Studies on the mechanism of mos1 transposition

Florin Dale

A thesis presented for the degree of PhD

University of Edinburgh

2005

I dedicate this thesis to my parents.

Declaration:

This thesis and all the work herein was composed by myself, unless otherwise stated.

Florin Dale

December, 2004

Acknowledgements

I would like to thank Professor David Finnegan for allowing me to carry out this research in his laboratory and for all his help, advice and encouragement.

Angela, this thesis would not have been possible without you. Thank you for your endless patience, kindness and help, and for making me laugh when things weren't going well. You made everything worthwhile. I would also like to thank Julia, Eve, Cheryl, Ivan, Theo and Natasha O'Hagen who have helped me on this journey.

Natasha and Camelia, thank you for too many things to be mentioned. John E, John B, John C, Phil, Elise, Emily, Katie, Julie, Anna, Laura, Jenna, Rabaab, Ewa, Amanda, and all the other people that lived on the 6th floor of Darwin Building, Jen, Gra and all the "Wellcome lot", thank you for being my companions on this ship.

I would like to express my gratitude and admiration to Noreen and Ken Murray. Your kindness and dedication will inspire me for all my life.

Finally I would like to thank my parents Nicolaie and Floare, my brother Marcel, Vic, Andrea and the future member of the family for all their unconditional love and support.

Abbreviations

A	Adenosine
ADE	adenine
Amp	ampicillin
ATP	adenosine 5'-triphosphate
bp	base pair
BSA	bovine serum albumin
C	cytosine
°C	degree Celsius
CJ	coding joint
cDNA	complementary DNA
Ci	curie
CIP	calf intestinal phosphatase
cm	centimetre
d	deoxy
dd	dideoxy
dH_2O	distilled H_2O
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	doexyribonucleoside triphosphate
DSB	double strand break
DTT	dithiothreitol
EDTA	ethylene diaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
g	gram
G	guanosine
GST	glutathione S-transferase
HEPES	N-[2-Hydroxyethyl]piperazine-N'-[2-ethane-sulphonic acid]
HMGB1	high mobility group protein B1
HR	homologous recombination
HRP	horseradish peroxidase
HTH	helix-turn-helix
IPTG	isopropyl-β-D-thiogalactoside
ITR	inverted terminal repeat
kb	kilobase
kD	kiloDalton
l	litre
L-Broth	Luria broth
LINE	long interspersed nuclear element
LiT	Lithium-Tris buffer
LTR	long terminal repeat
M	molar
MBP	maltose binding protein
mg	milligram
ml	millilitre
mM	millimolar
mol	mole

MOPS	3-[n-morpholino]propanesulphonic acid
mqH_2O	milli Q water
mRNA	messenger ribonucleic acid
MW	molecular weight
ng	nanogram
nt	nucleotides
NHEJ	non-homologous end joining
NLS	nuclear localisation signal
NP40	Nonidet-P 40
OD	optical density
ORF	open reading frame
OPI	overproduction inhibition
oz	ounce
p	plasmid
p	pico
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
POD	peroxidase
PVDF	polyvinylidene fluoride
Rev/min	revolutions per minute
RNA	ribonucleic acid
RNase	ribonuclease
rpm	rotations per minute
RSS	recombination signal sequence
RT	reverse transcriptase
SDS	sodium dodecyl sulphate
SINE	short interspersed nuclear element
ST	strand transfer
SJ	signal joint
T	thymine
Tris	tris (hydroxymethyl) aminomethane
TRP	triptophan
u	unit
µg	microgram
µl	microlitre
URA	uracil
UV	ultraviolet
V	volt
v	volume
v/v	volume per volume
w^{pch}	*white peach*
w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indol-βD-galactopyranoside

Definitions are not given for standard chemical symbols and amino acids.

Abstract

Transposable elements are mobile pieces of DNA that move inside the genome of their host and, can invade new genomes through horizontal transfer. They account for a major fraction of DNA in the genome of every organism including humans. Due to their mobility, transposable elements are major evolutionary forces that can shape the genomes of their hosts. The *Mos1* element from *Drosophila mauritiana* is a member of the *mariner* family of transposable elements being one of the very few active *mariner*s found to date. *Mariner* elements transpose through a cut-and-paste mechanism during which the element is excised from its original location and inserts into a new site. The same transposase enzyme catalyses both the excision and the integration steps. I have purified the *Mos1* transposase in a soluble form as a fusion to the Maltose Binding Protein of *E. coli* (MBP-*Mos1*). This fusion protein was used to characterise the steps leading to transposon excision with emphasis on the formation of a higher-order protein-DNA complex termed Paired End Complex. The results from the initial characterisation of MBP-*Mos1* transposase are in good agreement with those obtained using the transposase purified through a denaturation-renaturation process. Next, the stoichiometry of *Mos1* Paired End Complex was studied using a mixture of MBP-*Mos1* fusion transposase and *Mos1* transposase. The results suggest a dimeric structure of *Mos1* Paired End Complex. Gel filtration experiments showed that *Mos1* transposase exists in solution as a mixture of monomers and dimers. The characterization of these forms showed that the monomer is the active form of *Mos1* transposase for DNA binding. The possible role of dimer formation in the regulation of *Mos1* transposition is also discussed. Transposition is a source of DNA double strand breaks, which are lethal to the host organism if they remain unrepaired. An assay to study the repair of these double-strand breaks is presented that takes advantage of *Saccharomyces cerevisiae*'s ease of manipulation and wealth of knowledge regarding DNA repair processes.

Table of contents

Chapter 1. Introduction

1.1 Transposable elements – general introduction

Transposable elements are ubiquitous components of genomes that have the ability to move into new locations inside their host genome, and sometimes to bypass the barriers between evolutionarily distant organisms through horizontal transfer (Craig, 2002).

Surprisingly, the idea that genomes contain segments of mobile DNA was conceptualised before the discovery of the structure of DNA (Craig, 2002). The first suggestion that genetic material can actually be mobile came from work by Barbara McClintock in the late 1940s when she discovered that genetic determinants in maize, which she named "controlling elements", could generate chromosomal breaks. This led to the discovery of the *Ac/Ds* transposons in maize (McClintock, 1950).

Subsequently, transposable elements have been discovered in virtually every organism studied to date. In bacteria, they can mobilize antibiotic resistance determinants between plasmids and bacterial chromosomes, hence between different bacterial cells (Grindley, 1983). In humans, active copies and remnants of transposable elements represent an astonishing 45% of total DNA (Human Genome Consortium, 2001).

1.2 Classification of transposable elements

One of the most widely used means of classification takes into account the intermediates involved in the transposition reaction. According to this criterion, transposable elements belong to one of the two classes, Class I or Class II (Finnegan, 1989). Class I transposable elements transpose via an RNA intermediate that is converted into a cDNA copy through a reverse transcription reaction. This cDNA copy is then inserted at a new location inside the host genome. Class II transposable elements, also termed transposons or DNA transposons, transpose via a cut-and-paste mechanism, in which the transposon is excised from its original location, and inserted into a new sequence known as " the target site" inside the host genome (**Figure 1.1**).

Class I transposable elements can be further divided into Long Terminal Repeats elements, or LTR elements, and non-LTR elements. LTR elements resemble retroviruses in having long terminal repeats and open reading frames analogous to the *gag*, *pol* and in some cases *env* genes of retroviruses (**Figure 1.1a**). Examples of LTR elements include the *Ty1 and Ty3* elements from *Saccharomyces cerevisiae* (Sandmeyer *et al.*, 2002; Voytas and Boeke, 2002), the murine *intracisternal A particle* (*IAP*) (Dewannieux *et al.*, 2004), the *gypsy* and the *copia* elements from *Drosophila melanogaster* (Boeke and Corces, 1989).

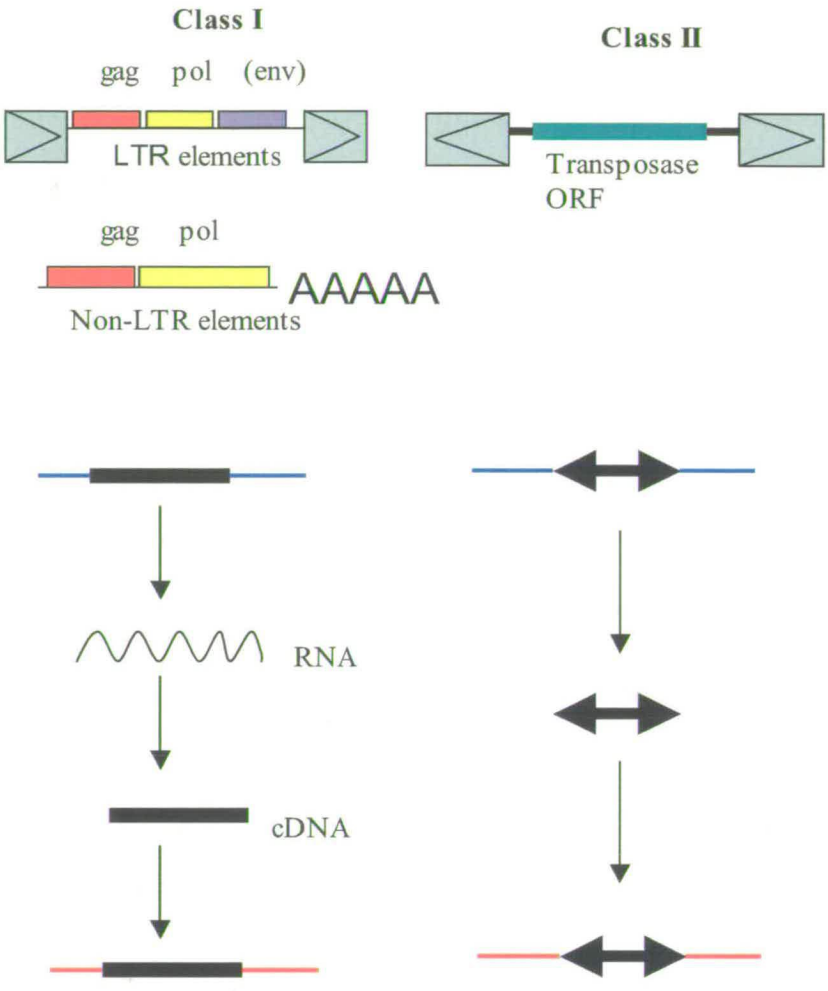

Figure 1a.
Transposition through an RNA intermediate

Figure 1b.
Cut-and-paste transposition

Figure 1.1 Class I and Class II transposable elements.

