage of Mboum9 transposon. To check this hypothesis we performed cleavage of the second strand of Mboum9
Figure 2-23 First strand cleavage of the Mboumar-9 IR

a) A scheme of the first strand cleavage of the Mboumar-9 inverted repeat from a random flanking DNA sequence. The 5' end of the non-transferred DNA strand (NTS) is fluorescently labelled. The product of the specific cleavage is 33 nucleotides.

(b) A denaturing polyacrylamide gel of the DNA after incubation with increasing concentration of transposases. M – DNA length marker, nts – nucleotides, no trp – negative control, no transposase was added to the reaction. The product of the specific cleavage is only present when the Mboumar-9 inverted repeat is incubated with Mboumar-9 transposase.
Figure 2-24 The second strand cleavage of Mos1 IR

(a) A scheme of the second strand cleavage of the Mos1 inverted repeat from a random flanking DNA sequence. The 5’ end of the transferred DNA strand (TS) is fluorescently labelled, the non-transferred strand (NTS) is mimicking the first strand cleavage has occurred 3 nucleotides within the transposon.

(b) A denaturating polyacrylamide gel of the DNA after incubation with increasing concentrations of Mos1 or Chim7 transposases. M – DNA length marker, nts – nucleotides, no trp – negative control, no transposase was added to the reaction. The product of the specific cleavage of 70 nucleotides is absent when the chimeric transposase is incubated with Mos1 inverted repeat DNA.
Improving mariner transposons for transgenesis

transposase (Figure 2-25). The expected band of estimated 60 nucleotides was the major product of Mbourmar-9 transposase cleavage. Both Mos1 and Chim7 transposases did not produce a characteristic band of 60 nucleotides, indicating that Chim7 transposase retained the specificity towards Mos1 inverted repeats. Change in one of the DNA recognition domains (HTH2) did not bring in the Mbourmar-9 IR specificity.

Notably, Figure 2-25 (b) showed that with the increasing concentration of Mbourmar-9 transposase its cleavage activity drops – the intensity of the 60 bp band decreases with the increased concentration of Mbourmar-9 transposase, showing non linear correlation of the activity vs concentration.

2.3 Discussion

We modified nucleic acid sequences of Mos1 and Mbourmar-9 transposases to be able to exchange parts of the proteins and decreased the number of rare E.coli codons to improve the yield of recombinant proteins. We have cloned and purified untagged and 6His-tagged forms of Mos1 and Mbourmar-9 transposases. Thrombin protease did not cleave the N-terminal 6His-tag, suggesting that the tag is folded in a way, which disables protease access to the cleavage site.

The results of thermal denaturation assay showed a peak of positive protein folding at 22.5°C. It is worth comparing the efficiency of the transposases in vitro at this temperature with the temperature for the established protocol – 30°C (Kaufman & Rio, 1992). The standard temperature for maintenance of Drosophila cell line is 23-25°C, which is close to the temperature we found for positive protein folding, indicating that this can be the optimal temperature for protein functioning. We would expect incubation of in vitro transposition reactions at 23-25°C to be more physiological and possibly show higher protein activity. In this study we were using the incubation temperature of 30°C as was published previously for Mos1 transposase (Zhang et al, 2001).

We showed that the first 55 N-terminal amino acids are essential for protein dimerization in solution. Himar1 and Sleeping Beauty transposases have also been
Improving mariner transposons for transgenesis

Figure 2-25 The second strand cleavage of Mboumar-9 inverted repeat (IR)

a) A scheme of the second strand cleavage of the Mboumar-9 inverted repeat from a random flanking DNA sequence. The 5’ end of the transferred DNA strand (TS) is fluorescently labelled, the non-transferred strand (NTS) is mimicking the first strand cleavage 3 nucleotides within the transposon.

b) A denaturing acrylamide gel of the DNA after incubation with increasing concentrations of Mboumar-9, Mos1 or Chim7 transposases. M – DNA length marker, nts – nucleotides, no trp – negative control, no transposase was added to the reaction. The product of the specific cleavage of 60 nucleotides is absent when Mos1 or Chim7 transposase is incubated with the Mboumar-9 inverted repeat DNA.
Improving mariner transposons for transgenesis

shown to interact via their N-terminal regions and form a tetramer or a dimer in solution respectively (Butler et al, 2006; Izsak et al, 2002).

Mos1 and Mboumar-9 transposases elute at the retention volume lower than it is predicted by calculated molecular mass of a dimer, suggesting that these proteins have an elongated shape in solution in the absence of DNA. The crystal structure of Mos1 transposase bound to two molecules of DNA in a paired end complex showed that the protein is in a compact, non-elongated form (Richardson et al, 2009). This means that when transposase binds DNA a conformational change happens to bring two inverted repeat ends together in a paired end complex. This conformation might happen either when a transposase dimer is bound to one of the inverted repeats or when both monomers are interacting with the inverted repeats before the compact paired end complex is formed.

The Chim7 transposase containing the HTH2 domain from Mboumar-9 transposase did not show specificity to perform the first or the second strand cleavages of Mboumar-9 inverted repeat. This can be explained by the fact that HTH1 of Mos1 is also involved in the sequence specific recognition of the inverted repeat DNA sequence. That is why the protein with HTH1 of Mos1 and HTH2 of Mboumar-9 transposase did not show specificity for the Mboumar-9 inverted repeat DNA. The cleavages of the Mboumar-9 inverted repeats by the chimeric transposase happened in the unspecific manner, replicating the pattern of Mos1 transposase cleavage. Recognition of the inverted repeat containing a nick was expected to be more specific, but Chim7 cleaved the second strand of the Mboumar-9 inverted repeat non-specific as well. This showed that only one DNA binding domain can not change the specificity towards the inverted repeats recognised by the transposase.

We could try swapping the HTH1 to check if this can change the specificity of the Chim7 cleavage. This will show if the catalytic domain is also involved in sequence recognition of the inverted repeats or it works as an unspecific nuclease.

Analysing the cleavage efficiency of Mboumar-9 transposase we noticed that with the increase of transposase concentration the efficiency of cleavage drops. The event of decrease of transposase activity with the increase of transposase concentration was
observed before for Mos1 transposon (Lohe & Hartl, 1996). This mechanism of regulation of transposase activity – overproduction inhibition, is shown for other mariner elements (Wu et al, 2006b). Here we showed the evidence that Mboumar-9 also follows overproduction inhibition mechanism of regulation. More experiments are required to quantify the transposition efficiency vs. transposase concentration in vitro.
Chapter 3: *In vivo* transposition
3.1 Introduction

There are a number of established methods to examine transposase activity. *In vitro* assays include studying of the transposase/DNA binding properties, cleavage activity and stability of a purified protein. *In vivo* methods use a helper plasmid, mRNA or purified protein as a source of transposase. The helper plasmid is a plasmid expressing transposase and it is co-transfected or injected together with a plasmid carrying transposon DNA in a cell or an organism (Fraser et al, 1996). The active form of transposase synthesised inside the cells, recognises inverted repeats of the transposon on the transposon donor plasmid, excises it and inserts in the genomic DNA. The disadvantage of this method is constant expression of the transposase, which can result in excision of the integrated transposons or lead to over production inhibition, shown for transposons of the *Tc1/mariner* family (Lampe et al, 1998). Purified protein or mRNA of the transposase can be injected in a living cell to promote integration or excision of transposons *in vivo* (Coates et al, 2000; Kawakami, 2007; Mates et al, 2009). The limitation of using mRNA is its short lifetime. Also the injection procedure is an invasive technique, which results in increased cell death and low throughput.

It has been shown that *in vivo* activity of Mos1 transposase expressed from the helper plasmid is much lower than for piggyBac, Sleeping Beauty or Tol2 transposases – the most promising representatives for use in mammalian transgenesis (Wu et al, 2006b). This might be due to the fact that Mos1 transposase is expressed in high quantities *in vivo* and forms aggregates in cells, reducing the transposition activity (Demattei et al, 2011). To date Mos1 transposase has not been shown to be efficient for mammalian transposition.

3.2 Results

The *in vitro* transposition assay was developed for *P* transposable element (Kaufman & Rio, 1992) and later on optimised for Mos1 transposition (Zhang et al, 2001). It consists of the incubation of a transposon donor plasmid with a target plasmid and purified transposase for one hour, inactivation of the reaction with proteinase K for one hour, phenol-extraction and ethanol precipitation of the DNA and transfection of
the competent cells with the DNA mixture. The products of transposition are scored by counting the number of colonies having the resistance markers of both transposon and target plasmid. The donor plasmid has a conditional origin of replication and is unable to replicate in the recipient strain. As we wanted to use this assay to quantify in vitro transposition, we reduced the number of steps to increase accuracy and reproducibility. We substituted phenol extraction to heat inactivation at 80°C for 20 minutes. The ethanol precipitation step is required to concentrate the DNA and remove salt to allow electroporation. To eliminate the DNA precipitation step we used chemical transfection of bacteria, which is not so sensitive to the DNA volume and salt concentration.

After incubating the donor plasmid with the target plasmid and transposase (Figure 3-1 a), the reaction was heated to 80°C for 20 minutes and the whole volume of the reaction was transfected into chemically competent E.coli DH10B cells, which were then plated on kanamycin and ampicillin containing agar. Plasmid DNA of the colonies with two resistances was isolated and analysed by restriction digestion with XhoI endonuclease. XhoI cuts once within the transposon sequence and once in the target plasmid. The expected product of transposition is 4.8 kb and after the digest we expected to observe bands that in sum are 4.8 kb. An agarose gel of the plasmid DNA from 4 clones (Figure 3-1 b and c) shows that only one clone had a product of transposition plasmid, while three others carry only target plasmid DNA. If this plasmid DNA was transfected into E.coli DH10B cells again, the cells were resistant to ampicillin and sensitive to kanamycin. The presence of two plasmid bands on the undigested DNA sample (Figure 3-1 b, sample 2) can be explained as more than one plasmid gets inside the cells during transfection. High copy number plasmids can co-exist together in one cell even if they are from the same group of incompatibility and have identical origin of replication (Velappan et al, 2007).

An explanation for having two resistances while carrying a target plasmid without an insertion could be that the transposon was inserted into the genomic DNA of E.coli DH10B having transposase entered the cells during transfection. Since we did not observe any evidence of the donor plasmid present in the selected clones an alternative explanation could be a recombination event between the donor plasmid

In vivo transposition
Improving *mariner* transposons for transgenesis

**Figure 3-1 In vitro transposition assay**

(a) A scheme of the *in vitro* transposition assay (not to scale). A donor plasmid carries kanamycin resistance cassette (kan\(^R\)) flanked with the inverted repeats of Mos1. This plasmid is incubated with a target plasmid and purified transposase. The whole reaction is transfected into *E.coli* DH10B strain. Products of transposition are analysed on agarose gel. The digestion sites for XhoI endonuclease are indicated with the black lines.

(b) Separation of the products of transposition plasmids of 4 clones on an agarose gel without restriction digest. M – supercoiled DNA molecular marker, 1-4 – clones No. 1-4.

(c) Digestion analysis of the products of transposition. M – linear DNA molecular marker, 1-4 – clones No.1-4. The expected product of transposition is 4.8 kb – clone No. 2, other three clones contained a target plasmid only.

*In vivo* transposition
and *E.coli* genomic DNA. This would result in the absence of the target site duplication as a sign of transposition.

### 3.2.1 Verification of the integration in genomic DNA

We observed kan^R^ colonies even when the target plasmid was not added to the reaction while the donor plasmid does not replicate in the *E.coli* DH10B strain. To find if the integration occurred in the genome of *E.coli* DH10B we isolated chromosomal DNA of these kan^R^ clones, digested the DNA with EcoRI endonuclease, which does not cut within the transposon, and separated the products on an agarose gel. The separated DNA was transferred to nylon membrane and hybridised with fluorescently labelled oligonucleotides, complementary to the transposon inverted repeat DNA. Each band in Figure 3-2 represents an integration of the transposon in the genomic DNA. As a negative control we used chromosomal DNA of *E.coli* DH10B cells, which had not been transfected with the donor plasmid and transposase and were kanamycin sensitive. All 16 kan^R^ clones analysed contained transposon inverted repeat DNA. This means that the transposon was cut out from the donor plasmid and integrated in the bacterial genome. This can only happen if the transposase protein is co-transfected with the DNA during the transfection step.

### 3.2.2 Factors reducing the number of resistant clones

To check if the transposase protein can get inside the bacterial cells during the transfection step we tried four different methods of protein inactivation prior to transfection – heat inactivation, column DNA purification, incubation with proteinase K and phenol extraction. We did not add a target plasmid to the mixture of donor plasmid and transposase to be sure that the only clones we recover are those, which carry kan^R^ in the genomic DNA (Figure 3-3 a). Figure 3-3 (b) shows that the maximum number of colonies was observed when no inactivation was performed. Incubation with proteinase K and 80°C heat inactivation resulted in six times and three time reduced amount of resistant colonies respectively, while DNA column purification, 90°C heat inactivation and phenol extraction resulted in the absence of the resistant clones. These observations mean that the transposase or the complex of
Figure 3-2 Southern blotting of the digested genomic DNA of kanR colonies with labelled transposon inverted repeat DNA

The genomic DNA of the kanR clones obtained with Mos1 transposon has been digested with EcoRI endonuclease, separated on an agarose gel, transferred to a membrane and hybridised with labelled transposon inverted repeat DNA. EcoRI endonuclease does not cut within the transposon ends. Each band represents a transposon present in the genomic DNA. M – linear DNA molecular marker, 1-16 genomic DNA from 16 kanamycin resistant clones, negative control – genomic DNA of E.coli DH10B strain.
Improving *mariner* transposons for transgenesis

**Figure 3-3** Minimised transposition assay (a) and factors reducing the number of kan\(^R\) colonies (b)

(a) Minimised *in vitro* transposition assay, where no target DNA is added. Donor plasmid has a conditional oriR6K origin of replication and can not replicate in a recipient *E. coli* strain.

(b) Graph showing comparison of four methods of protein inactivation following *in vitro* transposition. Column purification, incubation with proteinase K, heat inactivation at two different temperatures for 20 minutes and phenol extraction were tested. Column purification, 90°C heat inactivation and phenol extraction result in no kan\(^R\) colonies observed.
excised transposon and transposase is stable to heat inactivation. The reduced number of resistant colonies after heat inactivation might as well be due to the damaging effect of the heat to the donor DNA. The effect of the heating DNA prior to transfection on the efficiency of transfection can be tested to check this assumption. The absence of the kan^R colonies after phenol extraction or column purification indicated that transposase is required for resistant clones to grow after transfection. The integrations can only happen if transposase enters the bacterial cells during transfection. We named this event \textit{in vivo} transposition.

We wanted to find out which transfection method allows transposase to pass through the bacterial cell wall. We compared three transfection methods – electroporation of $\frac{1}{10}$th of the reaction, calcium transfection and rubidium transfection of all the reaction volume. Figure 3-4 shows that the most efficient transfection method is rubidium transfection and the least efficient is electroporation. It is easier to understand why rubidium transfection is more preferable for the highest number of kan^R colonies taking into consideration that rubidium transfection is more efficient method than calcium transfection and the buffer for preparing competent cells contains 50 mM MnCl\(_2\). Manganese ions are required for transposase activity \textit{in vitro} and transposon could be excised from the donor plasmid during incubation and prior to transfection, which results in higher colony number. The lowest amount of colonies was observed during electroporation, this can be due to the protein inactivation while applying a strong electric field. \textit{In vivo} transposition in bacteria is shown for both Mos1 and Mboumar-9 transposases (comparison of the efficiencies is shown on p. 102).

\section*{3.2.3 Transposition in other organisms}

\subsection*{3.2.3.1 Yeast}

We wondered if transposase/transposon complex could also pass through the cell walls of other organisms during transfection. The width of the gram negative bacterial cell wall is 5-10 nm (Beveridge, 1999), while yeast cell wall is 90-154 nm and contains the layer of chitin in its outer membrane (Dupres et al, 2010). The efficiency of bacterial chemical transfection is up to $10^8$-$10^9$ CFU/µg of DNA and
Figure 3-4 Comparison of different transfection methods for in vivo transposition
Graph represents the number of kan^R resistant colonies after modified transposition reaction followed by one of three transfection techniques. The higher number of colonies means higher number of cells had a transposon inserted in their genomic DNA. Rubidium transfection is the best method for in vivo transposition.
fission yeast *S. pombe* can be transfected by lithium transfection with the efficiency of $10^4$-$10^5$ CFU/µg of DNA only. We were curious if the *in vivo* transposition method discovered for *E. coli* is applicable for yeast cells. We used the assay shown in Figure 3-3 (a) to transfect yeast cells by lithium transfection. The amount of resistant clones screened on G418 (analogue of kanamycin for eukaryotic cells) was three times higher when the transposase was added to the reaction (Figure 3-5).

There were a few colonies appearing on the negative control plates, this could be due to the high recombination rate in *S. pombe* cells (Chua et al, 2000). Another limiting factor to obtain the maximum amount of resistant colonies is that the kanamycin cassette had a bacterial promoter and only if the transposon was integrated in the open reading frame of another gene expressed in *S. pombe* these clones can show resistance to the drug. The results for yeast integrations were obtained for Mos1 transposase. It is worth trying *in vivo* transposition of the transposon with the eukaryotic neo$^R$ promoter in yeast cells for both Mos1 and Mboumar-9 transposases. It is known that SV40 promoter is very well expressed in *S. pombe* cells (Jones et al, 1988), this promoter can be added to the transposon to be able to express the neo$^R$ cassette in yeast cells.

### 3.2.3.2 Mammalian cells

Expanding the application of the *in vivo* transposition into the mammalian cells could be potentially useful as a tool for generation of stable cell lines in the case of cells which can not express transposase from a helper plasmid or when the constant presence of transposase in the cells is not desirable. The efficiency of DNA transfection into mammalian cells is $10^4$-$10^5$ CFU/µg of DNA and highly depends on cell confluence prior to transfection.

#### 3.2.3.2.1 Transposons for mammalian cells

We have created new transposon donor plasmid that carries neomycin resistance cassette with a eukaryotic promoter. The neomycin gene (neo$^R$) under the regulation of SV40 promoter was flanked with the right inverted repeats of Mos1 or with the left inverted repeats of Mboumar-9 (reasons for using these inverted repeats explained later in Chapter 5).
Improving mariner transposons for transgenesis

Figure 3-5 In vivo transposition in different types of cells

(a) The amount of kanR or neoR colonies after addition of purified Mos1 transposase during transfection of E.coli, S.pombe and HeLa cells (1/12 of the reaction).
(b) Quantification of the number of resistant colonies per 1 µg of donor DNA added for the whole reaction.
Improving mariner transposons for transgenesis

To create neo<sup>R</sup> containing transposons we deleted the SalI site from the pBSKS(+) vector by site directed mutagenesis using primers No. 25 and 26. This gave plasmid pBSKS(+)noSalI. The kan<sup>R</sup> gene with the inverted repeats sequences from pEPMosRR (digested with XbaI) and from pEPMboLL (digested with SacI) was subcloned into pBSKS(+)noSalI using the appropriate sites. The kanamycin cassette was cut out from both of the transposons with SalI restriction endonuclease. The neomycin cassette of 1.6 kb was amplified from the pEGFP-C1 vector with the primers No. 17 and 18. Resulting product was cloned by SalI sites into the prepared vectors with two right inverted repeats of Mos1 or two left inverted repeats of Mboumar-9 transposons. The composition of the resulting constructs has been confirmed by sequencing.

We transfected HeLa cells with the mixture of donor plasmid and transposase as described on p. 146. At least 14 times difference in the number of stable lines was observed when the transposase was added (Figure 3-5). This means that transposase can be successfully co-transfected into mammalian cells and integrate sequence that lies between the inverted repeats into chromosomes, enabling straightforward creation of stable cell lines. The results for mammalian cells were obtained using Mos1 as well as Mboumar-9 transposons, which showed the same level of activity.

3.2.4 Mapping the integrations sites

To directly prove that integrations are happening in the genomic DNA we needed to sequence DNA flanking the transposon at its new location. We have done this in two ways – cloning of the genomic DNA into a plasmid vector and by inverted PCR on the self ligated genomic DNA.

Genomic DNA of the kan<sup>R</sup> bacterial cells was digested with EcoRI endonuclease, which does not have restriction sites in either the inverted repeats or the kan<sup>R</sup> cassette. The obtained fragments of the genomic DNA were cloned into the pBSKS(+) vector by EcoRI site. Products of ligation were selected by resistance to kanamycin (within the transposon) and ampicillin (within the vector). The plasmid DNA of these products was isolated and analysed by agarose gel separation and restriction digest with EcoRI (Figure 3-6). Digestion with EcoRI should release the
Improving *mariner* transposons for transgenesis

**Figure 3-6 Bacterial plasmid DNA with cloned genomic DNA fragments**

(a) separation on agarose gel of plasmid DNA either undigested (a) or digested with EcoRI (b). This plasmid DNA was obtained by ligation of the genomic DNA of kan^R^ clones digested with EcoRI and pBSKS(+) cloning vector. EcoRI endonuclease does not cut within the transposon. All 7 clones contain the DNA of plasmid vector and various size of inserts of the genomic DNA. These plasmids were sequenced to confirm the presence of transposon, genomic DNA and a target site duplication.

**Figure 3-7 Inverted PCR on genomic DNA with transposons insertions**

Separation on agarose gel of the products of inverted PCR on the self ligated genomic DNA of kan^R^ clones, digested with EcoRI endonuclease. PCR primers bind to the ends of the transposon and amplify genomic DNA flanking the integration site. The fragments were gel purified and sequenced to confirm the presence of transposon, genomic DNA and a target site duplication. M – linear DNA molecular marker, 1-9 – DNA from clones 1-9 was used for amplification.
Improving *mariner* transposons for transgenesis

fragment of genomic DNA from the plasmid backbone. All obtained plasmids contained the cloning vector and inserts of various sizes. We have sequenced these plasmids to find out the flanking transposon DNA.

We also performed inverted PCR on the self ligated genomic DNA using primers No. 11 and 12 to amplify the DNA joined to the ends of the inverted repeats in order to map the insertions. The resulted products of amplification (Figure 3-7) were gel purified and sequenced.

Sequencing results from both methods confirmed that transposon was integrated in the genomic DNA of *E.coli* DH10B strain with the duplication of TA nucleotides, which is a representative sign of the transposons of the *Tc1/mariner* family. Finding the flanking DNA enables us to map the integration sites on the genomic DNA (Figure 3-8 a).

There was only one insertion successfully sequenced for HeLa cells *in vivo* transposition, which occurred in the chromosome 1 (Figure 3-8 b). So far we did not get sequencing data with the yeast integrations.

### 3.3 Discussion

The optimisation of the *in vitro* transposition method lead to the discovery of the transposon integrations in the genomic DNA, which we called *in vivo* transposition. We assume that a very stable transposase/transposon complex is formed in a tube. DNA while going through the membrane during transposition pulls the protein inside the cell. It has been shown that in solution transposase has a shape of an elongated stick, which wraps around the DNA molecule (Cuypers et al, 2012). Such conformation of the complex minimises spatial obstructions and favours the transfection process. In case of mammalian cells transfection, lipid spheres are formed mediated by a transfection reagent, which fuse with the cell membrane. This process is less selective for DNA, that is why proteins, which are bound to DNA can be taken up by a cell.

*Tn5* has been extensively used for sequencing of genomic DNA. An *in vivo* method consists of a transposon, carrying a resistance cassette, and a conditional origin
Improving mariner transposons for transgenesis

In vivo transposition

Figure 3-8 Mapped insertions in *E. coli* and *H. sapiens* genome

(a) The whole *E. coli* genome and an example of 11 insertions of the transposons. Each of the coloured rectangles represent an insertion in one of the analysed kanR clones. The whole genome of *E. coli* is shown as a long black line, green dots underneath the line represent single genes. Coloured vertical lines indicate places of transposons integration. The colourful rectangles at the top of the line highlight the places of integration (built using NCBI web resource).

(b) The whole chromosome 1 of *H. sapiens* with one insertion mapped in the vacuolar protein sorting 13 homolog D – highlighted with a light grey rectangle (built using NCBI web resource).
of replication. After random integration of the transposon in the genomic DNA, it is isolated, digested, self-ligated and fragments containing a transposon insertion are selected in a bacterial strain, which enables functioning of the conditional origin of replication. This is a powerful and a widely used method for sequencing applications. Mos1 and Mboumar-9 transposase based systems can also be used for yeast and mammalian cells. More experiments are required to prove the integrations in yeast genome by sequencing as well as to prove the universal application of the system by finding more bacterial strains and mammalian cell lines in which transposons can be integrated. It is worth checking if the system is applicable for gram-positive bacteria as well. There are a number of applications, in which in vivo transposition could be useful – creation of stable cell lines, introducing genes in cells which do not propagate extra-chromosomal DNA, creation of the libraries of random gene knock-outs, sequencing.
Chapter 4: The inverted repeats distinction and target site selection
4.1 Introduction

Transposable elements are flanked with terminal inverted repeats that are recognised by transposase protein. Some of them are identical like in the P element from Drosophila (O'Hare & Rubin, 1983), Tc2 and CemaT1 from nematodes (Claudianos et al, 2002; Ruvolo et al, 1992), bacterial Tn4430 (Nicolas et al, 2010), or Pegasus from mosquitoes (Besansky et al, 1996). Other transposable elements are flanked with non-identical inverted repeats. These include piggyBac from baculoviruses (Cary et al, 1989), Tn5 from bacteria (Berg et al, 1982), ISY100 from the cyanobacteria (Urasaki et al, 2002), SleepingBeauty from fish (Cui et al, 2002), Mos1 from Drosophila (Medhora et al, 1991), Mboumar-9 from ants (Munoz-Lopez et al, 2008). The role of the inverted repeats difference is not yet fully understood. Several studies have shown, that the transposons with two identical inverted repeats are more active, than those found in nature (Jaillet et al, 2012; Zhou et al, 1998). The explanations that have been suggested to date are that the transposons were selected in a way to keep the transposition rates at a suboptimal level (Zhou et al, 1998) or that transposase has to bind to one of the repeats first while the second repeat stays unoccupied for the proper paired end complex formation (Jaillet et al, 2012).

The Mos1 transposon is flanked with 28 bp imperfect inverted repeats (Figure 4-1 a), which differ at the positions 3, 11, 13 and 28. The DNA-binding domain of Mos1 transposase has a preference for binding the right inverted repeat sequence (Zhang et al, 2001) and cleavage of this repeat is more efficient in vitro (Auge-Gouillou et al, 2001b). The first cleavage of Mos1 inverted repeats occurs at the 25th base of non-transferred strand and the second cleavage occurs at the 28th base of the transferred strand. Other studies have shown that full length Mos1 transposase fused at the N-terminal with maltose binding protein binds to the right inverted repeat 10 times more efficiently, than to the left (Auge-Gouillou et al, 2001b) and that a transposon with two right inverted repeats is 20 times more active in the whole in vitro transposition assay (Auge-Gouillou et al, 2001b; Jaillet et al, 2012).
Figure 4-1 Alignment of Mos1 and M bourgeois-9 transferred strand (TS) sequences of the left and right inverted repeats

(a) The Mos1 left and right inverted repeats are 28 bp long and differ in four bases: positions 3, 11, 13 and 28.

(b) M bourgeois-9 inverted repeats are 32 bp and differ in two bases: positions 13 and 27.
Mboumar-9, a naturally active transposon from ants, has 64% nucleotide sequence identity with Mos1 transposase (Munoz-Lopez et al, 2008). The inverted repeats of Mboumar-9 transposon are 32 bp long and differ at only two positions – 13 and 27 (Figure 4-1 b). It is not known whether Mboumar-9 transposase has any preference towards one of its inverted repeats and if it does, how this correlates with the difference of Mos1 transposase activity on the left and right inverted repeats.

4.2 Results
4.2.1 Cleavage efficiencies of the right and left inverted repeats
Here we investigate the difference in the cleavage efficiencies of the Mos1 and Mboumar-9 transposons. In this study we used constructed transposons containing kanamycin resistance cassette flanked with either 28 bp inverted repeats of Mos1 or 32 bp inverted repeats of Mboumar-9 transposon. The transposons of naturally found imperfect form (with left and right inverted repeats) are MosLR and MboLR. We also tested the activity of the transposons with two left inverted repeats – MosLL and MboLL, and two right inverted repeats – MosRR and MboRR. The scheme of the assay is presented in the Figure 4-2 (a). The transposon donor – supercoiled plasmid DNA carrying one of the constructed transposons – was incubated with the corresponding transposase for one hour (as described on p. 158). The products of the reaction were separated on an agarose gel. We observed either linearization of the plasmids (band of 5.6 kb), or full transposon excision (bands of 4.3 kb the plasmid backbone and 1.3 kb transposon) (Figure 4-2 b). We quantified the efficiency of cleavage by measuring the intensity of the band, corresponding to the released plasmid backbone after the transposon excision. The intensity of the band corresponding to free transposon was less than expected, as an excised transposon is a transient point, followed by an insertion of the transposon in a target DNA (any DNA present in the reaction). Quantification of cleavage efficiency showed that MosRR transposon is cleaved out 9 times more efficiently than MosLR, the form found in nature. Mboumar-9 transposase does not have such a significant preference towards the inverted repeats cleavage – less than two times (Figure 4-2 c). Mos1
Improving *mariner* transposons for transgenesis

**Figure 4-2 In vitro cleavage of the Mos1 and Mboumar-9 transposons**

(a) A scheme of the *in vitro* cleavage assay, (b) an agarose gel of the products of *in vitro* cleavage, (c) quantification of the backbone released. M – molecular weight marker, MosLL, MosLR, MosRR, MboLL, MboLR and MboRR – plasmids, containing transposons which are flanked either with two left terminal inverted repeats, or left and right or both right terminal inverted repeats.
transposase has preference for the right inverted repeat indeed while Mboumar-9 more actively cleaves its left inverted repeat.

4.2.2 Differences in the \textit{in vitro} transposition efficiency

To further investigate transposase preference towards one of the inverted repeats we performed the whole transposition reaction \textit{in vitro} as described on p. 158. The DNA donor of a transposon, carrying kanamycin resistance was incubated with the corresponding transposase and a target plasmid, carrying ampicillin resistance. After phenol extraction the DNA mixture was transformed into \textit{E.coli} DH10B and the cells carrying products of transposition were selected using both antibiotics. The donor plasmid DNA does not replicate in the recipient \textit{E.coli} strain, as it contains conditional oriR6K origin of replication, which requires P protein (product of \textit{pir} gene). The number of Amp\textsuperscript{R}Kan\textsuperscript{R} colonies gives a measure of the transposition efficiency. We calculate \textit{in vitro} transposition efficiency as the number of resistant colonies per 1 µg of donor DNA. In the Figure 4-3 (a) data is normalised to 1500 colonies per 1 µg of donor DNA (transposition efficiency of the MboRR transposon).

The results of \textit{in vitro} transposition assays indicate that each of the transposases has preference towards one of the inverted repeats for the whole transposition reaction, as well as for cleavage. Figure 4-3 (a) shows that the Mos1 transposon with two right inverted repeats (MosRR) is 25 times more active than the MosLR transposon, with a native inverted repeats composition. On the other hand the MosLR transposon is 50 times more active than MosLL, flanked with two left inverted repeats. The difference for Mboumar-9 transposition efficiency is not so big. The MboLL transposon is two times more active than the form of the native LR transposon, while the native MboLR form is three times more active than the MboRR transposon. All P-values calculated using T test are less than 0.0001.

4.2.3 Differences in the \textit{in vivo} transposition efficiency

Using the \textit{in vivo} transposition assay developed in this study (p. 159) we compared the activities of the Mos1 and Mboumar-9 transposons flanked with different combinations of the inverted repeats as used for the \textit{in vitro} cleavage and
Improving *mariner* transposons for transgenesis

**Figure 4-3 Transposition efficiencies of Mos1 and Mboumar-9 with the left and right inverted repeats in vitro and in vivo**

(a) Transposition efficiencies of the Mos1 and Mboumar-9 transposons *in vitro* normalised to 1500 colonies per 1 µg of donor DNA (P-values < 0.0001).

(b) Transposition efficiencies of the Mos1 and Mboumar-9 transposons *in vivo* normalised to 37 colonies per 1 µg of donor DNA (for Mos1 P-values < 0.01, for Mboumar-9 P-values > 0.05). Both *in vitro* and *in vivo* Mos1 and Mboumar-9 have preference for one of the inverted repeats. Mboumar-9 transposase shows weaker preference (two to three fold) compared to Mos1, with activity between transposons differing from 10 to 30 times.
Improving *mariner* transposons for transgenesis

Transposition assays described above. The transposition efficiency *in vivo* is normalised to 37 colonies per 1 µg of donor DNA (transposition efficiency of the MboRR transposon). Figure 4-3 (b) represents transposition efficiencies of Mos1 and Mboumar-9 transposons flanked with different inverted repeats *in vivo*. We noticed that in this assay the preference for one of the inverted repeats stayed the same as in the *in vitro* reactions, but the standard deviation error bars are much higher compared to the reactions carried out *in vitro*. This difference can be explained by the addition of a step of passing through the bacterial cell wall during *in vivo* transposition assay, which is unnatural for the transposases. This assay appeared to be more sensitive and variable while performing biological replicas.