Schematic representation of Class I (Fig.1a.) and Class II (Fig.1b.) transposable elements and their mechanisms of transposition.

Non-LTR elements have no terminal repeats, contain two ORFs, and a poly(A) stretch at the 3' end (**Figure 1.1a**). These elements were initially discovered in mammalian genomes and subsequently in nearly every eukaryote organism. This group includes the *I* and *jockey* elements from *Drosophila melanogaster* (Bucheton *et al.*, 2002; Priimagi *et al.*, 1988), the *Long Interspersed Nuclear Elements* (*LINE*s) from mammals (Moran and Gilbert, 2002), and *R1* and *R2* elements found in most arthropods studied to date (Burke *et al.*, 1993; Burke *et al.*, 1998; Eickbush, 2002; Jakubczak *et al.*, 1991).

A class II transposable element is cut from the donor DNA and inserted in the target DNA without the need for an RNA intermediate. This is why elements from this group are sometimes called DNA transposons, or simply transposons. Elements belonging to class II have a relatively simple structure, with inverted terminal repeats of various lengths flanking an ORF encoding the transposase enzyme. Examples of class II transposable elements are, from prokaryotes the bacteriophage *Mu*, Insertion Sequences (IS), and bacterial elements Tn*5* and Tn*10,* and eukaryotic elements belonging to the *Tc1* and *mariner* families (**Figure 1.1b**).

1.3 The class II transposition systems

1.3.1 Introduction

This chapter will briefly describe the mechanisms of transposition employed by some of the best-studied elements belonging to the second class. They transpose through a mechanism in which only DNA intermediates are involved. Two basic ways of DNA transposition have been described, a "replicative" mechanism and a non-replicative mechanism (**Figure 1.2**).

Elements that employ a replicative mechanism of transposition are the Tn*3*-like elements and the bacteriophage *Mu*. Some insertion sequences seem to be able to transpose by both replicative and non-replicative transposition, such as IS*1*.

The Tn*3*-like elements transpose by a replicative pathway forming an intermediate called a cointegrate. Apart from the transposase, they encode a serine or tyrosine recombinase that resolves the cointegrate into two components, the target DNA with a transposon insertion and the regenerated donor DNA. The *Mu* bacteriophage also transposes replicatively through an intermediate structure named a Shapiro intermediate, thought to be the same type of intermediate as used by the Tn*3* transposon (**Figure 1.2a**).

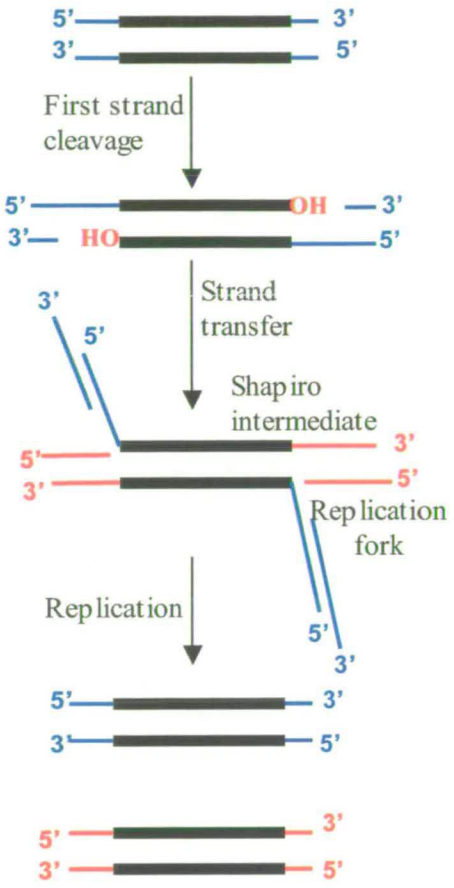

Figure 1.2a.
Replicative transposition

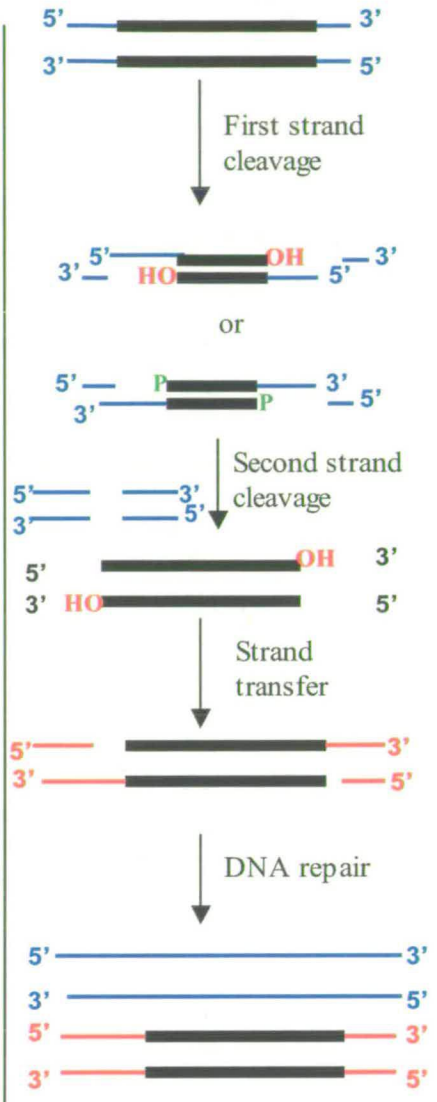

Figure 1.2b.
Cut-and-paste transposition

Figure 1.2. DNA transposition. Comparison between replicative transposition (Fig.1.2a.) and cut-and-paste transposition (Fig.1.2b.).

Note that in replicative transposition only one DNA strand is nicked, resulting in free OH groups at the two transposon ends.These are then used in the strand transfer step to attack the target DNA resulting in the formation of replication fork-like structures. In the cut-and-paste transposition, the transposon is completely excised from the donor DNA by cleavage of both DNA strands. Again, the resulting OH groups are used in the strand transfer step, inserting the transposon into the target DNA. The resulting gaps in the DNA are repaired by host enzymes. Note that in the cut-and-paste transposition, the donor DNA loses the transposon, as opposed to replicative transposition, where the donor DNA retains the original copy of the element. The free hidroxyl groups in the transposon DNA resulting after nicking are represented in red (OH). The corresponding phosphate groups are represented in green (P). The donor DNA is represented as blue lines. The target DNA as red lines. The transposon DNA is represented as black rectangles. Both DNA strands are represented, with the 5' and 3' ends of each strand marked.

The non-replicative transposition is also called "cut-and-paste" transposition and is used by members of the IS*630* / *Tc1* / *mariner* superfamily of transposable elements. This mechanism requires the complete excision of the transposon from the donor DNA and subsequent integration into the target DNA. No replicative intermediates are involved. During excision, the transposon is excised from the donor DNA by a pair of double-strand breaks (one for each transposon end), generating a free hydroxyl group at the 3' end of each strand of transposon DNA. These 3'-OH groups are used in the integration stage to attack the target DNA and cleave it in a staggered fashion. Through a trans-esterification reaction the transposon is inserted at the new location (**Figure 1.2b**).

In the next sections, the best-characterised transposition systems are discussed. As stated in section 1.6 one scope of this thesis is to address questions regarding a crucial intermediate in transposition of the *Mos1* transposon, namely the synaptic complex or Paired End Complex (PEC). In the next sections of the Introduction (sections 1.3.2 – 1.3.6) I will try to focus on the best characterized synaptic complexes in transposition systems available to date. Tn*3*-like elements and the *Mu* bacteriophage represent the replicative mode of transposition, with synaptic complexes characterised in depth. Their organization, roles in the transposition reaction, and models for assembly and function are the subject of the next two sections (section 1.3.2 for Tn*3*-like elements, and section 1.3.3 for bacteriophage *Mu*). Bacterial elements with a non-replicative mechanism of transposition are discussed in sections 1.3.4 (Tn*10*) and 1.3.5 (Tn*5*). Both Tn*10* and

Tn*5* are models for the mechanism in which one acive site of the transposase performs all the DNA cleavages at one transposon end through a hairpin intermediate, with the Tn*5* synaptic complex being the first for which a crystal structure is available. One alternative of the hairpin mechanism employed by Tn*10* and Tn*5* is described in section 1.3.6 discussing the V(D)J recombination reaction. In section 1.4 the *Tc1/mariner* family of transposable elements is introduced, with *Tc1* and *Tc3* elements discussed in section 1.4.1.1 and the *Sleeping Beauty* (*SB*) element in section 1.4.1.2. A brief introduction to the *mariner* transposons is given in sections 1.5.1 to 1.5.3. Their molecular mechanism of transposition is described in section 1.5.4, with some questions regarding this mechanism outlined at the end of section 1.5.4.

1.3.2 Tn*3*-like elements: transposition via a DNA co-integrate

Tn*1* and Tn*3* were among the first antibiotic-resistance carrying transposons identified. They form a large family of transposons called Tn*3*-like elements. The typical transposon from this class is Tn*3*, which is flanked by 38 bp ITRs, and contains transposase (*tnpA*), resolvase (*tnpR*) genes and a *bla* ORF encoding resistance to ampicillin. A region of approximately 220 bp separates the *tnpA* and *tnpR* ORFs, representing the site of resolvase action (**Figure 1.3**). A member of the same group is the $\gamma\delta$ transposon, component of the F fertility plasmid, whose resolvase functions have been studied in great detail (Grindley, 2002).

In vivo, the final product of Tn*3* transposition is represented by a simple insertion. In contrast to the cut-and-paste transposons such as Tn*10* and Tn*5*, Tn*3* does not excise from the donor replicon. *In vitro* studies have shown that transposition reactions using an internally deleted transposon result in accumulation of co-integrates between the donor and the target replicons (Gill *et al.*, 1978). These intermediates are formed through a mechanism proposed by Shapiro *et. al* (Shapiro *et al.*, 1979). In a first series of steps, the transposase binds to the transposon ends, and pairs them forming a synaptic complex. Nicks are introduced by the transposase at both 3' ends of the transposon, resulting in free hydroxyl groups. The complex captures a target DNA and, the 3' OH groups from the transposon ends attack the target phosphate backbone, linking them to the target DNA. This generates a three-way junction at each transposon end, representing a replication fork, at which the bacterial replication machinery initiates the replication of the transposon DNA. This leads to cointegrate formation.

The subsequent steps that resolve the cointegrate structure are performed by a different enzyme - the resolvase. For Tn*3* and $\gamma\delta$, this is a member of a larger family of recombinases, named serine-recombinases because covalently linked intermediates are formed between serine residues in the protein and the 5' ends of the recombining DNA strands (Reed, 1981a; Reed and Grindley, 1981b). Unlike the transposase, the resolvase does not act at the transposon ITRs, but at sequences located inside the transposon - recombination sites or *res* sites. The structure of one such site is presented in **Figure 1.3b**. A typical *res* site contains three binding sites, named I, II and III, with a dimer of

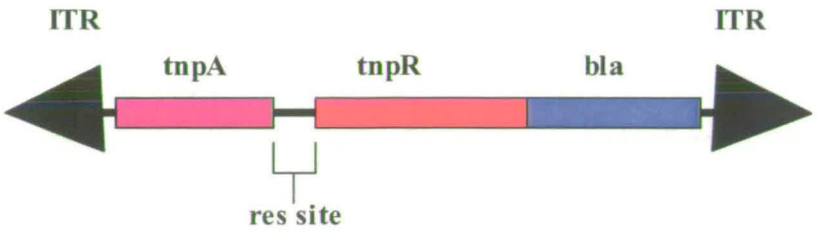

Figure 1.3a. Transposon Tn*3*

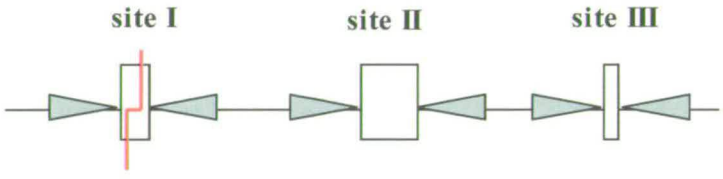

Figure 1.3b. Structure of a *res* site.
Open boxes represent resolvase binding sites. The strand exchange position is indicated by a staggered line (adapted from Burke *et al.*, 2004).

resolvase binding each site. Site I contains the site of DNA breakage and strand exchange. The recombination reaction takes place inside a synaptic complex containing two *res* sites held together by recombinase molecules. The $\gamma\delta$ resolvase synaptic complex contains twelve molecules of resolvase bridging a pair of *res* sites in a very complex structure (Sarkis *et al.*, 2001; Stark *et al.*, 1992). A dimer of resolvase binds to each of the six binding sites (three per *res* site), although only the site I-bound resolvase molecules are involved in the chemical steps of the reaction. The remaining sites, or accessory sites are still crucial for synaptic complex assembly.

Inside the synaptic complex, a pair of activated serine (Ser) residues attacks the phosphate backbone on the two strands of each DNA duplex. This generates free 3' OH groups and phosphodiester bonds between Ser residues and the 5' ends of the two DNA duplexes (that is, four covalent bonds per recombination complex). The DNA ends are switched from the original configuration into the recombinant configuration, thus allowing the free 3' OH groups to attack the serine-DNA phosphodiesters in a second set of transesterification reactions, re-ligating the DNA backbone to produce the recombinant sites (Stark *et al.*, 1992). Thus, the two transposon-containing replicons are separated, and the transposition process completed.

The recombinase activity requires the presence of two *res* sites on the same DNA molecule, with the *res* sites in a head-to-tail orientation, and the DNA negatively supercoiled. This ensures the assembly of a synaptic complex with a correct structure,

which is required to activate the recombinase. Two models have been proposed for the assembly of a correct synaptic complex or synaptosome, the slithering model and the topological filter model. The first model proposes that the two *res* sites on a supercoiled DNA molecule find each other by a movement of DNA segments in plectonemically interwound DNA (the model was proposed by Benjamin et. al. and cited by Grindley) (Benjamin and Cozzarelli, 1986; Grindley, 2002). This model, however, does not explain the observation that recombination can also occur with linear or relaxed plasmid DNA substrates, albeit with lower efficiency (Boocock *et al.*, 1987; Stark *et al.*, 1989). These observations showed that slithering is not required for synapsis and led to the proposal of a second model named the topological filter (Boocock *et al.*, 1986; Boocock *et al.*, 1987; Stark and Boocock, 1995; Stark *et al.*, 1989). The topological filter model predicts that two *res* sites are brought together by random collision, or slithering depending on the substrate, resulting in the formation of a mixture of synaptic complexes with different topologies. The correct synaptic complex requires three negative nodes to be trapped by the interwound *res* sites. This first necessitates a synapsis of the accessory sites II and III, and only subsequently of the two sites I (Kilbride *et al.*, 1999). The essence of the topological filter model predicts that if more (or less) than three negative nodes are trapped by the two synapsed *res* sites, the energy required to bring the synaptic complex to the productive topology (that is to three trapped nodes) is prohibitive (Grindley, 2002; Stark and Boocock, 1995; Stark *et al.*, 1994).

Crystal structures of the $\gamma\delta$ resolvase, alone and complexed with DNA, are available (Rice and Steitz, 1994a; Rice and Steitz, 1994b; Sanderson *et al.*, 1990; Yang and Steitz, 1995b). The structures have revealed an assembly of a "tetramer of dimers" in which essential contacts between resolvase dimers are made (Rice and Steitz, 1994a).

Rice and Steitz have proposed a first model for a complete synaptosome of $\gamma\delta$ resolvase (Rice and Steitz, 1994a). In this model, the crystallographic tetramer of dimers is responsible for the inter-wrapping of the accessory sites II and III. The catalytic site I was connected to the core tetramer of dimers by one interaction between the dimer bound to site I, and the dimer bound to site III. The two site I DNA segments would be in close proximity, passing between the catalytic domains of site I dimers, whereas the rest of the res site DNA would be wrapped around the outside of the tetramer of dimers core. Another model was proposed with an additional interaction between resolvase dimers bound to sites I and III (Sarkis *et al.*, 2001). In this model, all resolvase dimers are closely paired, each monomer making the same set of protein-protein interactions. The site I DNA in this case would be located on the outside of the catalytic resolvase dimers (Grindley, 2002).

One of the most interesting features of the resolvase synaptic complex lies in its ability to catalyse the strand exchange reaction. In this reaction, four DNA ends must be switched from the parental to the recombined position. Three mechanisms to explain how this might occur have been proposed. The Rice-Steitz model proposed that the

strand exchange reaction would take place in a rather rigid environment of resolvase subunits, without any major rearrangements of the complex. This is a consequence of the "DNA-in" position of site I DNA in the Rice-Steitz model(Rice and Steitz, 1994a). A second model, by Stark *et al.*, proposed that after cleavage, the four DNA strands are exchanged through a 180° rotation of a pair of resolvase subunits. This model predicts a "DNA-out" position for the site I DNA, although the model can accommodate the "DNA-in" configuration also. The later model has difficulties explaining how the synaptic complex is held together during such a dramatic conformational change (Grindley, 2002). This difficulty is explained in a third model proposed by Boocock as cited by Grindley (Grindley, 2002), and called the "domain swapping" model. The same DNA movement is predicted by this model, which requires the "DNA-out" configuration. The difference is that only the N-terminal parts of the resolvase subunits rotate by 180°. The rest of the resolvase subunits maintain their original position, and the stability of the synaptic complex (Burke *et al.*, 2004).

The Tn*3* synaptic complex exemplifies the high levels of complexity required from these complexes to coordinate the DNA cleavage and recombination steps. The next section describes another replicative system, namely that of bacteriophage *Mu*, in which a carefully orchestrated assembly of *Mu*'s synaptic complex ensures proper recombination.

1.3.3 *Mu* transposition reaction

Another transposition system that uses the replicative mechanism is represented by the bacteriophage *Mu*. *Mu* is one of the most studied transposons. The synaptic complex formation by the *Mu* transposase with *Mu* ends is well documented and reveals important roles for the synaptic complex in regulation of *Mu* transposition.

Mu is a temperate phage. As any temperate phage, upon infection of the host bacterial cell it can enter either the lytic or the lysogenic cycle (**Figure 1.4a**). In the first case, all the phage functions are expressed, the DNA is replicated, packaged into phage particles, and the host cell is lysed releasing between 50 and 100 phage particles. In the second instance, a repressor is synthesized that blocks the expression of most viral genes, and the phage enters the stable lysogenic cycle. The repressor also blocks the expression of a superinfecting *Mu*, so the cell becomes immune to infection by another *Mu* phage. A striking difference between *Mu* and other phages is that *Mu* integrates in the bacterial genome whether it enters the lytic or the lysogenic cycle. *Mu* integration is the first step after cell infection and generates 5 bp duplications of the target sequence. Once integrated, *Mu* can transpose by a replicative mechanism into new locations in the bacterial chromosome. Upon integration and excision mutations are generated, some of