During *in vivo* transposition assay the transposon DNA and the transposase protein were added to the vials of competent cell simultaneously and incubated for 30 minutes on ice before the 42°C heat shock for 60 sec. The heat shock could result in protein inactivation according to thermal denaturation assay which showed that Mboumar-9 transposase denaturates above 39.7°C (Figure 2-15). During thermal denaturation protein was exposed to a different type of incubation – the temperature of the plate with protein samples was raised from 4°C to 95°C by 1°C step with 30 sec hold. Thus, the protein was exposed to longer incubation at the temperatures above 4°C. Denaturation of the protein is not a momentary process (Graczer et al, 2007) that is why during heat shock a portion of Mboumar-9 transposase can remain active. Another possible explanation can be that partial aggregation of the Mboumar-9 transposase preserves the protein in an active state during the heat shock and these molecules are effective inside the bacterial cells and integrate transposons in the genomic DNA of *E.coli*. Both these hypotheses require further investigation.

### 4.2.4 Activity of the G32A Mboumar-9 transposon mutant

In the previous experiments we found that the cleavage at the right inverted repeat of the Mos1 transposon is 10 times more active than the left and that a transposon flanked with two right inverted repeats is 25 times more active than a transposon with left and right inverted repeats. It has been shown before that the most critical base for Mos1 transposase binding is 13th (Auge-Gouillou et al, 2001b), while in the
crystal structure of the paired end complex this base is not involved in specific interactions with transposase (Richardson et al, 2009). Single base substitutions of bases 3, 11 and 28 did not have a significant effect on Mos1 transposase binding in vitro (Auge-Gouillon et al, 2000). The 28th base of the transferred strand of the Mos1 inverted repeats is adenine in the right repeat and guanine in the left (Figure 4-1). The crystal structure of Mos1 paired end complex shows that the base at the 3’ end of the transferred strand is not involved in the sequence-specific recognition by the Mos1 transposase, but has a purine-specific contact with Arg183 (Richardson et al, 2009). DNasel footprinting has shown that 28th base is protected by Mos1 transposase (Auge-Gouillou et al, 2001a). Since the 3’ end is a reactive end of the inverted repeat and is a nucleophile for target integration, this base is highly important for transposase activity. Both Mboumar-9 inverted repeats have guanine at the 3’ end of the transferred strand (Figure 4-1 b). We wanted to know, whether changing the last base of the Mboumar-9 inverted repeat sequence to adenine affects cleavage and transposition activities.

4.2.4.1 Construction
The guanine nucleotide at the 3’ end of the left inverted repeat of Mboumar-9 was changed to adenosine in both inverted repeats of the MboLL transposon. Site directed mutagenesis was performed with one pair of primers No. 5 and 6 (list of primers p. 140) on the pEPMboLL plasmid, which resulted in two mutations, one in each of the identical inverted repeats of pEPMboLL. The resulted pEPMboLL G32A was sequenced to confirm the presence of the mutations of interest.

4.2.4.2 In vitro cleavage
We checked if there is preference for cleavage of the Mboumar-9 inverted repeat carrying adenine as the last base, compared to the wild type inverted repeat. The results of the in vitro cleavage assay (Figure 4-4) showed that the transposon with the G32A mutation is cleaved with the same efficiency as the native left inverted repeat. Indicating that introduction of adenine at the last position of the transferred strand has no effect on cleavage by transposase in vitro.
Improving *mariner* transposons for transgenesis

Figure 4-4 *In vitro* cleavage of the MboLL and the G32A mutant

(a) An agarose gel of the products of *in vitro* cleavage by Mboumar-9 transposase (Mbo) of the transposon with two left inverted repeats (LL) or carrying a G32A mutation in the left inverted repeat (LL G32A).

(b) Quantification of the cleavage products on lanes LL+Mbo and LL G32A+Mbo. All DNA – a sum of the intensities of all bands across the lane taken as 100%, % of DNA cleaved – intensity of the band corresponding to the released backbone relative to all DNA intensity. Quantification shows that there is no difference in the cleavage efficiencies of the native left inverted repeat and G32A mutant.
4.2.4.3 *In vivo* transposition

To investigate the effect of the G32A mutation on the whole transposition reaction we carried out the *in vivo* transposition assay, comparing MboLL transposon activity to the MboLL G32A mutant activity. Strikingly, the efficiency of the MboLL G32A *in vivo* transposition was enhanced three and a half times ($p < 0.0001$) compared to the transposon flanked with two native left inverted repeats (Figure 4-5). This finding suggests that the base at the 3’ end of the inverted has a significant effect at the stage of integration into a new target site. Presumably, an adenine base at the 3’ end position of the transferred stand makes a target site integration step more efficient and, as a result, amplifies the transposition efficiency without influencing the cleavage step.

4.2.5 Modeling of the Mbourmar-9 transposase PEC structure

Attempts to crystallise Mboumar-9 transposase in the presence or absence of DNA did not yield protein crystals. We used protein-DNA (1:1.2 ratio) complex in concentrations from 10 µM to 120 µM. The DNA used for crystallisation trials was dsDNA of the left inverted repeat with the three nucleotide stagger at the 3’ end, mimicking the cleaved transposon end. DNA oligonucleotides of 29 nucleotides of non-transferred strand and 32 nucleotides of transferred strand (Figure 2-14) of the left inverted repeat of Mboumar-9 transposon were annealed and purified by HPLC (as described on p. 147). Commercially available screens tested on the Mboumar-9-IRL DNA complex were: Morpheus, Natrix 1 and 2, SaltRx, Silver Bullets additive screen from Hampton Research; MIDAS, Clear Strategy Screen 1 and 2, JCSG-plus from Molecular Dimensions. More than 1000 different crystallisation conditions were tested including manually set grid screens.

The amino acid sequences of Mos1 and Mboumar-9 transposases are 68 % identical. The inverted repeats of Mos1 and Mboumar-9 have about 50 % sequence similarity. We used modelling approach to predict Mboumar-9 paired end complex (PEC) structure based on that of Mos1 transposase (PDB-ID 3HOT) (Richardson et al, 2009). Modelling was done on the Swiss-Model Workspace (Arnold et al, 2006; Guex & Peitsch, 1997; Schwede et al, 2003). We aligned the resulted protein
Improving *mariner* transposons for transgenesis

**Figure 4-5 In vivo transposition efficiency of MboLL and MboLL G32A mutant**

*In vivo* transposition efficiencies of the MboLL transposon (LL) and the G32A mutant of the transposon (LL G32A) normalised to 50 colonies per 1 µg of donor DNA. A transposon with the G32A mutations of the two left inverted repeats is three and a half times more active in the whole transposition reaction than the MboLL transposon carrying two native left inverted repeats.
model with the paired end complex structure of Mos1 transposase and mutated each of the inverted repeat residues to those of Mboumar-9 transposon. First four adenine residues of Mboumar-9 transposon inverted repeat are not added to the structure. Figure 4-6 (a and b) represents the alignment of the resulted structure of Mboumar-9 transposase and the Mos1 paired end complex. The predicted structure of the Mboumar-9 paired end complex looks almost identical to the Mos1 crystal structure. This might be due to the high level of identity of these proteins, or due to the algorithm of the modelling, where the predicted protein model is fit to match the template structure. Figure 4-6 (b and c) shows the arrangement of the Mboumar-9 transposase and the left inverted repeat DNA in a modelled paired end complex structure.

To access the quality of the generated model we used Anolea (Atomic Non-Local Environmental Assessment) server (Melo et al, 1997). Anolea represents the non-local interactions of each of the amino acids in the protein structure, using the protein chain only. The result represents the energy profile, where low energy corresponds to the most preferable amino acid conformation and high energy shows the errors or potential interacting zone of the protein. The Anolea profile of the monomer of Mos1 is calculated based on the crystal structure of the paired end complex (Figure 4-7 a). The total non-local energy of Mos1 transposase is -1468 E/kT. The Anolea profile of Mboumar-9 transposase is built based on the predicted structure of the paired end complex (Figure 4-7 b). The resulted total non-local energy for Mboumar-9 is -1183 E/kT, which is higher, than for Mos1 transposase. This indicates that the amino acids of Mboumar-9 transposase in the generated structure are in less favourable orientation, compared to the structure of Mos1 transposase. In the graph, representing the Anolea profile of Mos1 transposase the highest energy of -67 E/kT is localised in the first helix-turn-helix domain. This can be due to the protein-protein interactions between two monomers of Mos1 transposase, as the HTH1 is shown to be a dimerization domain of the protein (Section 2.2.8). The highest energy of the Mboumar-9 transposase of -14 E/kT is localised in the HTH1 as well, and is greater than those for Mos1 HTH1 transposase. This can be explained either as the structure of Mboumar-9 HTH1 transposase is more diverse from the Mos1 transposase HTH1
Improving mariner transposons for transgenesis

Figure 4-6 Alignment of the Mos1 and Mboumar-9 paired end complex structures
(a and b) two orthogonal views of the aligned paired end (DNA not shown) crystal structure of Mos1 transposase in violet (PDB-ID 3HOT) and predicted by modelling structure of Mboumar-9 transposase (in green). The N-terminal DNA binding motif (residues 1-112) contains two helix-turn-helix domains (HTH1 and HTH2), joined by a linker (residues 113-125) to the catalytic domain (residues 126-161 and 190-345) of a transposase.
(c and d) Two orthogonal views of the predicted Mboumar-9 paired end complex structure with the left inverted repeat DNA (IRL DNA).
Improving mariner transposons for transgenesis

Figure 4-7 Anolea of Mos1 and Mboumar-9 transposases
Anolea represents the energy of each of the protein residues in the protein 3D structure. Anolea of Mos1 transposase (a) is calculated for the crystal structure of the paired end complex. The Anolea of Mboumar-9 transposase (b) is based on the predicted by modelling structure, with Mos1 paired end complex structure as a template. The area of high energy corresponds to the unfavourable amino acid conformations or potential protein interactions. The HTH1 of Mboumar-9 has higher energy compared to the same motif of Mos1 transposase.
than it is for other domains. That is why the modelled structure may not represent the actual amino acid conformations for the Mboumar-9 HTH1 transposase. Another explanation is that the contact of the two HTH1 domains is tighter between the two monomers of the Mboumar-9 transposase, than it is in the paired end complex of Mos1 transposase.

4.2.5.1 Base-specific recognition of inverted repeats

Residues that are involved in the base-specific recognition of the inverted repeat DNA are present in all domains throughout the length of the transposase protein. The Mos1 transposase amino acid residues that are involved in sequence-specific and purine/pyrimidine-specific DNA recognition are presented in (Table 4-1) Based on the generated model of the Mboumar-9 PEC, we predicted the residues of Mboumar-9 transposase which may be involved in the base-specific recognition of Mboumar-9 inverted repeats (Table 4-2). Figure 4-8 shows the alignment of the Mos1 and Mboumar-9 transposase amino acid sequences with the amino acids involved in base-specific interactions are indicated with a black arrow and purine/pyrimidine-specific interactions with a grey arrow. Alignment of the inverted repeats sequences with the residues involved in specific DNA recognition is presented in Figure 4-9. Lys44 of Mos1 transposase and Arg44 of Mboumar-9 transposase interact with the guanine of the transferred strand of their inverted repeats in the same manner (Figure 4-10 a and b). Interaction of Arg48 of Mos1 with G7 of the non-transferred strand does not have an equivalent with the same in Mboumar-9 transposase (more details in section 4.2.5.2). Residue 100 of the HTH2 is important for specific interaction with A20 in Mos1 and G24 in Mboumar-9 of the non-transferred DNA strands (Figure 4-10 c and d). In the linker region of both transposases Arg118 interacts specifically with the thymine and guanine of the pre-cleaved transferred strand DNA, which are single stranded in the pair end complex (Figure 4-11 a and b). The DNA base at the 3’ end of the transferred strand is single stranded and makes a pyrimidine-specific interaction with Arg183 of the clamp loop of Mos1 transposase (Figure 4-11 c). In the case of Mboumar-9 transposase we see more specific interaction with the guanine at the 3’ transposon end by Arg183, where two hydrogen bonds can be made (Figure 4-11 d). The effect of the last base difference
Table 4-1 Base-specific interactions of Mos1 transposase with the inverted repeat DNA

Amino acid residues of Mos1 transposase, which specifically interact with the DNA bases of the inverted repeat in the crystal structure of the paired end complex. Base-specific interactions in black, pyrimidine-specific interactions in grey italics.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Amino acid residue</th>
<th>Atom</th>
<th>Nucleic acid residue</th>
<th>Atom</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTH1</td>
<td>Lys44</td>
<td>N\textsubscript{Z}</td>
<td>O\textsubscript{6}</td>
<td>G5 ts</td>
</tr>
<tr>
<td>HTH1</td>
<td>Lys44</td>
<td>N\textsubscript{Z}</td>
<td>N\textsubscript{7}</td>
<td>G5 ts</td>
</tr>
<tr>
<td>HTH1</td>
<td>Arg48</td>
<td>NH\textsubscript{2}</td>
<td>O\textsubscript{6}</td>
<td>G7 nts</td>
</tr>
<tr>
<td>HTH1</td>
<td>Arg48</td>
<td>NH\textsubscript{1}</td>
<td>N\textsubscript{7}</td>
<td>G7 nts</td>
</tr>
<tr>
<td>HTH2</td>
<td>Gln100</td>
<td>OE\textsubscript{1}</td>
<td>N\textsubscript{7}</td>
<td>A20 nts</td>
</tr>
<tr>
<td>HTH2</td>
<td>Gln100</td>
<td>NE\textsubscript{2}</td>
<td>N\textsubscript{6}</td>
<td>A20 nts</td>
</tr>
<tr>
<td>HTH2</td>
<td>Gln101</td>
<td>NE\textsubscript{1}</td>
<td>O\textsubscript{6}</td>
<td>G19 ts</td>
</tr>
<tr>
<td>Linker</td>
<td>Arg118</td>
<td>NH\textsubscript{1}</td>
<td>O\textsubscript{4}</td>
<td>T26 ts</td>
</tr>
<tr>
<td>Linker</td>
<td>Arg118</td>
<td>NH\textsubscript{2}</td>
<td>O\textsubscript{6}</td>
<td>G27 ts</td>
</tr>
<tr>
<td>Clamp loop</td>
<td>Arg183</td>
<td>NH\textsubscript{1}</td>
<td>N\textsubscript{7}</td>
<td>A28 ts</td>
</tr>
<tr>
<td>Clamp loop</td>
<td>Pro184</td>
<td>O</td>
<td>N\textsubscript{2}</td>
<td>G27 ts</td>
</tr>
<tr>
<td>Clamp loop</td>
<td>Pro184</td>
<td>O</td>
<td>N\textsubscript{1}</td>
<td>G27 ts</td>
</tr>
<tr>
<td>Catalytic</td>
<td>His293</td>
<td>NE\textsubscript{2}</td>
<td>N\textsubscript{2}</td>
<td>G25 nts</td>
</tr>
</tbody>
</table>

Table 4-2 Base-specific interactions of Mboumar-9 transposase with the inverted repeat DNA

Amino acid residues of Mboumar-9 transposase predicted to specifically interact with the DNA bases of the inverted repeat in the modelled paired end complex.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Amino acid residue</th>
<th>Atom</th>
<th>Nucleic acid residue</th>
<th>Atom</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTH1</td>
<td>Arg44</td>
<td>NH\textsubscript{1}</td>
<td>O\textsubscript{6}</td>
<td>G9 ts</td>
</tr>
<tr>
<td>HTH1</td>
<td>Arg44</td>
<td>NH\textsubscript{2}</td>
<td>N\textsubscript{7}</td>
<td>G9 ts</td>
</tr>
<tr>
<td>HTH2</td>
<td>Arg100</td>
<td>NH\textsubscript{1}</td>
<td>N\textsubscript{7}</td>
<td>G24 nts</td>
</tr>
<tr>
<td>HTH2</td>
<td>Arg100</td>
<td>NH\textsubscript{2}</td>
<td>O\textsubscript{6}</td>
<td>G24 nts</td>
</tr>
<tr>
<td>Linker</td>
<td>Arg118</td>
<td>NH\textsubscript{1}</td>
<td>O\textsubscript{4}</td>
<td>T30 ts</td>
</tr>
<tr>
<td>Linker</td>
<td>Arg118</td>
<td>NH\textsubscript{2}</td>
<td>O\textsubscript{6}</td>
<td>G31 ts</td>
</tr>
<tr>
<td>Clamp loop</td>
<td>Arg183</td>
<td>NH\textsubscript{1}</td>
<td>N\textsubscript{7}</td>
<td>G32 ts</td>
</tr>
<tr>
<td>Clamp loop</td>
<td>Arg183</td>
<td>NH\textsubscript{2}</td>
<td>O\textsubscript{6}</td>
<td>G32 ts</td>
</tr>
</tbody>
</table>
Improving mariner transposons for transgenesis

**Figure 4-8 Alignment of Mos1 and Mbourmar-9 transposase amino acid sequences**

Identical residues are indicated with a blue background. Black arrows indicate residues involved in base-specific recognition of the inverted repeats and grey arrows indicate pyrimidine-specific recognition. Thr48 is shown by a green arrow and was mutated to Arg to correspond to the T11G mutation of the inverted repeat of Mbourmar-9 transposon.

The inverted repeats distinction and target site selection 114
Improving *mariner* transposons for transgenesis

Figure 4-9 Base pairing of the transferred and non-transferred strands of Mos1 IRR and Mboumar-9 IRL.