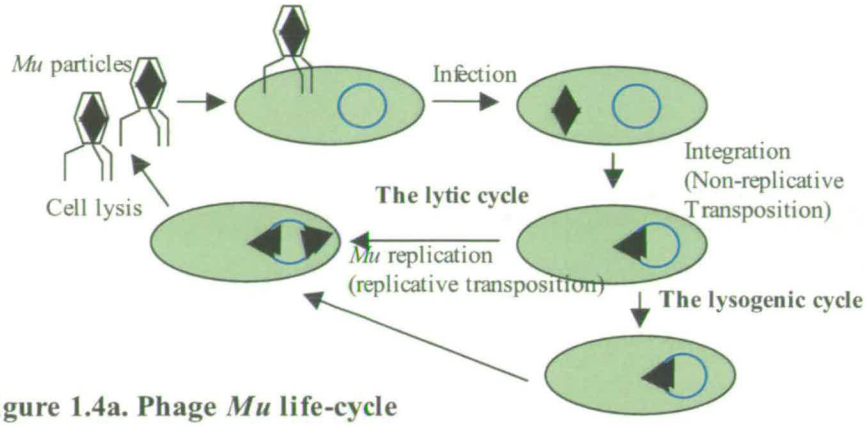

Figure 1.4a. Phage *Mu* life-cycle

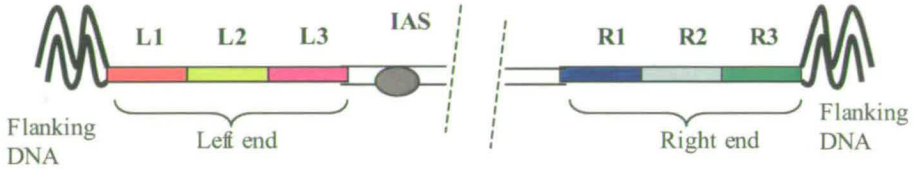

Figure 1.4b. Structure of *Mu* sites involved in synaptic complex formation

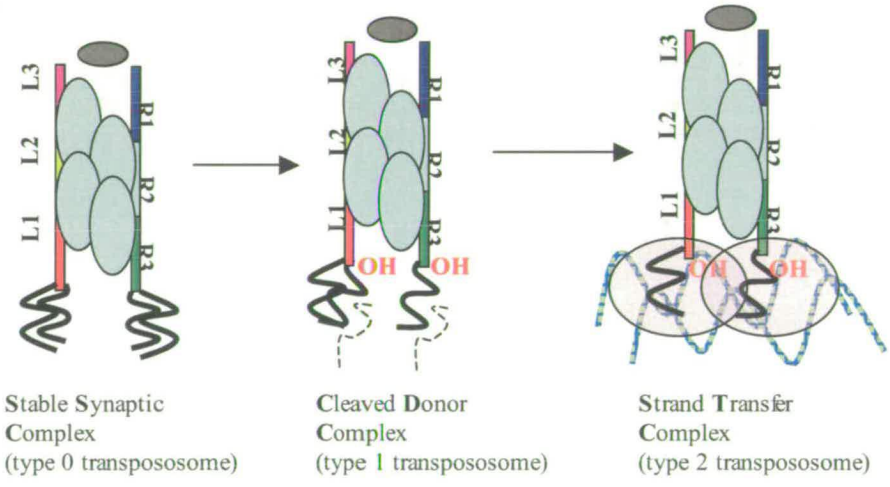

Figure 1.4c. Synaptic complexes during *Mu* transposition.

Figure 1.4. Bacteriophage *Mu*.

Figure 1.4a. Phage *Mu* life-cycle. After infection, *Mu* can undergo two cycles, the lytic cycle or the lysogenic cycle. In both situations, the phage integrates into the *E. coli* genome (blue circles).
Figure 1.4b. Structure of *Mu* sites involved in synaptic complex formation. Only the ends of *Mu* are represented. The Internal Activating Site (IAS) is represented as a grey oval.
Figure 1.4c. Synaptic complexes during *Mu* transposition. Note that the position of the MuA subunits (pale blue ovals) and the DNA-protein interactions inside the complexes are not represented in the figure. The free OH groups formed at the 3' ends of the transposon are represented in red. The cleaved DNA strand is represented as a dashed line.The MuB protein is represented as pale ovals on the target DNA.

which may manifest phenotypycally, hence the name *Mu* (mutator) (Taylor, 1963). *Mu* was the first prokaryotic transposon to be described (Taylor, 1963) and was also the first transposon to be genetically engineered (Casadaban and Cohen, 1979). The separation of *Mu* phage functions from those of transposition came with the construction of deleted variants of *Mu* called mini-*Mu* (Maynard-Smith *et al.*, 1980; Schumann *et al.*, 1979). These contain large internal deletions but retain intact *Mu* termini. All mini-*Mu*s are inactive but can be mobilised with the aid of a helper element. Two *Mu*-encoded proteins are required for transposition: MuA and MuB. *Mu* transposition is catalysed by the MuA transposase (Bukhari, 1975). MuA is a protein of 75 kDa, and contains 663 amino acids (Harshey *et al.*, 1985). Partial proteolysis experiments have identified domains in the transposase and assigned functions to some of them (Betermier *et al.*, 1987; Nakayama *et al.*, 1987). Two protease hypersensitive sites at residues 247 and 574 divide the protein into three domains denoted I, II and III as from the N terminus to the C terminus of the protein. The N terminal 76 amino acids of domain I (subdomain Ia) are involved in the binding of an enhancer-like sequence named Internal Activating Sequence (IAS) (Mizuuchi, 1989), while the remaining part of domain I until around amino acid 247 binds to the end-type MuA binding sequences (Nakayama *et al.*, 1987). The second domain between aminoacids 247 and 574 contains determinants for DNA cleavage and strand transfer (Betermier *et al.*, 1987). The most C terminal region of domain III has been shown to interact with the MuB protein (Baker *et al.*, 1991; Leung and Harshey, 1991). This has a molecular weight of 35 kDa, contains 312 amino acids (Miller *et al.*, 1984), and has an ATPase activity that is

stimulated by the presence of MuA and DNA (Maxwell *et al.*, 1987). MuB has been shown to bind DNA non-specifically and, in reactions where ATP and MuA are present, stimulates the formation of inter-molecular strand-transfer products (Maxwell *et al.*, 1987). In solution the MuA transposase exists principally as a monomer, and binds to single transposon ends (Kuo *et al.*, 1991). Each of the two *Mu* ends contains three binding sites for MuA denoted L1, L2 and L3 for the left end, and R1, R2 and R3 for the right end of the element, with each of these sites being bound by a monomer of transposase (Kuo *et al.*, 1991) (**Figure 1.4b**). The DNA binding step does not require metal ions and the binding seems to be reversible (Mizuuchi *et al.*, 1992a).

Mu transposition reaction also requires the formation of higher order protein-DNA complexes termed synaptic complexes. A series of three such complexes have been described, namely a stable synaptic complex (SSC), a Cleaved Donor Complex (CDC) and a Strand Transfer Complex (STC).

Divalent metal ions are required for the formation of the Stable Synaptic Complex (SSC) or type 0 transpososome (Mizuuchi, 1992b). This complex contains a tetramer of MuA protein bridging the two *Mu* ends, and accumulates in the presence of Ca^{2+}, which promotes the assembly of the complex but does not support the catalytic steps of the reaction (Baker *et al.*, 1991). Thus, it has been proposed that the divalent metal ions play a double role in *Mu* transposition: one in synaptic complex assembly and the other in the catalytic steps of cleavage and strand transfer (Mizuuchi, 1992b). Formation of a

Stable Synaptic Complex also requires the presence of the HU protein from *Escherichia coli* (Craigie and Mizuuchi, 1985a), and that of a *Mu*-internal enhancer-like sequence termed Internal Activating Sequence (IAS) (Mizuuchi and Mizuuchi, 1989). In the SSC, the junctions between *Mu* DNA and donor DNA are still intact, so the formation of this complex precedes the cleavage steps (Mizuuchi, 1992b). Inside this complex the *Mu*A transposase is tightly bound to the L1, R1 and R2 binding sites. *Mu*A does not stably bind the other three sites namely R3, L2 and L3 or the IAS. These sites are required only for the assembly of SSC but not for the subsequent cleavage reactions (Mizuuchi, 1992b).

In the presence of Mg^{2+} ions, the SSC is readily converted to the Cleaved Donor Complex (CDC) or type I transpososome in which the junction between *Mu* ends and the donor DNA is nicked at the 3' ends of the transposon. This generates free 3' OH groups at both ends of *Mu* (Craigie and Mizuuchi, 1987; Surette *et al.*, 1987). Protein crosslinking experiments have shown that this complex also contains a tetramer of MuA transposase (Lavoie *et al.*, 1991). The same sites are stably bound inside the CDC as in the SSC.

CDC complex accumulates in the presence of Mg^{2+} when MuB protein is absent from the reaction. In contrast, the CDC can readily interact with a target DNA that is bound by MuB, protein and undergo strand transfer inside a complex termed Strand Transfer Complex (STC) or type 2 transpososome (Craigie and Mizuuchi, 1987; Surette *et al.*,

1987). The strand transfer products can be detected in the absence of MuB albeit with a much lower frequency, and they are qualitatively different. When MuB is absent, *Mu* transposes to sites on the same DNA molecule generating intramolecular transposition products (Maxwell *et al.*, 1987). MuB has been shown to bind DNA which does not contain *Mu* sequences, a phenomenon named target immunity. MuB binding is non-sequence specific and requires the presence of ATP. MuB-bound DNA is a preferred site for *Mu* integration resulting in a high accumulation of intermolecular transposition products (Adzuma and Mizuuchi, 1988; Adzuma and Mizuuchi, 1991; Maxwell *et al.*, 1987) (**Figure 1.4c**).

After joining of the 3' ends of *Mu* to the target DNA during the strand transfer reaction, replication of the *Mu* genome initiates at the junction between the left end of the element and the target DNA (Wijffelman and Lotterman, 1977). This reaction requires the presence of *E. coli* proteins DNA gyrase (Sokolsky and Baker, 2003) and ClpX chaperone proteins (Kruklitis and Nakai, 1994; Levchenko *et al.*, 1995; Mhammedi-Alaoui *et al.*, 1994). DNA replication then generates a copy of *Mu* DNA at both the donor and the target loci.

In conclusion, *Mu*'s life cycle comprises two transposition steps different in mechanism: an integrative transposition and a replicative transposition process. After infection, *Mu* DNA is separated from the flanking DNA carried from the previous host, and integrated into the genome of a new *E. coli* cell through integrative transposition.

Both the excision and insertion steps require the *Mu*A protein. After integration 1-10% of phages enter the lysogenic cycle becoming an integral component of the host genome (Howe and Bade, 1975). The majority of phages enter the lytic cycle in which the phage genome undergoes multiple rounds of replicative transposition. This increases the number of phage particles to 50-100 per cell eventually causing the destruction of the host cell. The *Mu* repressor protein that binds to an operator sequence and shuts off early transcription, guards the entry to the lytic cycle. Thus, *Mu* can be regarded as a gigantic transposon of 37 kb that travels about disguised as a bacteriophage.

Apart from the replicative mechanism of DNA transposition, the cut-and-paste mechanism (**Figure 1.2b**) is used by a large number of elements such as the bacterial elements Tn*10* and Tn*5*, and the members of the *Tc1/mariner* superfamily of transposons. This mechanism also requires the formation of synaptic complexes between the two transposon ends. One important difference is that both DNA strands need to be cleaved during the cut-and-paste transposition. The next two sections discuss the mechanism employed by the Tn*10* and the Tn*5* elements to achieve this, using a so-called hairpin mechanism.

1.3.4 Tn*10* transposition reaction

Apart from the replicative DNA transposition, the cut-and-paste mechanism represents another modality used by transposons to mobilize their DNA (see **Figure 1.2b**). The best characterised bacterial elements which use this mechanism are the Tn*10* and Tn*5*

transposons. An important feature of their mechanism as discussed below is the formation of a hairpin intermediate required for complete transposon excision.

Tn*10* is a composite bacterial transposon. It is made up of two insertion sequences, IS*10*-Left and IS*10*-Right, in inverted orientation, flanking genes that confer resistance to tetracycline and several other ORFs, including a predicted sodium-dependent glutamate permease, one gene with homology to bacterial transcriptional regulators and others with unknown function. The Tn*10* transposon that has been studied in depth was isolated from the enteric bacterium *Shigella flexneri* (Watanabe *et al.*, 1968). Tn*10* is 9,147 bp in length and contains nine ORFs. IS*10* is 1,329 bp in length and only IS*10*-Right contains an ORF encoding the functional transposase protein. The two ends of IS*10* are denoted outside and inside end with respect to their position in Tn*10*. Both the outside and the inside ends contain binding sites for the transposase. The outside ends also contain binding sites for the *Escherichia coli* protein Integration Host Factor (IHF) (**Figure 1.5a**). Two roles have been documented for the IHF protein in Tn*10* transposition. The addition of IHF is required to form the transposition products when linear or plasmid DNA with reduced supercoiling are used as substrates. When the substrates are supercoiled plasmids the addition of IHF is not required. Thus, it was suggested that IHF serves as a "supercoiling relief factor" acting as a "molecular spring" (Chalmers *et al.*, 1998). At low concentrations of IHF, the strand transfer products are generated by a random collision between transposon ends and the target DNA. When present in high concentrations, IHF stimulates the formation of one

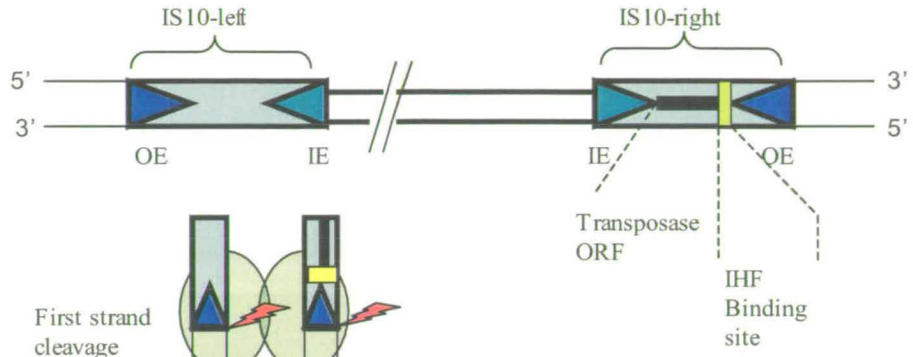

Figure 1.5a. Structure of Tn*10* and IS*10*

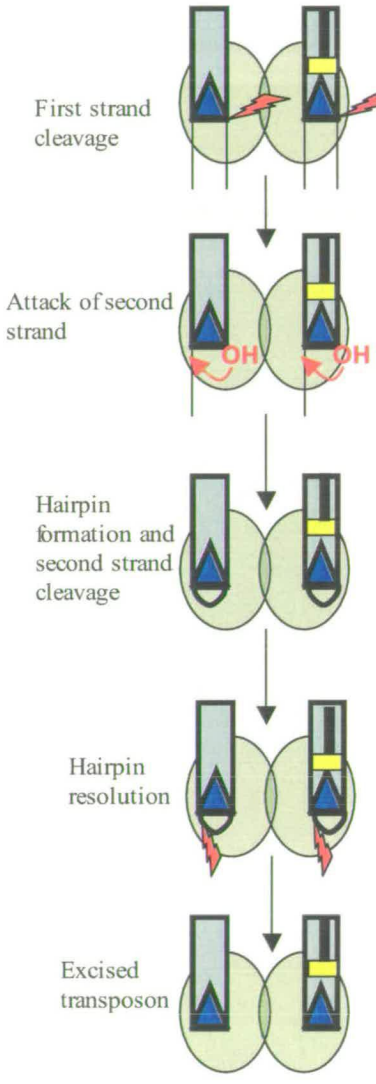

Figure 1.5b. The sequence of DNA cleavages during Tn*10* excision.

Figure 1.5 Transposon Tn*10*.

Figure 1.5a. Structure of Tn*10* and IS*10*. The abbreviations used are, OE for Outside End, IE for Inside End, IHF for Integrating Host Factor, ORF for open reading frame.
Figure 1.5b. The sequence of DNA cleavages during Tn*10* excision. The transposase monomers are represented as green ovals. Only the OE of IS*10* are shown. The 'red flashes' represent the DNA nicking steps.

particular strand transfer product imposing topological constraints on the reaction. This phenomenon is referred to as channeling (Chalmers *et al.*, 1998).

Tn*10* was the first transposon shown to transpose by a cut-and-paste mechanism (Bender and Kleckner, 1986; Kleckner, 1979). In this mechanism the transposon excises from its original location in the DNA molecule (donor site) and inserts into a new location (target site). Upon integration, Tn*10* generates gaps of nine nucleotides at each transposon end, gaps that when repaired result in 9 bp duplications of the original target sequence (Benjamin and Kleckner, 1989), *In vitro* studies performed with purified Tn*10* transposase, purified IHF protein and linear or supercoiled plasmid DNA substrates showed that Tn*10* excises through a series of three chemical steps. Nicking of the first DNA strand occurs at the 3' end of both strands of the transposon. This generates free 3' hydroxyl groups, which attack and cleave the opposite strand forming a hairpin structure at each transposon end. The hairpins are opened in a process usually termed hairpin resolution, generating free 3'-OH and 5'-P groups (Kennedy *et al.*, 1998) (**Figure 1.5b**).

All these cleavage steps occur inside a higher-order protein-DNA complex named Paired End Complex when it is formed with synthetic DNA substrates containing only the transposon ends or, synaptic complex when the full length transposon is present (for example on a plasmid DNA substrate). This complex contains the two transposon ends bridged together by transposase monomers. In the case of Tn*10* the PEC contains two

transposase monomers (Haniford , 2002). When the transposition reaction is analysed on native polyacrylamide gels or EMSA, three transposase-DNA complexes are observed, a pre-cleavage PEC in which both transposon-donor junctions are intact, a complex in which the flanking DNA has been completely removed from one transposon end (termed Single End Break Complex), and a complex in which the flanking DNA has been completely removed from both transposon ends (or Double End Break Complex). No complex containing the transposase bound to one transposon end has been observed in the Tn*10* transposition system. Complete transposon excision is required for the reaction to proceed to the next step – target capture (Sakai and Kleckner, 1997). Two distinct interactions with the target have been described, a non-sequence specific interaction, and a subsequent sequence specific interaction (Junop and Haniford, 1997; Sakai and Kleckner, 1997). An initial non-sequence specific interaction is converted to a sequence specific interaction, which in turn is converted to a strand transfer complex (Bolland and Kleckner, 1996). The strand transfer steps at the two transposon ends seem to be tightly coupled because one predominant form of strand transfer complex is observed, the double end strand transfer product (Bolland and Kleckner, 1996). This is in contrast with the excision step where the cleavage at the two transposon ends can be uncoupled.

Since the PEC formed by Tn*10* contains two monomers of transposase, a question was raised: how many active sites perform the chemical steps involved in transposition? Tn*10* was the first transposon to provide evidence that a single active site is required at

one transposon end to perform all the biochemical steps involved in cleavage and strand transfer (Bolland and Kleckner, 1996). By mixing wild type and catalytically inactive mutant transposase monomers and monitoring the outcomes of the reaction, it was found that one active site performs all the cleavage steps at one transposon end. Also, the same active site that performed the cleavage steps catalyses the strand transfer reaction. The same mechanism was subsequently documented for bacteriophage *Mu*, and the *HIV* integrase (Engelman *et al.*, 1991; Namgoong and Harshey, 1998; Williams *et al.*, 1999) with the difference that for *Mu* and *HIV* only the nicking of the first strand and the strand transfer steps are catalysed by the transposase/integrase active site. The cleavage of the first and second strands by the same active site in the Tn*10* system raises one problem, since the two DNA strands are of opposite polarities, there must be a significant conformational change in the transposase in order to accommodate the two cleavage reactions. This however remains to be demonstrated.

1.3.5 Tn*5*

The first transposition system for which a crystal structure of a synaptic complex is available is the Tn*5* transposon. As discussed below, this gave unprecedented insight into the structure of a synaptic complex involved in cut-and-paste transposition.

Tn*5* is a composite prokaryotic transposon resembling Tn*10* in the overall structure. It is composed of two insertion sequences (IS), IS*50*R and IS*50*L flanking a region of DNA containing resistance genes for kanamycin, bleomycin and streptomycin. Each IS

is delineated by 19 bp inverted repeats termed Outside End (OE) and Inside End (IE) with respect to Tn*5*. In this way Tn*5* is flanked by the two inverted OE repeats of 19 bp. These are imperfect copies of each other differing at seven nucleotides. Apart from the two OE, Tn*5* transposition requires the 476 amino acids transposase encoded by IS*50*R. Tn*5* transposes by a cut and paste mechanism that excises the transposon from the donor DNA and inserts it at another location. For Tn*5* as for Tn*10*, the mechanism involves four catalytic steps: transferred strand nicking, hairpin formation, hairpin resolution, and strand transfer (Bhasin *et al.*, 1999). All these steps occur inside a synaptic complex. Single transposase molecules bound to a single end are inactive for catalysis (Bhasin *et al.*, 2000; Davies *et al.*, 2000; Naumann and Reznikoff, 2000).