Base pairing of the transferred and non-transferred DNA strands of Mos1 IRR (a) and Mboumar-9 IRL (b). Black arrows indicate bases involved in sequence-specific interaction with transposase, grey arrows – pyrimidine-specific interactions, green arrow – T11, which was mutated to G to correspond to the Thr48Arg change in Mboumar-9 transposase.
Figure 4-10 Residues 44 and 100 involved in sequence-specific DNA recognition by transposase

Base-specific interactions of Lys44 of Mos1 (a) and Arg44 of Mboumar-9 (b) with the guanine of the inverted repeat DNA. Base-specific recognition of Gln100 (c) of Mos1 and Arg100 of Mboumar-9 (d) with adenine and guanine respectively.
Figure 4-11 Residues 118 and 183 involved in sequence-specific DNA recognition by transposase

Base-specific recognition of thymine by Arg118 of Mos1 transposase (a) and guanine by Arg118 of Mboumar-9 transposase (b). Arg183 of Mos1 transposase is involved in pyrimidine recognition of the last adenine base in the IRR (c). The last base of Mboumar-9 IRL, guanine, is recognised base-specifically by Arg183 of Mboumar-9 transposase (d).
was investigated in section 4.2.4. Change of guanine to adenine has no impact on the step of in vitro cleavage, but has a stimulating effect on the whole transposition reaction. This can be due to the fact that adenine is bound less strongly than guanine to Arg183 of the transposase and the dissociation of the protein-inverted repeat DNA complex is eased, therefore increasing the protein turnover and making the integration site accessible for repair by the host enzymes.

4.2.5.2 Improving the specificity of Mboumar-9 transposase

Based on the analysis of the generated model of Mboumar-9 transposase we found that amino acid residues involved in sequence-specific recognition of the inverted repeat DNA of Mos1 transposase structure play a similar role in Mboumar-9 transposase, except for the residue 48. In Mos1 transposase Arg48 is involved in base-specific recognition of the G7 base of the non-transferred DNA strand (Figure 4-12 a). In case of Mboumar-9 the corresponding residue is threonine and the closest base is thymine (Figure 4-12 b). We asked a question if the mutation Thr48Arg of Mboumar-9 transposase (Figure 4-12 c) can improve the recognition of the inverted repeats and increase transposition efficiency. We also investigated whether the complementary mutation of the inverted repeat DNA T11G will have an effect on the transposition efficiency.

4.2.5.2.1 Construction

The change of threonine to arginine at the position 48 of Mboumar-9 transposase was performed with the primers No. 21 and 22 (see the list of primers) by site directed mutagenesis. The presence of the mutation of interest was confirmed by sequencing. The Thr48Arg Mboumar-9 was expressed and purified as described before (p. 49).

The T11G mutation of the left inverted repeat of Mboumar-9 was introduced using site directed mutagenesis with one pair of primers No. 19 and 20 (see the list of primers) on the pEPMboLL plasmid. This resulted in two mutations, one in each of the inverted repeats, as the inverted repeats of pEPMboLL are identical. The resulted pEPMboLL T11G was sequenced to confirm the presence of the mutations of interest.
Figure 4-12 Interactions of the residue 48 of Mos1 and Mboumar-9 transposases with the inverted repeat DNA

(a) Base-specific interaction of Arg48 of Mos1 transposase with G7 of the inverted repeat DNA.
(b) Orientation of the Thr48 of Mboumar-9 transposase to T11 of the IRL DNA (a residue, which corresponds to G7 of Mos1 IRR DNA).
(c) Predicted interaction of the Mboumar-9 Thr48Arg mutant with T11 of the inverted repeat DNA.
4.2.5.2.2 Activity of the Thr48Arg Mboumar-9 transposase mutant

To assess the activity of the generated mutants we performed the \textit{in vivo} transposition assay comparing the wild type Mboumar-9 transposase with the Thr48Arg mutant and MboLL with MboLL T11G transposons. The data of the \textit{in vivo} transposition efficiency is normalised to 30 colonies per 1 µg of donor DNA (Figure 4-13). The graph shows that substitution of the threonine at the position 48 to arginine does not have an effect on the transposition levels \textit{in vivo}. However, the wild type Mboumar-9 transposase is 60% less active on the MboLL T11G substrate, while Mboumar-9 Thr48Arg transposase is only 30% less active in the reaction \textit{in vivo}. This indicates that Arg48 may potentially be involved in the recognition of the guanine at the position 11 of the inverted repeat DNA by Mboumar-9 transposase.

4.2.6 Target site selection

The final steps of the transposition reaction are capturing of the target DNA and inserting of the transposon within the target. This location has to be selected by a transposase. Transposons insert with a duplication of the target sites, which are reported to be from 2 to 15 base pairs long (Ivics et al, 2009). Specificity of the area around the duplication sites is not fully understood. There is evidence that transposases prefer more bendable regions of the target DNA (Vigdal et al, 2002). For transposons of the \textit{Tc1/mariner} family there is duplication of TA nucleotides and no sequence-specific recognition of the area around the duplication site has been detected.

\textit{In vitro} experiments on Mos1 transposase have shown that in presence of Mg\textsuperscript{2+} the duplication site is TA, while in presence of Mn\textsuperscript{2+} this specificity can be reduced (Zhang et al, 2001).

4.2.6.1 Target site selection \textit{in vitro}

We used an \textit{in vitro} transposition assay to integrate Mos1 and Mboumar-9 transposons into a target plasmid DNA. The products of transposition were sequenced with primers No. 11 and 12 (list of primers on p. 140) to establish the sites of transposon insertions. We compared the effect of two ions, Mg\textsuperscript{2+} and Mn\textsuperscript{2+}, on the
Figure 4-13 *In vivo* transposition efficiency of the Thr48Arg Mboumar-9 transposase mutant

*In vivo* transposition efficiencies of the wild type Mboumar-9 transposase (Mbo) and Thr48Arg mutant are almost identical on the MboLL transposon substrate. Both wild type and mutant transposases are less active on the MboLL T11G transposon substrate. The efficiency drop for the wild type Mboumar-9 transposase is 60%, while for the mutant, which was proposed to recognise G11 of the inverted repeat DNA – is only 30%.
Improving *mariner* transposons for transgenesis

selection of the target sites; 16 different integration sites were aligned for both of the transposases in these two conditions. In the presence of Mg\(^{2+}\), Mos1 and Mboutmar-9 transposons are inserted selectively with a duplication of TA nucleotides (Figure 4-14 a and c). In the presence of Mn\(^{2+}\), the specificity was lowered for both transposases (Figure 4-14 b and d). For Mos1 transposase the preference was reduced more dramatically than for Mboutmar-9. There are 94 possible TA sites available for integration within the target plasmid. From 30 sequences analysed for Mos1 and 31 for Mboutmar-9 transposons there are 11 common sites chosen by two transposases independently. With random selection of the integration sites the number of the sites selected twice under these conditions will be from 8 to 12.

Analysis of the sites of integration for both transposases and frequencies of the same sites used, showed that integration occurred completely randomly within the target plasmid.

**4.2.6.2 Target site selection *in vivo***

*In vivo* transposition is a convenient assay for establishing the integration sites of transposase in a live cell. We have isolated DNA from the *E.coli* clones with a kanamycin cassette integrated in the genomic DNA. This chromosomal DNA was digested with EcoRI restriction endonuclease and cloned into a pBSKS(+) vector. The resulting plasmids were sequenced with the primers No. 11 and 12 (list of primers on p. 140) to find out the integration sites. In total 14 sites of Mos1 integrations and 20 for Mboutmar-9 transposase have been sequenced. The WebLogo alignment of the 20 nucleotides around the duplication site is present in Figure 5-15. The only conservation we observed was the TA duplication site. No other specificity at the level of primary DNA sequence has been found. In case of *in vivo* transposition AA duplications for Mos1 and TT duplications for Mboutmar-9 transposase are found. This fact may suggest that in a living cell Tc1/mariner transposases can use both Mg\(^{2+}\) and Mn\(^{2+}\) ions for their activity, because in the case of Mg\(^{2+}\), the duplication sites *in vitro* are restricted to TA only. Other factors can affect target site selection *in vivo* as well and alter the specificity: accessibility of the DNA, level of translation, presence of the proteins that can interact with transposase.
Improving *mariner* transposons for transgenesis

The inverted repeats distinction and target site selection

Figure 4-14 Integration sites of Mos1 and Mboumar-9 transposases *in vitro*

WebLogo alignment of the integration sites *in vitro* of Mos1 (a and b) and Mboumar-9 (c and d) in the presence of Mg\(^{2+}\) (a and c) or Mn\(^{2+}\) (b and d) divalent ions. In the presence of Mg\(^{2+}\) ions the duplication sites are limited to TA sequence, while in the presence of Mn\(^{2+}\) ions this specificity is lost and other than T and A nucleotides can be duplicated. We also observed duplications of three nucleotides in the presence of Mn\(^{2+}\).
Improving *mariner* transposons for transgenesis

The inverted repeats distinction and target site selection

Figure 4-15 Integration sites of Mos1 and Mboumar-9 transposases *in vivo*

WebLogo alignment of the 42 bases of the integration sites *in vivo* for Mos1 (a) and Mboumar-9 (b) transposases. The duplication site is in the middle of the sequence. The only significant preference found is a TA site, which is duplicated are a result of transposon integration.
4.3 Discussion
In this chapter we investigated the effects of the difference in the sequence of the left and right inverted repeats of Mos1 and Mboumar-9 on the cleavage step of transposition as well as on the \textit{in vitro} and \textit{in vivo} transposition efficiencies. We also attempted to improve the efficiency of Mboumar-9 transposition and discovered target sites \textit{in vitro} and \textit{in vivo} for both of the transposases. We found that Mos1 transposase cuts out the right inverted repeat more efficiently than the left inverted repeat, which confirms previously reported results for this transposon. The Mos1 transposon flanked with two right inverted repeats is 20 times more active than the wild type in the whole transposition reaction both \textit{in vitro} and \textit{in vivo}. It was reported that one inverted repeat has to stay unoccupied while transposase dimer binds to the other as the transposon with two right inverted repeats is more sensitive to overproduction inhibition (Jaillet et al, 2012). Overproduction inhibition is a mechanism of regulation of transposon activity by inhibiting transposition at high concentrations of transposase (Lohe & Hartl, 1996). If one of the inverted repeats has to stay unoccupied we would not have observed strong activation of the transposition efficiency of the MosRR transposon both \textit{in vitro} and \textit{in vivo} and the native LR form of the transposon should be the most optimal.

We also discovered that Mboumar-9 transposase prefers one of the inverted repeats: the IRL. This preference is not so marked as in the case of Mos1. The efficiencies of the Mboumar-9 transposition vary two to three times between the transposons flanked with the left or right inverted repeats compared to the naturally found transposon with asymmetrical inverted repeats orientation. The difference between the inverted repeats of Mboumar-9 is only 6\% (2 base pairs of 32), while the difference for Mos1 transposon is 14\% (4 base pairs of 28). We hypothesise that preference towards one of the inverted repeat correlates with the level of the inverted repeats’ divergence.

The \textit{in vitro} cleavage assay enables study of the cleavage steps of the transposition reaction in isolation. While the \textit{in vitro} transposition shows the activity of the enzyme for the whole reaction (cleavage plus integration) in a tube. Many factors can
influence protein and DNA stability in a tube, that is why the \textit{in vivo} transposition assay gives an advantage in studying protein-DNA complex interactions in physiological conditions of a living cell. In this chapter the \textit{in vivo} integration assay was carried out in bacterial cells. Combination of these three methods together gives more conclusive results on the inverted repeats distinction for the transposon studies, than a single assay studies. Nevertheless we realise that the transposons we are studying have only inverted repeat DNA of the transposon and the internal sequence is changed to the kanamycin resistance cassette for the purposes of the experiments. In nature, transposons contain an ORF of the transposase protein flanked with the inverted repeats. It is possible that secondary or tertiary structure of the naturally found transposons significantly differs from the structure of artificially constructed transposons and has impact on the recognition of the inverted repeats by transposase.

The difference of guanine-adenine base at the 3’ end of the transferred strand of the inverted repeats is found for at least twelve members of the \textit{Tc1/mariner} (Jaillet et al, 2012). The most active Mos1 inverted repeat has an adenine at the last position of the transferred strand. We noticed that the last base does not differ for Mboumar-9 transposon (Figure 4-1) and is involved in a tighter interaction by Arg183 than adenine is in Mos1 paired end complex (Figure 4-11 c and d). We investigated if a change of guanine to adenine has an effect on the transposase activity. Unexpectedly the G32A mutation of the Mboumar-9 left inverted repeat had no effect on the cleavage step, but had a significant stimulation effect on the transposition reaction – 3.5 times (Figure 4-5). We presume, that this effect takes place at the stage of the transposon integration into a new target site. This hypothesis requires further investigations and is a promising platform for revealing the role of the inverted repeats imperfection in the \textit{Tc1/mariner} transposon family. Finding if the transposition efficiency \textit{in vitro} is increased for the G32A mutant transposon and whether the difference is the same as \textit{in vivo}, can give an idea of the stability of the paired end complex \textit{in vitro}. Maybe the tighter interaction of the Arg183 with the last G is required for the stability of the PEC after the transposon is excised and before a target DNA is captured. Investigating the efficiency of the transposons with the
Improving *mariner* transposons for transgenesis

Inverted repeats carrying thymine or cytosine as the last base of the transferred strand will give an insight of the “ideal” last base for the highest transposition efficiency.

We modelled a structure of the Mboumar-9 transposase paired end complex based on the solved crystal structure of Mos1 transposase and analysed the quality of the generated model using Anolea energy. The total non-local energy of Mboumar-9 is higher than those for Mos1. There are two possible explanations, which can co-exist. The model may not represent the actual amino acid orientation of Mboumar-9 transposase or the interaction between two monomers of Mboumar-9 is tighter than it is for Mos1 transposase. Using this model we predicted the amino acid residues that interact base-specifically with the inverted repeat DNA and those correspond to the same residues of Mos1 transposase. Residues 44, 100, 118 and 183 are involved in base-specific recognition of the Mboumar-9 inverted repeats. Based on the generated model we did not find Pro184 and His293 to be oriented close to the DNA strands.

The attempt to improve the specificity of Mboumar-9 transposase to its inverted repeat – introduction of Thr48Arg mutation of the protein, resulted in no change in the protein activity *in vivo*. A corresponding T11G mutation of the left inverted repeat reduced the transposition efficiency by 60% for the wild type Mboumar-9 transposase and by 30% for Thr48Arg mutant. This observation suggests that the residue 48 might interact with the 11th base of the inverted repeat. Gln101 of Mos1 transposase correspond to Val101 of Mboumar-9 transposase. We did not observe possible interactions of Val101 with inverted repeat DNA of Mboumar-9. A Val101Gln mutation and corresponding A23G mutation in the inverted repeat might improve the specificity and the efficiency of transposition. Further experiments are required to check this hypothesis.

We used *in vitro* and *in vivo* transposition techniques to assess the integration sites for Mos1 and Mboumar-9 transposases. Both transposons integrate with the duplication of TA sequence. The selectivity is 100% to TA if Mg$^{2+}$ was used for *in vitro* transposition. When Mn$^{2+}$ was used as a divalent ion *in vitro* the selectivity to TA is reduced. For Mos1 transposition only 44% and for Mboumar-9 53% of integrations happen with duplication of TA sites. *In vivo* transposition assay gives an improved view on the transposase selectivity in a living cell. The integrations are

The inverted repeats distinction and target site selection
found mainly and TA sites, with only 5% insertions in TT or AA sites. Mos1 and Mboumar-9 transposases are representable members of Tc1/mariner family of transposable elements. They integrate with the duplication of TA nucleotides and do not shown any other preference of the DNA sequence around these sites.
Chapter 5: Discussion and future work
Transposable elements compose a big proportion of genomes and are significant contributors to genome diversity and evolution. Migration of transposable elements can result in loss of function of essential genes, activation of silenced genes or can bring in new phenotypes such as antibiotic resistance in bacteria. Since transposases can cut out, carry and integrate DNA sequences into new locations, nowadays they are exploited as potential tools for genome engineering. There are a number of critical parameters which have to be annotated and optimised in order for transposable elements to be able to serve as vectors for transgenesis. These include protein stability, precision of excision, stability of the protein-DNA complex, preferences for the integration site and absence of footprints after excision.

In this work we studied two transposases: Mos1 and Mboumar-9. These transposases are the most closely related active transposases of the Tc1/mariner family. We have designed codon and restriction sites optimised Mos1 and Mboumar-9 genes, enabling swap of the protein domains between two transposases. Mos1 and Mboumar-9 were selected in order to be able to create chimeric protein with the DNA-binding domain of Mboumar-9 and the catalytic domain of Mos1 to establish the orientation of the first and second strand cleavages (more details in section 2.2.10).