Transposition is a dangerous event for the host cell, so that a highly active transposon would be evolutionarily selected against. Tn*5* transposition is in fact a very inefficient process *in vivo*. This is due to at least three levels of regulation described so far. The first mechanism involves a protein identical to the transposase but lacking the N-terminal 55 amino acids (Isberg *et al.*, 1982; Johnson and Reznikoff, 1984; Johnson *et al.*, 1982). This inhibitor forms inactive multimers with the transposase, "poisoning" its activity. The second inhibitory mechanism is enforced by the transposase itself. It has been proposed that the C-terminal domain of the transposase inhibits the N-terminal DNA binding domain via an interaction between residues 40 and 450 (Weinreich *et al.*, 1994b). This blocks the protein immediately after translation. As a consequence, the wild type protein is completely inactive *in vitro*. A mutation has been identified at the

position 372 that changes the Leu residue to a Pro residue allowing the N-terminus to move away from the C-terminus and thus bind DNA (Weinreich *et al.*, 1994a), (Davies *et al.*, 2000). The third regulatory mechanism is enforced by the Dam DNA-methylation process, which down-regulates transposase expression, up-regulates the expression of the inhibitor protein, and inhibits the recognition of the transposon IE by the transposase (Reznikoff, 2002).

Tn*5* transposase is an inefficient protein, evolved towards a reduced level of activity. This enabled the identification of hyperactive variants of the proteins. One such mutant changes the Glu at position 54 into a Lys. This mutation EK54 together with LP372 defines the hyperactive EK/LP transposase. The EK/LP double mutant generated crystal structures of transposase-DNA complexes. The first crystal structure was that of a dimer of Tn*5* inhibitor protein (Davies *et al.*, 1999) followed shortly by crystal structures of the Tn*5* Paired End Complex (Davies *et al.*, 2000; Lovell *et al.*, 2002; Steiniger-White *et al.*, 2002). All these represent crystal structures of post-cleavage synaptic complexes (also named Double End Break complexes) and have given unprecedented insights into the transposition mechanism. The transposase comprises three domains: the N-terminal DNA-binding domain consisting of the first 70 amino acids, the catalytic region containing the DDE catalytic motif in a region spanning 300 amino acids and, the C-terminal domain responsible primarily for protein-protein interactions, although this delimitation of functions is far from complete (Davies *et al.*, 2000). The Tn*5* synaptic complex contains two monomers of transposase that contact

the two transposon ends in *cis* and in *trans*. That is, the N-terminal DNA binding domain of one monomer binds to one transposon Outside End (*cis* protein-DNA contacts) and, through residues in the catalytic region contacts the opposite transposon end (*trans* protein-DNA contacts). This organisation shows that DNA cleavages occur *in trans* inside the Tn*5* synaptic complex (Davies *et al*, 2000; Naumann and Reznikoff, 2000). Each monomer of transposase contains three amino acids in the active site: Asp97, Asp188 and, Glu326 forming the so-called DDE motif. This is conserved in the transposase/integrase family of proteins, and in the Tn*5* crystal structure has been shown to coordinate two Mn^{2+} ions (Lovell *et al.*, 2002), one by the residues Asp188 and Asp97, and the other by Asp97 and Glu326. The two divalent metal ions most likely to be Mg^{2+} *in vivo*, serve to activate water molecules required as nucleophiles at the transferred strand nicking and hairpin opening steps. The OH group from the transposon 3' ends acts as a nucleophile at the other two steps: hairpin formation and strand transfer.

The available crystal structure is in good agreement with the cleavage via a hairpin structure proposed for Tn*10* and Tn*5* (Bhasin *et al.*, 1999; Kennedy *et al.*, 1998). This mechanism requires the scissile phosphodiester bond of the non-transferred strand to be close enough to the 3'OH of the transferred strand to permit a transesterification reaction to occur (hairpin formation). Indeed, the distance between the two components is only 3.5 Å. The proximity of the two DNA strands seems to be facilitated by protein-DNA contacts near the active site that stabilize a bend in the backbone of the non-

transferred strand. Moreover, a thymidine base next to the 5' end is rotated away from the interior of the DNA molecule such that the base is held inside a hydrophobic binding pocket in the transposase. This phenomenon, termed 'base flipping', has been observed in other enzymatic systems that act upon DNA structures (Steitz *et al.*, 1994). The thymidine base is maintained in the 'flipped' position through interactions with amino acid residues forming a so-called YREK motif, conserved in the IS*4* family of transposases. Mutational studies in Tn*5* and Tn*10* transposase have shown that this motif is required for catalytic activities in both systems.

Tn*5* inserts the 3' OH groups at each transposon end in the two opposite strands of a target DNA at a distance of 9 bp from each other. The distance between the scissile bonds in the B DNA form would be approximately 35 Å whereas the distance between the two transposon 3' OH groups in the crystal structure is 41 Å. This discrepancy might be explained by Tn*10*'s preference to insert in negatively supercoiled DNA as this has an increased distance between the two scissile bonds targeted by the Tn*5* transposon (Davies *et al.*, 1999).

As discussed above, the transposition mechanism employed by Tn*10* and Tn*5* involves the formation of hairpin structures at the transposon ends. An alternative to this mechanism is represented by the RAG1/RAG2-catalysed recombination reaction discussed in the next section.

1.3.6 V(D)J recombination reaction

A process that bears striking similarities to the transposition reaction is the eukaryotic recombination required for the assembly of functional immunoglobulin (Ig) and T cell receptor genes, also named V(D)J recombination. It has been suggested that V(D)J recombination has evolved from an ancient transposon (Lewis and Wu, 1997; Gellert, 2002; Zhou *et al.*, 2004). The genes for the immunoglobulins and T-cell receptor genes of vertebrates have an interesting structure. They exist initially as a linear array of gene fragments that require recombination to generate the functional genes (Gellert, 2002; Tonegawa, 1983). These gene fragments named variable (V), diversity (D) and joining (J) are present in a very large number, and during V(D)J recombination are joined to form functional variable region exons in a V-J or V(D)J array. Later, RNA splicing links the recombined variable region exon to a constant region. The V, D and J gene fragments are flanked by specific DNA sequences termed Recombination Signal Sequences (RSS), where the DNA is cleaved. An RSS contains a conserved heptamer (CACAGTG) and a conserved nonamer sequence (ACAAAAACC) separated by 12 or 23 bp of non-conserved DNA named spacer DNA. Thus, RSSs can be divided into 12RSSs and 23RSSs (**Figure 1.6a**). This division has important consequences on the reaction. The V(D)J recombination is catalysed by two enzymes: Recombination Activating Gene 1 (RAG1) and Recombination Activating Gene 2 (RAG2) (Oettinger *et al.*, 1990). These two proteins are only expressed in lymphoid cells, and are the only lymphoid-specific factors required for V(D)J recombination. RAG1 and RAG2 can

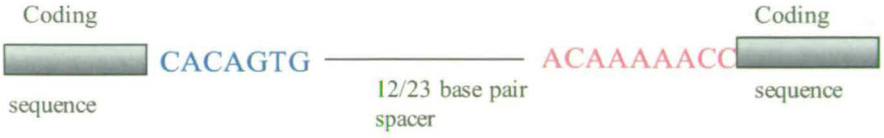

Figure 1.6a. Structure of Recombination Signal Sequences (RSS)

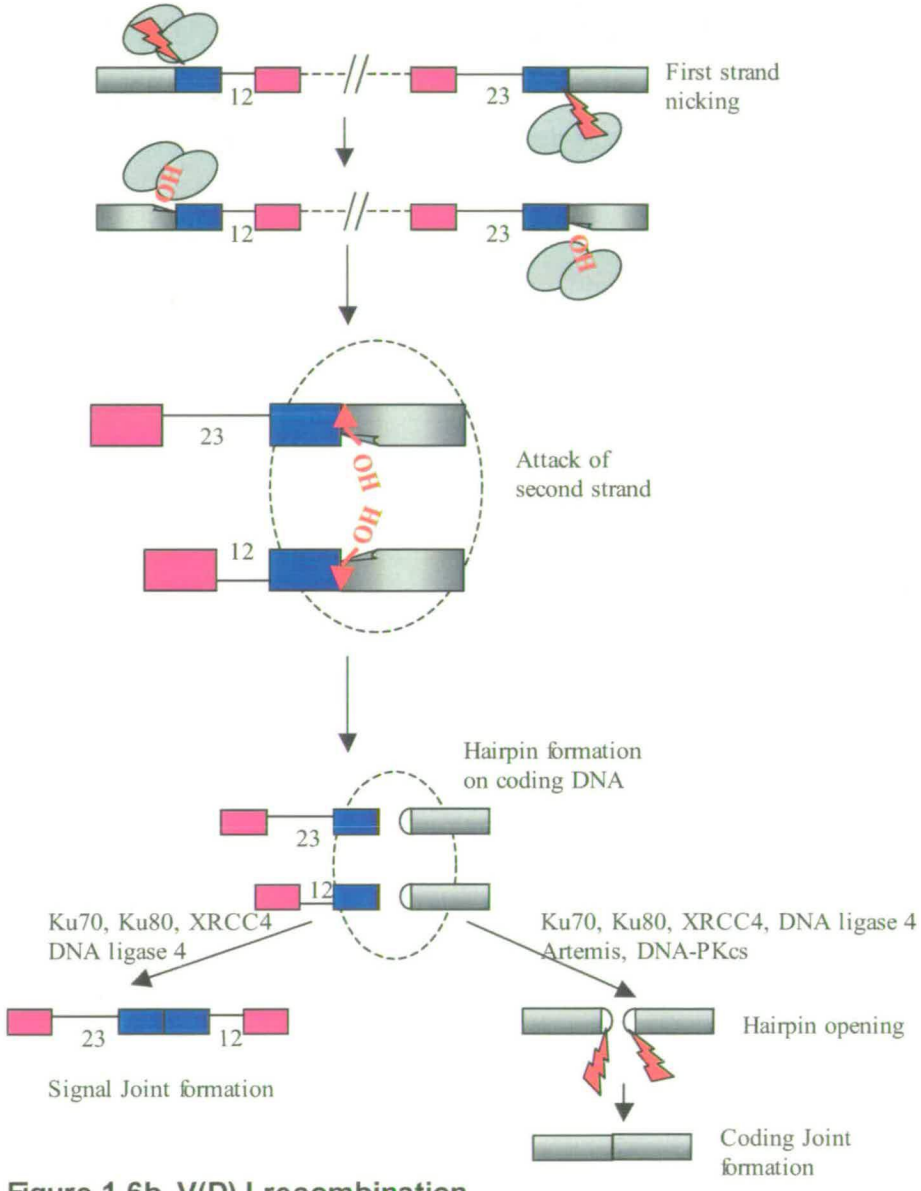

Figure 1.6b. V(D)J recombination

Figure 1.6. The V(D)J recombination reaction.