Mos1 and Mboumar-9 transposases were expressed in *E.coli* and purified using HPLC. N-terminal six histidine residues interfered with the protein activity *in vitro*, thus both transposases were purified as untagged proteins. Mos1 transposase is more stable in solution than Mboumar-9 as shown by TDA and SEC-MALLS. Mboumar-9 undergoes proteolysis at room temperature within one hour. Proteolysis of Mos1 transposase has been observed before: crystal trials of the full-length protein without DNA resulted in the formation of crystals of the Mos1 catalytic domain only (Richardson et al, 2006). Proteolysis of Mos1 occurred at the position 118 (Arg). Mboumar-9 transposase also contains arginine at the same position, which is within the flexible linker between the DNA-binding and catalytic domains. It is possible that proteolysis of Mboumar-9 occurs at the same position, but faster than for Mos1. Rapid Mboumar-9 degradation results in drop of protein activity after storage over short periods of time. Mutation of the residues around Arg118 of Mboumar-9 might stabilize the protein in solution and preserve its activity.
Analysis of the oligomeric state of Mos1 and Mboumar-9 transposases showed that in solution both transposases are dimers. Estimation of the oligomeric state of Mos1 truncations showed that the first helix-turn-helix domain (amino acids 1-55) is involved in the protein dimerization in solution. Knowing this fact we could eliminate putative structures in which the catalytic domain is involved in the protein dimerization (structures 2, 3, 7 and 8 from Figure 1-8).

Analysis of the first strand cleavage showed that Mboumar-9 cleaves three nucleotides inside the transposon end, similar to Mos1 first strand cleavage. Another band of a higher molecular mass (~51 bp) was registered for Mboumar-9 first strand cleavage. This could be due to the presence of Mboumar-9 recognition sequence in the flanking DNA of the oligonucleotides used in this assay or due to the formation of a hairpin intermediate. To test this an internal radioactive label can be introduced in the donor oligonucleotide, which could shed light on the position of an unexpected cleavage by Mboumar-9 transposase.

Notably Mboumar-9 transposase cleaved the first strand of Mos1 inverted repeat specifically, whereas Mos1 transposase did not cleave Mboumar-9 inverted repeat at the expected positions. Similar data has been observed for Himar1 transposase. Himar1 transposase cleaves Mos1 inverted repeat specifically, while Mos1 did not cleave Himar1 inverted repeat (O’Hagan, 2007). The cross-recognition of Mos1 inverted repeats by both Mboumar-9 and Himar1 transposases could be due to the presence of consensus sequence in the inverted repeats of Mos1 or due to lower specificity of Mboumar-9 and Himar1 towards a recognition sequence. Inability of Mos1 transposase to cleave the inverted repeats of Mboumar-9 or Himar1 could be due to either higher specialization of Mos1 transposase or higher specificity of Mboumar-9 and Himar1 inverted repeats.

Change of the second helix-turn-helix domain of Mos1 transposase to Mboumar-9 second helix-turn-helix resulted in the retention of the Mos1 pattern of bands after the first and the second strand cleavages of Mboumar-9 inverted repeat. This suggests that a change only in one of two DNA-recognition helix-turn-helix domains is not enough to gain the specificity towards Mboumar-9 inverted repeats. Change of
Improving *mariner* transposons for transgenesis

the HTH1 in Chim7 to Mboutmar-9 might be enough for a chimeric protein to perform a specific cleavage of the Mboutmar-9 IR. In this case a chimera protein might be a potential tool to establish the orientation of the subunits during the second strand cleavage to find out which orientation of transposase subunits in Figure 1-8 is the correct one (6, 9 or 10). If change in both DNA-binding domains is not enough to perform a specific cleavage this will suggest that the catalytic domain might play a role in the sequence specific recognition.

To investigate the difference in cleavage and transposition activities of the two imperfect inverted repeats of Mos1 (4 differences) and Mboutmar-9 (2 differences), we created transposons flanked with either identical left or right inverted repeats or both left and right as native transposons. Our data confirmed that Mos1 cleaves the right inverted repeat 10 times more efficiently that the left one in the *in vitro* cleavage assay. We found that Mboutmar-9 also has preference towards one of the inverted repeats, the left one. The whole transposition reaction of the element flanked with two right inverted repeats of Mos1 is 20 times higher than of the native left-right form and the element with two left inverted repeats is 50 times less efficient for *in vitro* transposition. For Mboutmar-9, preference for one of the inverted repeats is not so marked. This can be due to the fact that only two bases of the 32 bp inverted repeats differ, compared to Mos1 (4 bases of 28 bp inverted repeats differ). The conclusion is that if the inverted repeats are different most likely transposase binds, cleaves and/or integrates one of the inverted repeat more efficiently than the other one.

We found that changing the last base of Mboutmar-9 inverted repeat from guanine to adenine did not have any effect at the cleavage steps, but increased the whole transposition reaction 3.5 times, suggesting that the efficiency of integration might have increased. This could be because of the formation of an additional hydrogen bond between the 3’-OH terminal guanine and Arg183, making the complex of cleaved out transposon more stable (Figure 5-11, c and d). In the case of adenine only one hydrogen bond is formed with Arg183. This could ease the step of integration of excised transposon into the target TA dinucleotide sequence. Since the right inverted repeat of Mos1 has adenine as the 3’-OH terminal base and the left
inverted repeat has guanine at the terminal position it is possible that the stabilized complex of the transposase with the left inverted repeat is a coordinator for the strand transfer. It could be that the right inverted repeat can be integrated first, followed by the 3’-OH nucleophilic attack of the left inverted repeat.

Optimisation of the *in vitro* transposition assay resulted in an unexpected observation: growth of resistant colonies when the target plasmid was not present in the reaction. Finding the reason of this unexpected result led us to discover the *in vivo* transposition activity. We developed an *in vivo* assay in which transposase protein is co-transfected with the transposon donor plasmid, which results in transposon integrations into the genomic DNA of the recipient strain. We have validated this assay for both Mboumar-9 and Mos1 transposases and have direct confirmation of the integrations in bacterial and human cells. This assay can be further developed as a tool for random mutagenesis, genome sequencing, stable integration of genes in the cells that do not maintain plasmids and creation of libraries of random genes knock-outs. Future experiments would be to prove and optimise integration in other cell lines. Tn5 has similar properties of entering cells during electro-transfection, but it recognizes an 8 bp target site, while Mos1 and Mboumar-9 recognize only TA dinucleotide and can enter cells during chemical transfection.

How transposases of the *Tc1/mariner* family recognize TA dinucleotide target sequence still remains unclear. A structure of the target capture complex might give missing answers and lead to the generation of highly specific or target non-specific transposases. Directing the transposon integrations to particular sites or sequences would allow Mos1 or Mboumar-9 to be competitive candidates for the gene therapy. This could be achieved by fusing Mos1 or Mboumar-9 transposase to DNA-binding domains. Target non-specific transposases will be useful for improved coverage in random mutagenesis.

Future work to be continued on this project includes characterisation of the orientation of the second strand cleavage using chimera protein. Mutation of two mismatched residues of Mbo9 inverted repeats might improve transposition activity,
as well as mutating residues around Arg118 to improve Mboumar-9 transposase stability. Crystallization of Mboumar-9 with or without DNA and its direct comparison to the structure of Mos1 paired end complex would be a step forward in our understanding of the inverted repeat DNA recognition.
Chapter 6: Materials and Methods
6.1 Materials

6.1.1 Strains

\textit{E.coli} DH10B (Invitrogen)

Str\textsuperscript{R} F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 endA1 recA1 deoR Δ(ara,leu)7697 araD139 galU galK nupG rpsL λ- rpsL nupG tonA

\textit{E.coli} BL21 (DE3) (Stratagene)

fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS

λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5 (r\textsubscript{B}-, m\textsubscript{B}-)

\textit{E.coli} C43 (DE3) (Lucigen)

F\textsuperscript{-} ompT gal dcm hsdSB(r\textsubscript{B}- m\textsubscript{B}-)(DE3)

\textit{E.coli} S17-1 λpir (a gift from Garry Blakely)

Tmp\textsuperscript{R} Str\textsuperscript{R} hsdR pro recA RP4-2-Tc::Mu-Km::Tn7 λ pir

\textit{Schizosaccharomyces pombe} 972 (a gift from Ken Sawin)

HeLa cell line, stock from Oxford (a gift from Gracjan Michlewski)

6.1.2 Plasmids

pMos1 (Ap\textsuperscript{R}, lac\textsuperscript{Q}, P\textsubscript{trc}, rrnB, ColE1) 8.0 kb (lab stock)

pEP185.3 (Cm\textsuperscript{R}, mobRP4, oriR6K), 4.3 kb (a gift from Garry Blakely)

pBSKS(+) (Ap\textsuperscript{R}, lacPOZ', ColE1), 3 kb (lab stock)

pUC4K (Kn\textsuperscript{R}, Ap\textsuperscript{R}, ColE1), 3.9 kb (a gift from Garry Blakely)

pEGFP-C1 (Kn\textsuperscript{R}, Neo\textsuperscript{R}, ColE1, SV40 ori, f1 ori), 4.7 kb (a gift from Gracjan Michlewski)

6.1.3 Media

The percentage mentioned is weight per volume unless otherwise stated.
2xTY
1.6% Difco Bacto tryptone, 1% Difco Bacto yeast extract, 0.5% NaCl, adjust pH to 7.0.

GYT
10% (v/v) glycerol, 0.125% Difco Bacto yeast extract, 0.25% Difco Bacto tryptone.

SOC
2% Difco Bacto tryptone, 5% Difco Bacto yeast extract, 0.05% NaCl, 25 mM KCl. Adjust pH to 7.0, before use add 10 mM sterile MgCl₂ and 20 mM sterile glucose.

YE5S pH 5.6
5% Difco Bacto yeast extract, 3% glucose, 0.0225% of each supplement: Histidine, Leucine, Adenine, Uracil, Lysine.
For YE5S agar: also add 2% Difco Bacto Agar.

YPD
1% Difco Bacto yeast extract, 2% peptone, 2% dextrose.

6.1.4 Buffers

Antibiotic stocks
Kanamycin (50 mg/ml), Carbenicillin (100 mg/ml) and Streptomycin (50 mg/ml) in water, filter sterilised and stored at -20°C. Chloramphenicol (50 mg/ml) in 70% ethanol, stored at -20°C. Trimethoprim (50 mg/ml) in DMSO, stored at +4°C. Solution of G418 100 mg/ml was prepared in DPBS (Gibco®, No. 14190), filter sterilized and stored at -20°C.

Cell resuspension buffer (for protein purification)
20 mM PIPES pH 6.8, 400mM NaCl, 1 mM DTT, 5 mM MgCl₂.

6xLoading Buffer
0.25% bromophenol blue, 40% sucrose.
Improving *mariner* transposons for transgenesis

**Harju buffer**
2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris-Cl pH 8.0, 1 mM EDTA, filter sterilise.

**Hybridization solution**
10% dextran sulphate, 5xSSPE, 2% SDS, 1x Denhardt’s solution (Sigma), 0.1% herring sperm DNA. Add Denhardt’s solution and salmon sperm DNA last, before use.

**Protein storage buffer**
Mos1: 50 mM PIPES pH 7.5, 250 mM KCl, 1 mM DTT. For long term storage add glycerol to final concentration 50% (v/v), store at -20°C.
Mboumar-9: 50 mM PIPES pH 6.8, 500 mM NaCl, 5 mM MgCl₂, 1 mM DTT. For short or long term storage snap freeze in liquid nitrogen, store at -80°C.

**20xSSC**
3 M NaCl, 30 mM sodium citrate pH 7.0

**20xSSPE**
200 mM sodium phosphate pH 7.5, 3 M NaCl, 25 mM EDTA.

**1xTAE**
40 mM Tris pH 7.6, 20 mM acetic acid, 1 mM EDTA. Store as 50x stock.

**TE pH 7.5**
10 mM Tris-Cl, pH 7.5, 1 mM EDTA, filter sterilise.

**TEN**
10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA pH 8.0. Store as 10x stock.

**TFB1**
30 mM CH₃COOK, 100 mM RbCl₂, 10 mM CaCl₂, 50 mM MnCl₂, 10% glycerol. Adjust pH to 5.8 with diluted acetic acid, filter sterilise and store at 4°C for up to one month.
TFB2
10 mM MOPS pH 6.5, 75 mM CaCl₂, 10 mM RbCl₂, 15% glycerol. Adjust pH to 6.5 with KOH, filter sterilise and store at 4°C for up to one month.

5xTTE
445 mM Tris, 142.5 mM taurine, 2.5 mM EDTA.
### 6.1.5 Primers

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 After IRL BamHI</td>
<td>ctt acg ggt cga cac agt tgg tac tgg ttc g</td>
<td>SalI site underlined</td>
</tr>
<tr>
<td>2 After IRR BamHI</td>
<td>gtt agt gga tcc ccc att gct gga ata tct ttc gc</td>
<td>BamHI site underlined</td>
</tr>
<tr>
<td>3 Before IRL SacI</td>
<td>att cag age tca aag age gaa cgc aga cga g</td>
<td>SacI site underlined</td>
</tr>
<tr>
<td>4 Before IRR Sall</td>
<td>ata tgg gtc gac tgc ggc tta cet acc aga cc</td>
<td>Sall site underlined</td>
</tr>
<tr>
<td>5 CITFor</td>
<td>gag ctc TAT CAG GTG TGT CGG TAA TTC CTT</td>
<td>SacI site underlined, IRL in capitals, mutation in lower case</td>
</tr>
<tr>
<td>6 CITRev</td>
<td>AAG GAA TTA CCG ACA CAC CGT aTA gag ctc</td>
<td>SacI site underlined, IRL in capitals, mutation in lower case</td>
</tr>
<tr>
<td>7 d1-55MosFor</td>
<td>tac gtc ata tgg gcg att tgg atg tgg atg</td>
<td>NdeI site in bold</td>
</tr>
<tr>
<td>8 Left Xbal IRRMboKan</td>
<td>ggc act cta gAT ACC AGG AGT GTC GTT AAT TCT TTT CCG GTT TTT gtc gac cgc tga ggt ctc cct cg</td>
<td>Xbal site underlined, Sall site in italics, IRR in capitals</td>
</tr>
<tr>
<td>9 Left Xbal IRRMosKan</td>
<td>ggc act cta gAT ATC AGG TGT ACA AGT ATG AAA TGT CGT TTg tcg acc gtc gag gtc tgc etc g</td>
<td>Xbal site underlined, Sall site in italics, IRR in capitals</td>
</tr>
<tr>
<td>10 MosRev</td>
<td>tac ggc teg agt tat cta aag tac tgg cc</td>
<td>XhoI site in bold underlined</td>
</tr>
<tr>
<td>11 pMK1870F</td>
<td>gtt tcc cgt tga ata tgg ctc</td>
<td>for sequencing of flanking DNA</td>
</tr>
<tr>
<td>12 pMK867R</td>
<td>act ttc tgg ctc gat gat gg</td>
<td>for sequencing of flanking DNA</td>
</tr>
<tr>
<td>13 Right SacI IRLMboKan</td>
<td>gta ctc age tcT ACC AGG TGT GTC GTT AAT TCC TTT CCG GTT TTT gtc gac ggg aaa ggc agc tgg tgc ctc c</td>
<td>SacI site underlined, Sall site in italics, IRL in capitals</td>
</tr>
<tr>
<td>14 Right SacI IRLMosKan</td>
<td>gta ctc age tcT ACC AGG TGT ACA AGT AGG GAA TGT CGG TTg tcg acg gga aag cca cgt tgg gtc cgg</td>
<td>SacI site underlined, Sall site in italics, IRL in capitals</td>
</tr>
<tr>
<td>15 SacI IRL MboKan</td>
<td>gta ctc age tcT ACC AGG TGT GTC GTT AAT TCC TTT CCG GTT TTT gtc gag cgc tga ggt ctc cct cg</td>
<td>SacI site underlined, Sall site in italics, IRL in capitals</td>
</tr>
<tr>
<td>16 SacI IRL MosKan</td>
<td>gta ctc age tc TAC CAG GTG TAC AAG TAG GGA ATG TCG GTT gtc gag cgc tga ggt ctc cct cg</td>
<td>SacI site underlined, Sall site in italics, IRL in capitals</td>
</tr>
<tr>
<td>17 Sall NeoGen For</td>
<td>tcc ata gtc gag ctc cct gag ggc gaa aga acc</td>
<td>Sall site in italics</td>
</tr>
</tbody>
</table>
Improving *mariner* transposons for transgenesis

<table>
<thead>
<tr>
<th>Step</th>
<th>Oligo/IRR/BS</th>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>Sall NeoGen Rev</td>
<td>tcc ata gtc gac atgagt aac ctc agg eta tgg c</td>
<td>Sall site in italics</td>
</tr>
<tr>
<td>19</td>
<td>T22GFor</td>
<td>GTC GGT AAT TCC TgT CCG GTT TTT gtc gac</td>
<td>Sall site in italics, IRL in capitals, mutation in lower case</td>
</tr>
<tr>
<td>20</td>
<td>T22GRev</td>
<td>gtc gac AAA AAC CGG AcA GGA ATT ACC GAC</td>
<td>Sall site in italics, IRL in capitals, mutation in lower case</td>
</tr>
<tr>
<td>21</td>
<td>T48RFor</td>
<td>CATTCGCACCTGCAGAAAggTG GTTCGCAGTTCAATGC</td>
<td>Mutated bases in lower case</td>
</tr>
<tr>
<td>22</td>
<td>T48RRev</td>
<td>GCATTTCAGTGGCCAAACC AccTTTCGCAGTTCAATGC</td>
<td>Mutated bases in lower case</td>
</tr>
<tr>
<td>23</td>
<td>Xbal IRRMboKan2</td>
<td>ggc act cta gat ACC AGG AGT GTC GGT AAT TCT TTT CCG GTT TTT gtc gac aaa gcc agc tgt tgt etc</td>
<td>Xbal site underlined, Sall sire in italics, IRR in capitals</td>
</tr>
<tr>
<td>24</td>
<td>Xbal IRRMosKan2</td>
<td>ggc act cta gat ATC AGG TGT ACA AGT ATG AAA TGT CGT TTg acg cca cgt tgt gtc tc</td>
<td>Xbal site underlined, Sall sire in italics, IRR in capitals</td>
</tr>
<tr>
<td>25</td>
<td>NoSalI pBS For</td>
<td>gtttatcgataccgacgactcgaggggg</td>
<td>For deletion of Sall site</td>
</tr>
<tr>
<td>26</td>
<td>NoSalI pBS Rev</td>
<td>cccctcgggtcgtcgtatcagaage</td>
<td>For deletion of Sall site</td>
</tr>
<tr>
<td>27</td>
<td>ForKan</td>
<td>tatggetcaacacccc</td>
<td>For sequencing of flanking DNA</td>
</tr>
<tr>
<td>28</td>
<td>RevKan</td>
<td>gtcaccactaacaaga</td>
<td>For sequencing of flanking DNA</td>
</tr>
</tbody>
</table>

### 6.2 Methods

#### 6.2.1 Bacterial work

##### 6.2.1.1 Bacterial growth, storage

Bacterial cells were grown in liquid medium LB, or on LB agar with addition of the appropriate antibiotics at the following concentrations: Kanamycin 50 µg/ml, Carbenicillin (Ampicillin analogue) 100 µg/ml, Streptomycin 50 µg/ml, Trimethoprim 50 µg/ml, Chloramphenicol 50 µg/ml.

##### 6.2.1.2 Bacterial transfection

#### 6.2.1.2.1 Rb²⁺ transfection

A single colony was inoculated in 5 ml LB over night culture from a freshly streaked plate, incubated at 37°C with 250 rpm agitation. The culture was diluted 1:200 into pre-warmed 100 ml LB with 20 mM MgCl₂ and incubated at 37°C with 250 rpm
agitation until OD_{600} reached 0.48 (for *E.coli* DH10B strain) or 0.6 (for *E.coli* S17 \( \lambda \) pir strain). The culture was transferred to a chilled 250 ml centrifuge bottle and incubated on ice for 10 min. The cells were harvested for 5 min at 5,000 rpm at 4\(^\circ\)C. The pellet was resuspended in 40 ml of cold TFB1 buffer by gently pippeting and incubated on ice for 5 min, the cells were pelleted for 10 min at 3,000 rpm at 4\(^\circ\)C. The pellet was gently resuspended in 4 ml of cold TFB2 per 100 ml of culture, incubated on ice for 15 min. Aliquots of 100-200 \( \mu \)l were dispensed into pre-chilled Eppendorf tubes. Tubes were stored at -80\(^\circ\)C after snap freezing in liquid nitrogen.

For transfection cells were thawed on ice, 1-100 ng of DNA was added in \( \leq 20 \) \( \mu \)l volume to 100 \( \mu \)l of cells, mixed by gentle flicking, incubated on ice for 30 min. Heat shock was performed for 90 sec at 42\(^\circ\)C, the tubes rested on ice for 1 min. For recovery 400 \( \mu \)l SOC was added prior to incubation at 37\(^\circ\)C for 60 min. Cells were plated onto suitable selective media.

**6.2.1.2.2 Ca\(^{2+}\) transfection**

A single colony was inoculated in 5 ml 2xTY medium, incubated at 37\(^\circ\)C with 250 rpm agitation. The over night culture was diluted 1:100 or 1:200 in 100 ml 2xTY medium, incubated at 37\(^\circ\)C with 250 rpm agitation until OD_{600} was 0.2-0.4. The cells were harvested at 4,000 rpm for 5 min at 4\(^\circ\)C and were kept on ice during the next steps. The pellet was resuspend in 50 ml of 0.1 M ice-cold sterile CaCl\(_2\), incubated on ice for 1-2 hours (30 min minimum). The cells were harvested at 4,000 rpm for 5 min at 4\(^\circ\)C. The pellet was gently resuspended in 1 ml of 0.1 M ice-cold sterile CaCl\(_2\). For storage sterile glycerol was added to a final volume of 10\% (v/v), the cell suspension was mixed, incubated on ice for 30 min, stored at -80\(^\circ\)C after snap freezing in liquid nitrogen.

For transfection 1-10 ng of DNA was added to 100 \( \mu \)l aliquots, incubated on ice for 10-30 min. Heat shock was performed at 42\(^\circ\)C for 90 sec. For recovery 900 \( \mu \)l of SOC medium was added prior incubation for 30-60 min at 37\(^\circ\)C with 250 rpm agitation. The cells were plated out on appropriate selective media (Dagert & Ehrlich, 1979).
6.2.1.2.3 Electro-transfection
We inoculated over night culture from a single colony on a fresh agar plate, incubated at 37°C with 250 rpm agitation. Diluted over night culture 1:100 into 500 ml of pre-warmed LB medium and grew until OD600 is 0.35-0.4. Transferred the flasks on ice-water bath for 15-30 min, swirled occasionally, transferred culture to ice-cold centrifuge bottles. Harvested cells at 2,500 rpm (1,000 g) for 15 min at 4°C. Resuspended the cell pellet in the same volume (500 ml) of ice-cold sterile H2O. Harvested cells at 1,000 g for 20 min at 4°C. Resuspended the cell pellet in \( \frac{1}{2} \) volume (250 ml) of ice-cold sterile 10% (v/v) glycerol. Harvested cells at 1,000 g for 20 min at 4°C. Resuspended the cell pellet in \( \frac{1}{50} \)th volume (10 ml) of ice-cold sterile 10% (v/v) glycerol. Harvested cells at 1,000 g for 20 min at 4°C, carefully removed all the supernatant and resuspended the pellet in 1 ml of ice-cold GYT medium by gentle swirling. Measured OD600 of a 1:100 dilution and diluted the cell suspension to a concentration of 2 \( \times 10^{10} \) to 3 \( \times 10^{10} \) cells/ml. The OD600 of 2.5 \( \times 10^8 \) cells/ml is 1.0.

For transfection added 10 pg to 25 ng of DNA in a final volume of 1-2 µl to 40 µl of cells, incubated on ice for 30-60 sec. Transferred the DNA/cell mixture to an ice-cold electroporation cuvette (0.2 cm gap). Delivered an electrical pulse of 25 µF capacitance, 2.5 kV and 200 ohm resistance for 4-5 msec (Bio-Rad GenePulser). As quickly as possible after the pulse added 1 ml of SOC medium at room temperature, transferred the cells to 15 ml Falcon tubes and recovered for 1 hour at 37°C with 200 rpm agitation. Plated out different volumes on LB agar plates containing the appropriate antibiotics (Sambrook & Russell, 2001).

6.2.2 Yeast manipulations
6.2.2.1 Yeast transfection
We inoculated a pre-culture in 5 ml of YE5S medium, incubated at 25°C for 24 hours. Inoculated the ‘real culture’ in a desired volume (20 ml for one transfection), incubated at 32°C in YE5S until OD595 is 0.8-1.0. To work out the volume of real culture (V) we used the Equations 2 and 3:
Improving mariner transposons for transgenesis

\[ V = \frac{OD_{wanted} \times V_{wanted}}{OD_{preculture} \times 2^n} \]  
\[ n = \frac{t}{gt} \]

Equation 2

Equation 3

\( V_{wanted} \) is 20 ml per one transformation, \( t \) – in how many hours culture should be ready, \( gt \) – generation time, which is 2.17 hours for \( S.pombe \) in YE5S medium at 32°C growth temperature.

We pelleted 20 ml of cells for each transfection at 4,000 rpm, 2-5 min at room temperature. Resuspended the pellet in 20 ml of sterile dH2O and spun again. Resuspended the pellet in 1 ml of sterile dH2O, transferred into an Eppendorf tube. Spun 1 min at 13,000 rpm, resuspended in 1 ml 0.1 M LiOAc/TE pH 7.5. Spun 1 min at 13,000 rpm, resuspended in the total volume of 100 µl of 0.1 M LiOAc/TE pH 7.5. Added DNA (20-30 µg for linear DNA, 2 µg for plasmid DNA), incubated at room temperature for 10 min. Add 260 µl of 40% PEG (MW 3350 or 4000)/LiOAc/TE, mixed gently, incubated for 30 min at 30°C. Added 43 µl of DMSO, heat shock was performed at 42°C for 5 min. Spun for 1 min at 13,000 rpm, resuspended the pellet in 1 ml of sterile dH2O. Spun again, resuspended in 500 µl of sterile dH2O. Plated onto two YE5S agar plates (250 µl each). Incubated for 24 hours at 32°C until the dense lawn of cells appeared. Replica plated to YE5S agar containing 100 µg/ml of G418, incubated for 48 hours at 32°C. Replica plated again to YE5S agar with G418, incubated for 24 hours at 32°C. Picked individual colonies and streaked on sectored plates, incubated for 48 hours until single colonies appear. Replica plated on YE5S agar with G418 and incubated for further 24 hours at 32°C. Discarded the clones in which no colonies had grown or in which only some colonies had grown throughout the streak. From each stable clone picked a colony and patched it on a fresh YE5S agar plate.

6.2.2.2 Replica plating

Fixed two Whatman filter papers on a replica-plating block, placed a plate on the filters, pressed lightly, removed the plate. Placed a lid of the used plate on the filter
and pressed hard, removed the lid. Placed a fresh plate with YE5S agar on the filter, pressed lightly, removed. Removed the filters from the block and discarded.

### 6.2.2.3 Isolation of yeast genomic DNA
We grew liquid culture for 20-24 hours at 30°C in YPD. Pelleted 1.5 ml at 13,000 rpm for 1-5 min. Added 200 µl of Harju buffer, immersed tubes in a dry ice-ethanol bath for 2 min, transferred to in a 95°C water bath for 1 min. Repeated dry ice-ethanol bath and the hot bath steps. Spun for 3 min at 13,000 rpm, room temperature. Transferred the upper aqueous phase to an Eppendorf tube, containing 400 µl ice-cold ethanol. Mixed by inversion, incubated at room temperature for 5 min (or -20°C to increase the yield). Spun down the DNA for 5 min at 13,000 rpm, removed the supernatant. Washed the pellet with 500 µl of 70% ethanol. Spun for 5 min at 13,000, removed the supernatant. Air dried the pellet at room temperature or for 5 min at 60°C in a vacuum dryer. Resuspended in 25-50 µl of TE of water (Harju et al, 2004).

### 6.2.3 Mammalian cells manipulations

#### 6.2.3.1 Growing and maintenance of cell culture
A frozen stock of cell was thawed in 37°C water bath and inoculated into T75 plate containing 20 ml of pre-warmed (37°C) DMEM (Gibco®, No. 41966), containing 10% fetal bovine serum (SIGMA, F7524) and 1% penicillin-streptomycin (SIGMA, P4333). The cells were incubated for 24 hours at 37°C until confluent.

To split the cells from T75 plate, the growing medium was aspirated and the cells were washed with 10 ml (2 ml for one well of the 6-well plate) of pre-warmed DPBS (Gibco®, No. 14190-094). Two ml (0.5 ml for one well of 6-well plate) of pre-warmed trypsin (TrypLE Express, Gibco®, No. 12605-010) were added, incubated for 10 min at 37°C. After the cells are dissociated (using gentle taping to dislodge the cells), 10 ml (1.5 ml for one well of 6-well plate) of DMEM was added to the T75 flask and desired volume of cells was inoculated for further applications. Usually for inoculation of one 6-well plate 1 ml of trypsinized cells mixed with 12 ml of DMEM medium was used (to dispense 6x2ml) or 2 ml of cells and 11 ml of medium.
To prepare a stock of cells for storage, cells were trypsinized, spun down for 5 min at 400 rpm at room temperature, medium was aspirated and the pellet was carefully resuspended in 1 ml of the pre-warmed freezing solution (40% (v/v) fetal bovine serum, 10% (v/v) DMSO in DMEM). Aliquots (500 µl) were dispensed in cryotubes and slowly frozen in polystyrene box at -80°C for ≥24 hours before transferring to storage in liquid nitrogen.

6.2.3.2 Transfection
Cells were inoculated a day before transfection in several dilutions. For transfection with Lipofectamine® 2000 (Life Technologies™) cells of 60-70% confluence were used. The growth medium was aspirated, cells were washed with 2 ml of DPBS (Gibco®), then with 2 ml of Opti-MEM® Reduced Serum (Gibco®, No. 31985).

For each transfection reaction 1 tube containing 245 µl of Opti-MEM® with 5 µl of Lipofectamine® 2000 and 1 tube with 230 µl of Opti-MEM® with 20 µl of DNA (if DNA is in less volume the volume of Opti-MEM® was increased to the final volume of 250 µl) were prepared, incubated at room temperature for 5 min. The solution containing DNA was added drop-wise to the tubes containing Lipofectamine® 2000 (no vortexing or mixing). Incubated at room temperature for 30 min and each 500 µl was added drop-wise to 1.5 ml of DMEM in each well, the plate was rocked gently and incubated for 24 hours prior medium changing and/or splitting the cells for selection.

6.2.3.3 Fixing and staining of colonies
When the colonies are 1-3 mm in diameter, medium was aspirated, colonies were washed with DPBS and fixed with ice-cold methanol for 5-10 min and stained with 0.1% brilliant blue in DPBS for 15 min. The stain was removed and the wells were rinsed with water to clear the background.

6.2.4 Nucleic acids work
6.2.4.1 Agarose gel electrophoresis
1% (w/v) agarose (Invitrogen) gels were made using 1xTAE buffer with addition of 5 µl per 100 ml of SafeView nucleic acid stain (NBS Biologicals). After the gel is
Improving *mariner* transposons for transgenesis

set, the DNA samples were loaded with 6xLoading buffer in a maximum volume of 24 µl. Electrophoresis was performed in 1xTAE buffer at the constant voltage of 70 V at room temperature. The gel was visualised with BioRad GelDoc XR System. Quantification of bands was performed with ImageJ ver. 10.2 software.

The molecular markers used in this work are: 1 kb DNA ladder (NEB N3232S) and supercoiled DNA ladder (NEB N0472S).

6.2.4.2 Annealing and purification of oligonucleotides

Desalted oligonucleotides (IDT®) were dissolved in dH₂O to a final concentration of 1 mM – 100 µM and stored at -20°C. Annealing was performed in 1xTEN buffer by slow cooling from 95°C to room temperature over 3-4 hours.

Annealed oligonucleotides were purified by anion exchange chromatography on MiniQ 4.6/50 PE column attached to Ettan ÄKTA Purifier (GE Healthcare Life Sciences). Sample was loaded in 100 µl volume at 0.25 ml/min flow rate in buffer containing 25 mM Tris pH 7.5/300 mM NaCl, which is 30% of the high salt buffer (25 mM Tris pH 7.5/1 M NaCl). After 3 CV of equilibration elution was performed with linear gradient of high salt up to 100% of the high salt buffer. The DNA peak eluted at the conductivity of 70-91 mS/cm depending on the oligonucleotide length. The DNA was ethanol precipitated and diluted in the desired volume of TE or dH₂O.

6.2.4.3 Denaturating DNA gel electrophoresis

Polyacrylamide gel (8%) was prepared as follows: 18 g urea, 8 ml 5xTTE, 8 ml 40% (19:1) acrylamide, dH₂O to 40 ml, 280 µl of 10% APS, 40 µl TEMED. Before adding APS and TEMED, the solution was filtered through a 0.22 µm filter. The gel was prepared on Bio-Rad Protean® II xi electrophoresis cell. After incubation for 40 min at room temperature to set, the comb was carefully removed under running cold water to preserve straight wells. The gel was pre-run in 0.5xTTE buffer at constant power of 10 W for at least one hour. Before sample loading wells were washed with 0.5xTTE using 20 ml syringe and a 0.45 mm needle. Usually 10 µl of the sample was loaded into each well without dye. To visualise migration of the gel 2 µl of loading buffer containing 0.25% bromophenol blue and 0.25% xylene cyanol
was loaded into the outer wells. Migration was performed at constant power of 5 W for 5-10 min until bromophenol blue and xylene cyanol dyes were separated, then at 10 W for further 1-1.5 hours and stopped when bromophenol blue dye front was $\frac{1}{3}$rd from the end of the gel.