Figure 1.6a. Structure of Recombination Signal Sequences (RSS). As discussed in more detail in text, the RSSs are the site of DNA binding and cleavage by the RAG proteins.They are composed of a conserved heptamer sequence (blue) and a conserved nonamer sequence (red) separated by a less conserved spacer sequence of 12 or 23 bp (black). The coding sequences that flank the RSS are represented by grey bars.

Figure 1.6b. V(D)J recombination. A complex between RAG1 and RAG2 is active in V(D)J recombination (RAG2 not shown). First strand nicking occurs outside the synaptic complex. Second strand cleavage occurs through a hairpin intermediate formed in the coding DNA. Hairpin formation requires synaptic complex formation. Note that the stoichiometry of the synaptic complex is not represented. The DNA nicking steps are represented by red flashes.

initiate V(D)J recombination when co-expressed in non-lymphoid tissues. Mice with disruption of the RAG1 and RAG2 genes are completely defective for V(D)J recombination and have no mature B and T cells. Thus, completion of V(D)J recombination is required for the maturation of lymphoid cells (Mombaerts *et al.*, 1992; Shinkai *et al.*, 1992). Full length RAG1 and RAG2 have been found to be insoluble and/or inactive when purified. For this reason, truncated versions of the proteins have been used extensively for *in vitro* biochemical studies. The truncated versions contain the amino-acids 384-1008 of murine RAG1 and 1-383 of murine RAG2, and are able to catalyse all the chemical steps involved in V(D)J recombination. Mutational studies have identified three acidic residues in RAG1 that are required for catalysis (D600, D708, E962) (Landree *et al.*, 1999). A complex of RAG1 and RAG2 cleaves the RSS at the junction between the heptamer and the neighbouring coding sequence. The cleavage occurs in two steps. First, a nick is introduced 5' of the heptamer sequence. This generates a free phosphate group on the heptamer and a free hydroxyl group on the coding sequence. The second step is represented by the attack of the free hydroxyl group on the opposite strand generating two types of cleavage products: a hairpin-terminated coding end, and a blunt signal end (McBlane *et al.*, 1995). The cleavage reactions require the presence of divalent metal ions. In the presence of Mg^{2+}, the first strand cleavage occurs on an isolated RSS which can be either a 12RSS or an 23RSS (van Gent *et al.*, 1996; Yu *et al.*, 2000). In contrast, the second strand cleavage occurs only inside a synaptic complex or Paired End Complex containing a pair of RSSs: one 12RSS and one 23RSS. This preference has been termed the 12/23 rule (Eastman *et al.*,

1996; Kim and Oettinger, 1998) and seems to be enforced by the HMG1 proteins at the stage of hairpin formation (West and Lieber, 1998). HMG proteins are non-specific DNA binding and bending proteins, thus presumably distorting the DNA so that the RAG proteins can bind better (van Gent *et al.*, 1997). When the divalent metal is changed to Mn^{2+}, double cleavage can occur on a single end without synaptic complex formation (**Figure 1.6b**).

The stoichiometry of complexes formed by the RAG proteins with their DNA substrate has been analysed by several laboratories (Ciubotaru *et al.*, 2003; Godderz *et al.*, 2003; Landree *et al.*, 2001; Mundy *et al.*, 2002; Rodgers *et al.*, 1999; Swanson and Desiderio, 1999). The results are not yet entirely clear. Two main models have been proposed. One model suggests that a dimer of RAG1 assembles on a RSS and this is the precursor forming the synaptic complex (Bailin *et al.*, 1999; Swanson, 2002a; Swanson, 2002b; Swanson, 1999). A second model proposes that a higher number of RAG1 molecules bind to a RSS (Mundy *et al.*, 2002). Some support for the second model comes from two other studies. In one study, Size Exclusion Chromatography was used to analyse the behaviour of RAG1 in solution (Godderz *et al.*, 2003). This study found that at low ionic strength (0.2 M NaCl) RAG1 exists as a mixture of mainly dimers and tetramers, with the latter increasing upon addition of RSS. A second study using a mixture of active and catalytically inactive mutant RAG1 found that a single active site is sufficient for both nicking and hairpin formation at one RSS. The same study showed that a PEC formed by heterodimers of wild type and mutant RAG1 proteins is able to

perform nicking and hairpin formation at both 12 and 23 RSSs. This suggests at least a tetramer of RAG1 inside the synaptic complex (Landreeet al., 2001). The number of RAG2 subunits is also uncertain with either one molecule (Swanson, 1999) or one or two molecules under certain conditions (Mundy *et al.*, 2002; Swanson, 2002a; Swanson, 2002b).

The V(D)J recombination can be divided into two main types of reactions: DNA cleavage reactions and DNA joining reactions. The DNA cleavage reactions discussed above require only the RAG and one of the HMG proteins (McBlane *et al.*, 1995; van Gent *et al.*, 1997). As a consequence of RSS cleavage by the RAG proteins, two kinds of DNA ends are generated: blunt end signal ends, and coding ends that terminate in hairpin structures. Both kinds of ends are ligated, generating Signal Joints (SJ) and Coding Joints (CJ), respectively. The processing of CJ and SJ is quite different. SJ are simple, usually precise end-to-end fusions of two heptamer sequences (Lewis *et al.*, 1985; Lieber *et al.*, 1988). CJ are much more variable. Both addition and loss of nucleotides has been found in CJ. Two types of nucleotide additions are present: templated and non-templated. The templated addition of nucleotides (called P nucleotides, from palindromic) are caused by an off-center nicking of the hairpin coding ends (Lafaille *et al.*, 1989; McCormack *et al.*, 1989). The non-templated addition of nucleotides generates sequences of up to 15 nucleotides in length (the so-called N regions). These are a result of a lymphoid-specific activity termed terminal deoxynucleotidyl transferase (TdT) (Gilfillan *et al.*, 1993; Komori *et al.*, 1993). Less is

known about the causes of nucleotide loss in CJ. These deletions do not seem to be generated by a lymphoid-specific activity, occurring in all cell types in which V(D)J recombination has been induced. Both endo-and exo-nuclease activities have been suggested (Besmer *et al.*, 1998; Lewis *et al.*, 1985). The addition and deletion of nucleotides from coding joints has important consequences, increasing the diversity of antigen binding sites beyond that generated by combinatorial joining of the gene segments.

The DNA joining reactions require an array of proteins that have recently started to be identified. In eukaryotes a majority of double-strand breaks (DSBs) are resolved through two repair pathways: homologous recombination (HR) and non-homologous end joining (NHEJ). Homologous recombination uses information present on a homologous template to repair the breaks. Non-homologous end joining re-joins broken DNA ends with little or no sequence homology and it is the predominant pathway for repair during G1 phase of the cell cycle, when RAG associated DSBs are generated (Rothkamm *et al.*, 2003; Takata *et al.*, 1998).

The best-studied transposition systems have been discussed above. What is already known for these systems however is largely still an open question for what is arguably the most widespread family of transposons in nature – the *mariner* family of transposons. These eukaryotic elements belong to a very wide superfamily including the *Tc1* family with its best-studied members *Tc1*, *Tc3* and the artificially reconstructed

element *Sleeping Beauty*. In the next section (section 1.4) the transposition of members of the *Tc1* family is discussed (section 1.4.1.1 for *Tc1* and *Tc3* and section 1.4.1.2 for *Sleeping Beauty*).

1.4 The Tc1/mariner *family of transposable elements*

The name *Tc1/mariner* comes from the two best studied members. *Tc1* is a transposon discovered in *Caenorhabditis elegans* where it represents the most important source of mutations and gene inactivations (Emmons *et al.*, 1983; Liao *et al.*, 1983). Since its discovery in *C. elegans*, seven other members of the *Tc1* family have been described only in this organism (Collins *et al.*, 1989). Members of the *Tc1* family of transposons have been identified in a multitude of organisms ranging from *Drosophila* (Caizzi *et al.*, 1993; Robertson and Lampe, 1995b) to the trout *Epatretus stouti* (Heierhorst *et al.*, 1992).

The first *mariner* element was discovered in *Drosophila mauritiana* (Jacobson and Hartl, 1985; Jacobson *et al.*, 1986). Since then, related elements have been found in many other species of arthropods (Robertson, 1993; Robertson and MacLeod, 1993); *C. elegans* (Sedensky *et al.*, 1994); flatworms: *Dugesia tigrina*, *Stylochus zebra* and, *Bdelloura candida*; hydras: *Hydra littoralis* and *Hydra vulgaris* (Robertson, 1997); fungi (Langin *et al.*, 1995); plants (Feschotte *et al.*, 2003; Feschotte *et al.*, 2002) and, the human genome (Auge-Gouillou *et al.*, 1995; Morgan, 1995; Oosumi, 1995; Smit

and Riggs, 1996). The *mariner* family is probably the most widespread family of transposons in nature. Most *mariner* elements are inactive due to deletions or mutations. Only one naturally occuring element has been shown to be active: *Mos1* from *D. mauritiana* (Bryan *et al.*, 1987). An artificially reconstructed element from the horn-fly *Haematobia irritans* and named *Himar1* has also been shown to be active (Lampe *et al.*, 1996; Medhora *et al.*, 1988; Robertson, 1995a).