The gel was removed, rinsed with dH$_2$O and fixed in 10% ethanol/5% acetic acid for 10 minutes at room temperature without shaking prior to imaging on the LI-COR system using settings: ProteinGel, resolution: 169 µm, quality: medium, focus offset: 0.5 mm, channels: 700, intensity: 8.5.

**6.2.4.4 Ethanol precipitation**

To precipitate DNA from solution we added CH$_3$COONa pH 5.2 to a final concentration of 300 mM, MgCl$_2$ to a final concentration of 10 mM. Added 3 volumes of 96% ethanol at room temperature, incubated on ice for ≥1 hour. Spun down at 13,000 rpm 4°C for 30 min. Discarded supernatant, added 1 ml of 70% ethanol at room temperature, spun down at 13,000 rpm for 15 min at 4°C. Discarded supernatant, spun again for 1 min and aspirated the residual ethanol with pipette. Dried the pellet at room temperature, resuspended in desired volume of TE or dH$_2$O.

**6.2.4.5 Inverted PCR**

Approximately 1 µg of genomic DNA was digested with 1 µl of EcoRI endonuclease (NEB R0101S) in 20 µl for 3 hours at 37°C, the endonuclease was heat inactivated for 20 min at 65°C. Two dilutions (1:10 and 1:20) were prepared for self-ligation of the genomic DNA fragments in a final volume of 20 µl at room temperature with Rapid DNA Ligation Kit (Roche) for 20 min.

PCR was performed with Phusion High-Fidelity Polymerase (ThermoScientific) in a final volume of 20 µl: 400 µM dNTPs (Invitrogen), 500 µM of primers No. 27 and 28, 1 µl of DNA ligation mixture, 3% DMSO, 0.2 µl of Phusion Polymerase, dH$_2$O to 20 µl. Initial denaturation step was 98°C for 30 sec, followed by 30 cycles of 98°C for 30 sec, 56°C for 20 sec, and final elongation of 72°C for 7 min. We added 4 µl of 6xLoading buffer to 20 µl of PCR reaction and separated DNA fragments on
1% (w/v) agarose gel, visualised DNA with SafeView stain. The brightest bands were cut out, gel purified (Qiagen) and sequenced.

6.2.4.6 Isolation of the genomic DNA from bacterial cells
We grew 5 ml of over night culture in LB medium, harvested cells from 2 ml of the culture in a 2 ml Eppendorf tube for 2 min at 13,000 rpm. The cell pellet was resuspended in 567 µl TE by repeated pipetting, 30 µl of 10% SDS and 3 µl of 20 mg/ml proteinase K were added, mixed and incubated for 1 hour at 37°C. We added 100 µl of 5 M NaCl and mixed thoroughly, added 80 µl CTAB/NaCl solution (65°C), mixed and incubated for 10 min at 65°C. Added equal volume (780 µl) of chloroform/isoamyl alcohol (24:1), vortexed thoroughly, spun for 5 min, transferred the supernatant to a new Eppendorf tube. Added equal volume of phenol;chloroform;isoamyl alcohol (25:24:1), vortexed vigorously, spun for 5 min and transferred the supernatant to a new Eppendorf tube. Added 0.6 volume of isopropanol and mixed gently until DNA precipitated. Spun for 1 min at 13,000 rpm, discarded the supernatant, added 1 ml of 70% ethanol, spun for 5 min at 13,000 rpm. Discarded the supernatant, air-dried the pellet and resuspended in the desired volume (for example 100 µl) of dH2O or TE. The average yield was around 50 µg.

To remove RNA we added 1 µl of RNase A (Qiagen) and incubated for 1 hour at 37°C prior to restriction digest, gel electrophoresis or Southern blotting.

6.2.4.7 Isolation of plasmid DNA
Isolation of plasmid DNA from 1.5 ml of over night culture was performed using QIAprep Spin MiniPrep Kit according to manufacturer protocol. Elution of DNA from the column was done with pre-warmed to 70°C dH2O, the column was incubated at room temperature for 0.5-2 hours to increase the DNA yield.

Isolation of DNA from 250 ml of over night culture was performed by the protocol developed in Finnegan’s Lab by Eve Hartswood and Jim Brodie:
We grew 250 ml of over night culture and pelleted at 4,000 rpm for 10 min, drained the pellet thoroughly and disrupted by vigorous vortexing prior to adding 10 ml of resuspension solution (50 mM glucose/25 mM Tris pH 8.0/10 mM EDTA/50 mg
lysozyme). Incubated at room temperature for 10 min. Lysis solution (200 mM NaOH/1% SDS) was added and the mixture was inverted gently until solution cleared and became even (no vortexing), incubated on ice for 10 min. Neutralisation solution (3 M CH₃COOK/1.3 M CH₃COOH) was added and mixed promptly by inversion, incubated on ice for 15 min. The cell debris were pelleted for 15 min at 6,000 rpm. The supernatant was filtered by passing through glass wool in a small funnel. Added 0.6 volume of isopropanol, mixed, incubated for 5 min at room temperature. Spun for 10 min at 3,500 rpm and discarded the supernatant. Briefly rinsed the pellet with 70% ethanol and resuspended still wet pellet in 1 ml dH₂O, incubated on ice for 5 min. Added equal volume of 5 M LiCl (stored at -20°C), mixed and incubated on ice for 5 min, spun for 5 min at 13,000 rpm. Transferred supernatant to a new Eppendorf tube on ice, added equal volume of isopropanol to the supernatant. Incubated on ice for 10 min, spun for 5 min at 13,000 rpm and discarded the supernatant. Air-dried the pellet and dissolved in 300 µl of 10 mM Tris pH 8.0. Added 1 µl of DNase free RNase (10 mg/ml, Sigma) and incubated for 30 min at 37°C, then transferred to ice. Added an equal volume of 15% PEG/1.6 M NaCl, incubated on ice for 5 min, spun for 5 min at 13,000 rpm, discarded the supernatant and spun again to remove all the residual liquid by pipette. Resuspended still wet pellet in 300 µl of 10 mM Tris pH 8.0. Added equal volume of phenol:chlorophorm:isoamyl alcohol (25:24:1), vortexed and spun down for 4 min at 13,000 rpm. Kept the upper phase avoiding any precipitant. Added 1/20 volume 3 M CH₃COONa pH 5.6, mixed, added 2 volumes of 96% ethanol, mixed well (no vortexing). Large visible precipitant was pelleted for 5 min at 13,000 rpm, the pellet was washed with 1 ml of 70% ethanol. The ethanol was removed and the pellet was air dried prior to resuspension in the desired volume (for example 100 µl) of dH₂O or TE.

6.2.4.8 Ligation
Ligation of around 100 ng of DNA was performed with NEB BioLabs T4 Ligase (M0202T) at 4°C over night. The whole ligation mixture was used to transfrect the competent cells.
6.2.4.9 Phenol extraction
Equal volume of TE saturated Phenol (Sigma 77607) was added to the sample, vortexed vigorously for 1-2 min and spun down for 4 min at 13,000 rpm, room temperature. The upper phase was carefully transferred to a new Eppendorf tube.

6.2.4.10 Restriction digest
Restriction digest was performed with NEB BioLabs enzymes following the manufacturer protocol.

6.2.4.11 Sequencing
The BigDye® Terminator v3.1 Cycle Sequencing (Applied Biosystems) was done as follows: 3 µl of dH₂O, 2 µl of sequencing buffer, 2 µl of BigDye, 1 µl of 3.2 pmole/µl primer, 2 µl of DNA were placed in 0.2 ml PCR tubes. The PCR program was 95°C for 30 sec (denaturation), 50°C for 20 sec, 60°C for 4 min (25 cycles), 60°C for 1 min 15 sec (final extension). Sanger sequencing run was performed by The GenePool Facilities, Edinburgh. The chromatograms were analysed with ApE software– A plasmid Editor v.1.17.

6.2.4.12 Southern Blotting
Genomic DNA (1 µg) was separated on 1% (w/v) agarose gel after endonuclease digestion with EcoRI enzyme (NEB R0101S). The gel was washed twice with dH₂O to wash out the SafeView stain. The gel was then incubated with the depurinisation solution (0.2 M HCl) for 20 min with gentle shaking, rinsed with dH₂O. The DNA was denaturated by 2x15 min incubation with 0.5 M NaOH/ 1.5 M NaCl with gentle shaking, rinsed with dH₂O. Neutralised the hydroxide by 30 min incubation in 0.5 M Tris pH 7.5/ 1.5 M NaCl with gentle shaking, rinsed with dH₂O. Equilibrated for 30 min in 20xSSC solution (3 M NaCl/ 30 mM sodium citrate) with gentle shaking.

Capillary transfer was performed over night at room temperature as follows: a tray with 20xSSC solution, a glass plate on the top of the tray, long blotting paper soaked in 20xSSC (falling into the solution in the tray), 2x blotting papers (size of the gel), gel facing down, nylon transfer membrane for nucleic acids (size of the gel), 2x blotting papers (size of the gel), a pile of tissue papers (approximately 10 cm height),
Improving \textit{mariner} transposons for transgenesis

glass plate, weight (approximately 3 kg). After incubation ≥20 hours the membrane 
was rinsed with dH\textsubscript{2}O and cross-linked for 90 sec under UV (UV Stratalinker 1800).

Prehybridization was performed in a plastic bag at 55°C for 1.5 hours in 8 ml of 
hybridization solution with 120 rpm agitation. Hybridization was performed at 55°C 
for 15 hours in a plastic bag containing 5 ml of hybridization solution with 3 µl of 
100 µM fluorescently labeled 28 bp Mos1 IRR (5’ IRDye® 700, IDT) with 120 rpm 
agitation. The membrane was washed for 5 min at room temperature in 2xSSPE, 
5 min at 40°C min 2xSSPE, 2x for 15 min at 50°C in 2xSSPE/ 1% SDS, 2x for 
15 min at 50°C in 0.2xSSPE. Fluorescent dye was visualized on LI-COR Imaging 
System using Odyssey software.

\textbf{6.2.4.13 Transposons}

\textbf{6.2.4.13.1 Transposons for cleavage assays}

Mos1 transposons for cleavage assay were created by four-way-ligation (Figure 6-1).
The right hand inverted repeat of Mos1 was amplified with the primers No. 2 and 4 
from pMos1 template (Zhang et al, 2001), generating a product of 0.5 kb. The left 
hand inverted repeat of Mos1 was amplified with the primers No. 1 and 3, resulting 
in 0.6 kb fragment. To be able to select the products of transposition we introduced a 
kanamycin resistance (kanR) gene of 1.3 kb from pUC4K plasmid in between the 
inverted repeats of new transposons. The kanR gene was cut out by SalI restriction 
endonuclease. Two inverted repeats and a resistance gene were ligated with the 
pBSKS(+) plasmid (3 kb), digested with BamHI and SacI, in mass ratio 5:5:5:1 to 
generate pKanLR plasmid, where transposon was carrying both inverted repeats as in 
a naturally found form. Only one of the PCR products (containing IRL or IRR), kanR 
cassette and the plasmid backbone were used to create pKanRR plasmid (transposon 
is flanked with two right hand inverted repeats) and pKanLL plasmid (transposon is 
flanked with two left inverted repeats). The presence of the inverted repeats in all 
three resulting plasmids was confirmed by sequencing (See Appendix, sequences 
No. 1, 2 and 3).

Since pBSKS(+) plasmid contains ColEI origin of replication, the newly generated 
plasmids could not have been used for \textit{in vitro} transposition assay described before
Improving mariner transposons for transgenesis

Materials and Methods

**Figure 6-1 Four-way-ligation to create three transposons for Mos1 cleavage.**
Inverted repeats were amplified from pMos1 plasmid with the primers introducing SalI (cyan), BamHI (green) or SacI (violet). Kanamycin resistance cassette was cut out by SalI sites from pUC4K plasmid. pBSKS+ plasmid digested with BamHI and SacI was used as a backbone for new transposons. Four fragments were ligated together to create three different transposons (pKanLL, pKanLR, pKanRR) for *in vitro* cleavage assay.
(Zhang et al, 2001), as the transposon donor plasmid would have been replicating in the recipient strain together with the products of transposition.

6.2.4.13.2 Transposons for integration assays
As a backbone for transposons for integration assay we used pEP185.2 plasmid with a conditional origin of replication oriR6K. This plasmid could only replicate in special strains such as *E.coli* S17 lambda pir, which expressed P1 protein. The transposons were cut out with BamHI and SalI (from pKanLR) or BamHI alone (from pKanRR) or SalI alone (from pKanLL) and ligated with pEP185.2 plasmid by the appropriate sites. Resulting plasmids pEPLL, pEPLR and pEPRR can be used for *in vitro* integration assay (See Appendix, sequences No. 4, 5 and 6).

6.2.4.13.3 Easy-to-sequence Mos1 and Mboumar-9 transposons
The sequencing of the transposition products of pEPLL, pEPLR and pEPRR plasmids (primers No. 11 and 12) was obstructed by the presence of 12G/C nucleotides in a row flanking kanamycin resistance gene. The inverted repeats were amplified within 200-300 bp of their flanking DNA from pMos1. During Sanger sequencing the signal drops down significantly after the 12G/C loop and the quality of the sequence is not enough to cover 100-150 following bases to map an insertion of the transposon into a new location (Figure 6-2).

To create Mos1 and Mboumar-9 transposons that would be easy to sequence we have amplified the kanamycin resistance cassette from pUC4K excluding the 12G/C repeat with the primers carrying the sequence of the Mos1 inverted repeat (primers No. 9, 14, 16 and 24) and Mboumar-9 inverted repeats (primers No. 8, 13,15 and 23). The products of amplification of 1.3 kb were cloned into pEP185.2 vector to create six new transposons, three for Mos1 and three for Mboumar-9. The resulted constructs were sequenced to confirm the presence of the correct inverted repeats and absence of mutations in the kanR gene. The created transposons: pEPMosLL, pEPMosLR, pEPMosRR, pERMboLL, pEPMboLR and pERMboRR are flanked with either two left inverted repeats, two right or both left and right inverted repeats (See Appendix, sequences No. 7-12) to investigate the activity of each of the repeats *in vitro* and *in vivo*. 
Figure 6-2 Drop of the sequencing reads after the 12G/C loop.

The 12G/C loop flanking the kanamycin resistance gene from pUC4K plasmid obstructs mapping of new integration sites of Mos1 transposon in vitro and in vivo.
6.2.4.13.4 Transposons for mammalian cells

The neomycin gene (neo<sup>R</sup>) under the regulation of SV40 promoter was flanked with the right inverted repeats of Mos1 and with the left inverted repeats of Mboumar-9 to generate transposons for in vivo integration in the genome of HeLa cells.

To create neo<sup>R</sup> containing transposons we have deleted the SalI site from pBSKS(+) vector with site directed mutagenesis using primers No. 25 and 26 to create pBSKS(+)noSalI. Then subcloned kan<sup>R</sup> gene with the inverted repeats sequences from pEPMosRR (digested with XbaI) and from pEPMboLL (digested with SacI) into pBSKS(+)noSalI by the appropriate sites. The kanamycin cassette was cut out of both transposons with SalI restriction endonuclease. The neomycin cassette of 1.6 kb was amplified from the pEGFP-C1 vector with the primers No. 17 and 18. Resulting products was cloned by SalI sites into the prepared vectors with two right inverted repeats of Mos1 or two left inverted repeats of Mboumar-9 transposons, creating pMosNeo and pMboNeo plasmids. The correct order and absence of mutations of the resulting constructs were confirmed by sequencing (See Appendix, sequences No. 13-15).

6.2.5 Protein work

6.2.5.1 Storage conditions

Mos1 transposase was stored in aliquots as 50% glycerol stocks at -20°C. Mboumar-9 transposase was stored in aliquots in a storage buffer at -80°C after snap freezing in liquid nitrogen. Mbo9 was thawed under cold running water and snap frozen straight after the necessary amount was taken.

6.2.5.2 Gel electrophoresis

Proteins were separated using SDS-PAGE on Bio-Rad Mini-Protean III Electrophoresis Cell. To prepare 12% resolving acrylamide gel we used: 1.6 ml dH<sub>2</sub>O, 2 ml 30% acrylamide, 1.3 ml 1.5 M Tris pH 8.8, 50 µl 10% SDS, 50 µl 10% ammonium persulfate, 2 µl TEMED. To prepare stacking gel we used: 680 µl dH<sub>2</sub>O, 170 µl 30% acrylamide, 130 µl 1 M Tris pH 6.8, 10 µl 10% SDS, 10 µl 10% ammonium persulfate, 1 µl TEMED. The proteins were separated at constant
Improving mariner transposons for transgenesis

Materials and Methods

6.2.5.3 Induction of protein expression

E. coli BL21 (DE3) strain transformed with expression plasmids was grown on LB agar plates with addition of 50 µg/ml of kanamycin (Kan50). Over night cultures in LB Kan50 were inoculated from single colonies. Over night cultures were used to inoculate (1:200) 2xTY media for induction of expression. Cultures were induced at $\text{OD}_{600} = 0.3$ - 0.8 with 0.5 mM IPTG and incubated for 16-24 hours at 18°C for Mboumar-9 protein and 25°C for Mos1 protein expression with 250 rpm agitation.

6.2.6 Biophysical characterisation methods

6.2.6.1 Size exclusion chromatography

Size exclusion chromatography was performed in protein storage buffer at 4°C at 0.5 ml/ml flow rate on Superdex 200 10/300 GL column (GE Healthcare) attached to ÄKTA HPLC system. Sample was loaded in a volume $\leq 500$ µl after centrifugation for 10 min at 13,000 rpm at 4°C (to remove aggregates).

6.2.6.2 Dynamic Light Scattering

Dynamic Light Scattering was measured on Zetasizer Auto Plate Sampler (Malvern Instruments) in 384-well plate. Each protein sample was spun down for 10 min at 13,000 rpm at 4°C prior dispensing into the plate. Triplicates of 50 µl were measured at 4°C three times for each of the samples. We used protein in 0.5-1 mg/ml concentration in a protein storage buffer.

6.2.6.3 Thermal Denaturation Assay

Protein at the concentration of 5 µM in a final volume of 45 µl was mixed with the fluorescent dye Sypro Orange (ex/em 485 nm/575 nm, SIGMA) at a final
Improving *mariner* transposons for transgenesis

The concentration of 5x, spun down at 10 min at 13,000 rpm. If DNA was added (7.5 μM) the protein-DNA mixture was incubated for 30 min on ice before dye was added. The assay was performed in the Bio-Rad IQ5 ICycler, in a sealed 96-well plate, at temperature gradient from 4°C to 95°C with 1°C step and 30 sec hold. Each sample was prepared in triplicates.

6.2.7 *In vitro* plasmid cleavage assay

Donor plasmid of transposon (5.6 kb, 500 ng, 7.24 nM) was incubated with 200 nM of Mos1 transposase or 18 nM Mboumar-9 transposase (optimal protein:DNA ratio varies in different protein stocks and has to be adjusted) in 20 µl final volume for 1 hour at 30°C in a buffer containing: 25 mM HEPES pH 7.5/ 12.5 μg/ml BSA/ 2 mM DTT/ 100 mM NaCl/ 10% glycerol/ 10 mM MnCl₂ or MgCl₂. To stop the reaction 0.4 µl of 500 mM EDTA was added. The whole volume was loaded on 1% (w/v) agarose gel (containing SafeView) and DNA was separated at 70 V for 2 hours at room temperature.

6.2.8 *In vitro* oligonucleotide cleavage assay

The first and the second strand cleavages were performed in 20 µl final volume. Reactions contained: 25 mM HEPES pH 7.5/ 50 mM CH₃COOK/ 10% (v/v) glycerol/ 0.25 mM EDTA/ 1 mM DTT/ 10 mM MgCl₂/ 50 µg/ml BSA/ 20% (v/v) DMSO/ 15 nM DNA/ 100 nM transposase (usually three different concentrations of transposase were used). Incubated for 1 hour at 30°C, stopped reaction by addition of 20 µl of loading buffer (90% (v/v) formamide/ 20 mM EDTA). Samples were denaturated at 95°C for 5 min, transferred to ice, 10 µl was loaded into each well of denaturating polyacrylamide gel.

6.2.9 *In vitro* transposition assay

Transposition was performed essentially as described before (Zhang et al, 2001). Transposon donor plasmid (5.6 kb, 500 ng, 7.24 nM) was incubated with pBSKS+ recipient plasmid (3 kb, 300 ng) and 72.4 nM transposase for 1 hour at 30°C in 20 µl final volume in buffer containing 25 mM HEPES pH 7.5/ 100 mM NaCl/ 10% (v/v) glycerol/ 2 mM DTT/ 200 µg/ml acetylated BSA/ 10 mM of MnCl₂ or MgCl₂. The
buffer and DNA was mixed first, vortexed and spun down, transposase was added last just before incubation, the tube was briefly flicked and spun down. After 1-hour incubation the reaction was stopped by addition of 80 µl of buffer containing 50 mM Tris pH 7.5/500 µg/ml proteinase K/10 mM EDTA/6.25 µg/ml yeast tRNA and incubated for 1 hour at 37°C. DNA was phenol extracted and ethanol precipitated usually over night. The DNA pellet was gently resuspended in 10 µl of pre-warmed to 70°C dH2O. Competent cells were transfected with 10 µl of DNA and plated out on appropriate antibiotics (approximately 4,000 colonies were observed at the optimal conditions).

6.2.10 **In vivo transposition assay**

*In vivo* transposition assay was created and optimised in this work. *In vivo* transposition assay was performed as *in vitro* transposition without addition of stopping buffer, phenol extraction and ethanol precipitation steps. After 1-hour incubation at 30°C the whole reaction was transfected into the cells of interest using the DNA transfection protocol and plated out on appropriate antibiotics (for bacteria cells). Approximately 400 resistant colonies were observed at the optimal conditions.

For mammalian cells no antibiotic was added on the day of transfection. In 24 hours cells were split $\frac{1}{12}$ into 6-well plate and incubated for 48 hour at 37°C. Then the medium was changed and G418 was added to final concentration 1200 µg/ml (concentration was optimised for specific batch of G418 and the cell line). Selection was performed until 1-3 mm in diameter resistant colonies were formed (10-14 days), medium was changed every 2-3 days following DPBS (Gibco®, No. 41966) wash.

The assay also worked for bacterial cells without pre-incubation of transposase with transposon donor DNA. DNA and purified protein were added to a tube of competent cells without pre-incubation in the transposition buffer and the transfection was carried out by the protocol for DNA transfection (approximately 300 colonies were obtained). Formation of resistant colonies was observed even when incubation on ice was omitted from the standard transfection protocol (approximately 50-60 colonies were obtained).
Improving *mariner* transposons for transgenesis

References
Improving mariner transposons for transgenesis


Improving mariner transposons for transgenesis


References 162
Improving *mariner* transposons for transgenesis


Improving mariner transposons for transgenesis


Duncan L, Bouckaert K, Yeh F, Kirk DL (2002) kangaroo, a mobile element from Volvox carteri, is a member of a newly recognized third class of retrotransposons. *Genetics* 162: 1617-1630


Elick TA, Bauser CA, Fraser MJ (1996) Excision of the piggyBac transposable element in vitro is a precise event that is enhanced by the expression of its encoded transposase. *Genetica* 98: 33-41

Improving mariner transposons for transgenesis


Fraser MJ, Ciszczon T, Elick T, Bauser C (1996) Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. *Insect molecular biology* 5: 141-151


References
Improving *mariner* transposons for transgenesis


Improving mariner transposons for transgenesis


References


Improving *mariner* transposons for transgenesis


Kirchner J, Sandmeyer SB (1996) Ty3 integrase mutants defective in reverse transcription or 3'-end processing of extrachromosomal Ty3 DNA. *Journal of virology* **70:** 4737-4747


Le Breton Y, Mohapatra NP, Haldenwang WG (2006) In vivo random mutagenesis of *Bacillus subtilis* by use of TnYLB-1, a mariner-based transposon. *Applied and environmental microbiology* **72:** 327-333

References
Improving mariner transposons for transgenesis


Li R, Zhuang Y, Han M, Xu T, Wu X (2013b) piggyBac as a high-capacity transgenesis and gene-therapy vector in human cells and mice. *Disease models & mechanisms*


References
Improving mariner transposons for transgenesis


Lyras D, Rood JI (2000) Transposition of Tn4451 and Tn4453 involves a circular intermediate that forms a promoter for the large resolvase, TnpX. *Molecular microbiology* **38**: 588-601


References
Improving mariner transposons for transgenesis


References

174
Improving mariner transposons for transgenesis


References
Improving *mariner* transposons for transgenesis


References
Improving mariner transposons for transgenesis


References
Improving mariner transposons for transgenesis


Improving *mariner* transposons for transgenesis


References
Improving mariner transposons for transgenesis


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

Wu SCY, Meir YJJ, Coates CJ, Handler AM, Pelczar P, Moisyadi S, Kaminski JM (2006b) piggyBac is a flexible and highly active transposon as compared to Sleeping Beauty, Tol2 and Mos1 in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 15008-15013


Yin J, Bi X (2013) Generation and characterization of virus-free reprogrammed melanoma cells by the piggyBac transposon. *Journal of cancer research and clinical oncology*


References
Improving mariner transposons for transgenesis


Appendix
Sequences of the plasmids and inserts created in this study:

>1. pKanLL, SacI, IRL, SalI

cacctaaattgaacgcttaaatattttgttaaaattgcgttaaatatttttcgcgttaaatatttttttaattcgcgttaaatcagctcattttttaaccaataggccgaaatc
gcggcaaaataccttataacatcaaaagagagacagagaattggttaatgttggttgctcagtttggccaaagagaatggcagacgctgggctgcaagtgtagcg
cctagctcgcgcaacggttaagcgtttaaatattttgttaaaattcgcgttaaatcagctcattttttaaccaataggccgaaatc
gcggcaaaatcccttataaatcaaaagaatagaccgagatagggttgagtgttgttccagtttggaacaagagtccactattaa
tttattttgtttatttttttgtttatgtatttttttggttcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccccgatttagagcttgacggggaa
tagccggcgaacgtggcgagaaaggaagggaagacgccagatttttttggttcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccccgatttagagct
tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
Improving mariner transposons for transgenesis

Appendix

2. pKanLR, SacI, IRL, IR, BamHI

cacttaaatgtgaacgttaaatattttgtataaatcagcctaatattttttcataaatcagcctaattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
Improving *mariner* transposons for transgenesis

>3. pKanRR, BamHI, IRR, SalI
Improving mariner transposons for transgenesis

Appendix 186
Improving mariner transposons for transgenesis

Appendix

>4. pEPLL, Sacl, IRI, SalI, this 2392 bp fragment was inserted into pEP185.2 plasmid (4275 bp)
Improving mariner transposons for transgenesis

Appendix

>5. pEPLR, \textbf{SacI}, \textbf{IRL}, \textbf{SalI}, \textbf{IRR}, \textbf{BamHI}, this 2330 bp fragment was inserted into pEP185.2 plasmid (4275 bp)

gagctccccagacgagtgaagaagaaaaacgaatgtatgctgtgggttaaaaacccacctttaaatgtgtatgcagatgtgtaaacaagacagacttgagttcgaatagttcgcaaaaccgtgaatattatagtacagtttgtatgtatgctaatgcgttttttggtagataaatgtcggctgcgcagtttcccttttccttaggaatgaccatgcgcacttccactttcgtgggataaaaatggtaaatccggctttttctttaaagctaccaaatgatgtagataggcatccacagtacgggctccagtcggtgaatatagaaactaatcaggtgtacaagtagtatgaaatggtcttttttaaatcaaaaaacacgttaaattttggaayaagaaaaagaacattttctgcattccagagtcacctttttcccatctctcggaataaacgtgttttttgatttaaaaaaacgacatttcatacttgtagttttctatattcaccgactggagccctgtactgtg

ggatcc

>6. pEPRR, \textbf{BamHI}, \textbf{IRR}, \textbf{SalI} this 2268 bp fragment was inserted into pEP185.2 plasmid (4275 bp)

ggatcc

ggatcc
Improving mariner transposons for transgenesis

Appendix

189
Improving mariner transposons for transgenesis

Appendix 190

tcaaccaacgctattatcattctctgtgatgacgccagacgaattacagcagacctgcacgcaaaaatattctctacggtaattcattccttgataatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcaggatattcttctaatacct~
gggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcggaaagggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattggccacgctacctttgccatgtttcagaaacacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacacccctttgattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacaacaacctctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttccgttgaatatggctcataacac
Improving mariner transposons for transgenesis

>11. pEPMboLR SacI, IRL, SalI, IRR, XbaI. This 1304 bp fragment was inserted into pEP185.2 plasmid (4275 bp)

gagctctacagggtgtctgcgttaaccttcctctgtgttttgctgac
cgctgaggtctccctgtgaagaaggttgtgctgactc
atacagggcctgaatgcgcccatccgacagccggaaaggtgaggtggtggtctgacagtgtgaatgggtggtgattttgaacttttgctttgccacgggaacggtctgactgatccttcaactcagaa
agtttggtgattttgaacttttgctttgccacggaacggtctgcgttgtcgggaagatgcgtgatctgatccttcaactcagaa
aagttcgatttattcaacaaagcccaccgatgcacatccatctctgtgacacagtgaatccg
gtgagaatggcaaaagcttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgactcaaccaaaacgcttcatctgtgaatcaggatattcttctaatacct
ggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcggagaggcataaatcggcttcccatacaatcgatagattgtcgcacctgattgccgacattatcgcgagcccatttata
cccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttcccgttgaatatggctcataacaccc
ttgtatatgattatatagcaagcaagggtatttcatttcgactgatgatatatttttctcttgcaatgtaacatcagagattttgagacacaacgtggcctgtgtgacdaaaaccgggaaaagaattacgacactcctgg
tctaga

>12. pEPMboRR XbaI, IRR, SalI. This 1304 bp fragment was inserted into pEP185.2 plasmid (4275 bp)

tctagatacaggagtctgcgttaaccttcctctgtgttttgctgac
cgctgaggtctccctgtgaagaaggttgtgctgactc
atacagggcctgaatgcgcccatccgacagccggaaaggtgaggtggtggtctgacagtgtgaatgggtggtgattttgaacttttgctttgccacgggaacggtctgactgatccttcaactcagaa
agtttggtgattttgaacttttgctttgccacggaacggtctgcgttgtcgggaagatgcgtgatctgatccttcaactcagaa
aagttcgatttattcaacaaagcccaccgatgcacatccatctctgtgacacagtgaatccg
gtgagaatggcaaaagcttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgactcaaccaaaacgcttcatctgtgaatcaggatattcttctaatacct
ggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcggagaggcataaatcggcttcccatacaatcgatagattgtcgcacctgattgccgacattatcgcgagcccatttata
cccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttcccgttgaatatggctcataacaccc
ttgtatatgattatatagcaagcaagggtatttcatttcgactgatgatatatttttctcttgcaatgtaacatcagagattttgagacacaacgtggcctgtgtgacdaaaaccgggaaaagaattacgacactcctgg
tctaga

>13. pBSKS(+)noSalI
cacctaatgtaaagcgccttaatatatttttgataaatctgcctctctgtgttttgctgac
cgctgaggtctccctgtgaagaaggttgtgctgactc
atacagggcctgaatgcgcccatccgacagccggaaaggtgaggtggtggtctgacagtgtgaatgggtggtgattttgaacttttgctttgccacgggaacggtctgactgatccttcaactcagaa
agtttggtgattttgaacttttgctttgccacggaacggtctgcgttgtcgggaagatgcgtgatctgatccttcaactcagaa
aagttcgatttattcaacaaagcccaccgatgcacatccatctctgtgacacagtgaatccg
gtgagaatggcaaaagcttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgactcaaccaaaacgcttcatctgtgaatcaggatattcttctaatacct
ggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcggagaggcataaatcggcttcccatacaatcgatagattgtcgcacctgattgccgacattatcgcgagcccatttata
cccatataaatcagcatccatgttggaatttaatcgcggcctcgagcaagacgtttcccgttgaatatggctcataacaccc
ttgtatatgattatatagcaagcaagggtatttcatttcgactgatgatatatttttctcttgcaatgtaacatcagagattttgagacacaacgtggcctgtgtgacdaaaaccgggaaaagaattacgacactcctgg
tctaga

Appendix
Improving mariner transposons for transgenesis

Appendix

192
Improving *mariner* transposons for transgenesis

Appendix

>15. pMboNeo
cacttaatttgaaagctctaattttgtaatttttgaattttttgcatttaatgatttcgacggctggagttgctctcaatcagctcagacgtcagccggtgcagtttcgtccatatctgtggtggtctgctctggtgtctgttcgttcttccctggttctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctgtggtggtctgctcctggtggtctg
Improving mariner transposons for transgenesis

gattaagttgtaacgccagggtttttccagtcagctgaagtttaaaacagcggctagttgtattgtcaaattttatgatatgcagatcagtcatcaattttcatcttataatttatcattatatgatggtccccccaccacccagcagcagtcatcaattttcatcttataattttagttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
Improving *mariner* transposons for transgenesis

gtgctcatcattggaaacgttcctggggcgaacttaaggatctattaccgtgtgtgagatctgccatgtgtacccca
cctcgtgccaccaacttgcagctacccattttactttaccacccgtttctggcttcaagtaaaaaacaggaaggeaaatgcgc
aaaaaaaggaataagggcagacggaaatgttaactctactactactcttcctccttttcaatattatggggcttttacaggtttat
tgtctcatcgccgacatattgtattagactaataaaataacaaataggggtccgcgcacattttgccgaagaafgc