Elements belonging to the *Tc1/mariner* family of transposons have a relatively simple structure containing a gene encoding the transposase required for their mobilisation. The gene is flanked by inverted terminal repeats (ITRs) of various lengths (**Figure 1.7**), (**Table1**).

More recently, another class of elements has been added to the *Tc1/mariner* family of transposons. The bacterial IS*630* is a 1.15 kilobase sequence isolated from *Shigella sonnei* with terminal inverted repeats of 32 bp. It encodes a 343 amino acids transposase. This element does not mediate co-integration of flanking sequences as expected if a replicative mechanism would have been used for transposition, and it exclusively inserts at TA dinucleotide sequences. This resembles the *Tc1/mariner* family, which has thus become IS*630-Tc1-mariner* superfamily of transposable elements (Doak *et al.*, 1994).

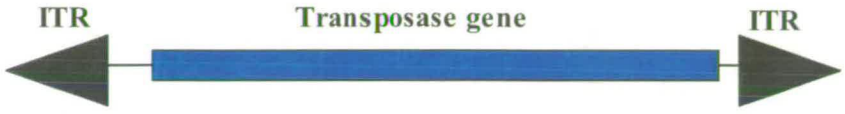

Figure 1.7. General structure of the Tc*1/mariner* elements

Family	Element	Transposase	ITR (bp)	Organism
Tc*1*	Tc*1*	343	54	*C. elegans*
	Tc*3*	1015	462	*C. elegans*
	Sleeping Beauty	341	230	fish
mariner	*Mos1*	346	28	*D. mauritiana*
	Himar 1	348	27	*Haematobia iritans*

Table 1. Best-studied members of the *Tc1/mariner* family of transposons

1.4.1 The transposition of *Tc1* family of transposons

1.4.1.1 *Tc1* and *Tc3* transposition

At present seven groups of Tc-family members have been identified in *C. elegans*. The best-studied members are *Tc1* and *Tc3*. *Tc1*-like elements are about 1.3 to 2.4 kb in length, are flanked by inverted terminal repeats of various lengths and, encode a single transposase protein. The transposase contains a DDE motif conserved in most transposases and integrases. Transposition of both elements has been analysed *in vivo*, in a cell-free system, and biochemically using transposase expressed and purified from *E. coli*. Both elements transpose by a cut-and-paste mechanism. It has been shown that *Tc3* excises through a pair of staggered double strand cleavages at both transposon ends. Cleavage at the 3' ends occurs precisely at the end of the transposon sequence, and exposes a 3'-OH group, whereas the cleavage at the 5' ends occurs two nucleotides within the *Tc3* inverted repeats. The 3' transposon ends attack the target DNA with a 2 bp stagger so that upon integration 4 base gaps are generated. These gaps are repaired by the host DNA repair activities. The same mechanism has been documented for *Tc1* (Vos *et al.*, 1996). In both cases the transposase is the only protein required for transposition, at least *in vitro* (Vos *et al.*, 1996). Host proteins are involved however, in the repair of DNA breaks after transposon insertion and excision, and maybe in the regulation of transposition.

For *Tc1*, the ITRs are 54 bp and are perfect copies of each other. The minimal domain of *Tc1* transposase required for binding to the inverted repeat is contained within the N-terminal 63 amino acids (Vos *et al.*, 1993). Further studies have shown that the N-terminus of *Tc1* transposase has a bipartite structure with two helix-turn-helix motifs (HTH). The first HTH motif is similar to the paired domain of some transcription factors, and is followed by a second HTH embedded in a homeo-like DNA binding domain (Ivics *et al.*, 1996; Plasterk *et al.*, 1999). The paired domain is composed of two subdomains that bind DNA in adjacent major grooves on the same side of the DNA helix (Czerny *et al.*, 1993). This domain is present in a very large number of proteins including the mammalian Pax proteins (Gehring, 1993). The homeobox domain binds DNA through a helix-turn-helix structure and was first identified in a number of *Drosophila* homeotic and segmentation proteins. Most of these are known to be sequence-specific DNA-binding transcription factors, and include members of the antennapedia and bithorax complexes, involved in the determination of antero-posterior axis in *Drosophila* and the mammalian HOX genes (Gehring *et al.*, 1994).

The two DNA-binding domains in *Tc1*-like elements have different DNA-binding properties. The first HTH motif located in the first 63 amino acids is necessary and sufficient to bind specifically to the transposon ITRs. The second HTH motif located within amino acids 71 and 207 binds DNA non-specifically (Vos *et al.*, 1993). A nuclear localization signal (NLS) partially overlaps the second HTH. The NLS is flanked by putative casein-kinase II (CK II) phosphorylation sites (Ivics *et al.*, 1996).

Phosphorylation at these positions might be involved in transposition regulation, although this possibility has not been investigated. DNase I footprinting analysis indicates that transposase makes contact with ITRs DNA between positions 3 to 29 in the top strand, and between nucleotides 6 and 31 in the bottom strand. The terminal TACAGT sequence that is conserved in the *Tc1/mariner* family is not protected by contacts with the transposase in these assays (Vos *et al.*, 1993). This suggests that the terminal nucleotides are not required for the DNA binding step, but probably are involved in later steps in the transposition reaction.

Tc3 is 2335 bp in length and has inverted repeats of 462 bp at its ends. *Tc3* transposase, like that of *Tc1*, has two DNA binding activities: a sequence-specific DNA-binding region within the first 65 amino acids and a non-sequence-specific DNA-binding region within the first 159 amino acids (Colloms *et al.*, 1994). *Tc3* is unique among the *Tc1/mariner* elements because it has long, almost perfect inverted repeats. DNase I footprinting analysis has indicated that two regions within each inverted repeat are protected, one located between positions 1 and 32 and, one located around position 180 from the transposon end. The role of the internal binding site is not yet clear, but it is dispensable for transposition (Colloms *et al.*, 1994).

The structure of the first 65 amino acids from the N-terminal domain of *Tc3* transposase has been determined in complex with a DNA duplex containing 20 bp of transposon end (van Pouderoyen *et al.*, 1997). The structure shows that this region of the

transposase contains three α-helices. The second and the third helix form the HTH motif and are involved in DNA recognition in the major groove. The first helix is involved in the dimerization of the protein fragment in the crystal with each protein fragment binding one DNA. The N-terminal domain of the *Tc3* transposase also forms dimers in solution in gel-filtration experiments. The DNA in the structure is in a non-linear B conformation (van Pouderoyen *et al.*, 1997).

More recently, the full bipartite DNA-binding domain of the *Tc3* transposase has been crystallised with a 26 bp DNA substrate (Watkins *et al.*, 2004). This contains the first 135 amino acids of *Tc3* transposase encompassing the first DNA binding domain (between amino acids 1 and 44), and the second DNA binding domain (between amino acids 63 and 135). A linker of 18 amino acids separates the two DNA binding domains. Within this crystal structure two distinct dimeric interfaces are observed. The interfaces occur in both N-terminal and C-terminal domains although, it is not clear whether they are relevant *in vivo*. It is conceivable that the *Tc3* transposase is a tetramer inside the synaptic complex since two DNA binding sites have been described (Colloms *et al.*, 1994). One has to bear in mind though, that the internal binding site seems to be dispensable for transposition (Colloms *et al.*, 1994).

1.4.1.2 Transposition of the *Sleeping Beauty* transposon

Not a single active *Tc1/mariner*-type transposon was found in vertebrate genomes. For example, of the nearly 10,000 *Tc1/mariner* elements in the haploid human genome,

none appears to encode an active transposase (Lander *et al.*, 2001; Venter *et al.*, 2001). From inactive elements found in salmonid fish, a functional *Tc1*-like transposon was reconstructed and named *Sleeping Beauty* (Ivics *et al.*, 1997). More recently, a similar approach was used to reconstruct a transposon from *Rana pipiens* (Miskey *et al.*, 2003).

Sleeping Beauty (*SB*) is flanked by non-identical 230 bp Inverted Terminal Repeats containing binding sites for the transposase, each inverted repeat containing two transposase binding sites in direct orientation (DR) (Ivics *et al.*, 1997). This organization of the inverted repeats, denoted IR/DR, is conserved in a group of *Tc1*-like transposons but not in *Tc1* itself (Plasterk, 1999). Both transposase binding sites are absolutely required for transposition (Izsvak *et al.*, 2000). Transposon versions containing two left inverted repeats are more active than the wild type transposon. Conversely, transposons with two right inverted repeats are severely impaired for transposition (Izsvak *et al.*, 2002). An enhancer-like sequence denoted HDR is found within the left DR (Izsvak *et al.*, 2002).

The N-terminal domain of *SB* transposase contains the DNA-binding domain. This domain is further divided into the PAI sub-domain and the RED sub-domain. Both contain HTH motifs and are separated by an AT-hook motif involved in DNA binding, conserved in the *Tc1/mariner* family of transposable elements (Aravind *et al.*, 1998). The PAI sub-domain has been shown to bind the HDR enhancer sequence and to mediate protein-protein interactions with other transposase subunits (Izsvak *et al.*,

2002). In solution the N-terminal DNA-binding domain of *SB* transposase is principally monomeric according to sedimentation equilibrium experiments (Izsvak *et al.*, 2002). This transposase fragment is able to form tetramers on an oligonucleotide DNA substrate containing the transposase binding sites (Izsvak *et al.*, 2002). Interestingly, these protein-DNA complexes seem to contain two DNA molecules, thus resembling a Paired End Complex with a tetrameric stoichiometry of the transposase protein.

The *SB* transposase has been shown to interact with the High Mobility Group Protein 1 (HMGB1), suggesting that the transposase might recruit HMGB1 to transposon DNA. It has been suggested that this would help the assembly of the synaptic complex, ensuring that transposon-transposase complexes are formed first at the internal DRs (Zayed *et al.*, 2003).

1.5 The mariner transposable elements

1.5.1 Introduction

The first member of the *mariner* family of transposons was isolated from *Drosophila mauritiana* as responsible for an unstable mutation of the *white* gene (Jacobson and Hartl, 1985). This mutation generated flies with peach-coloured eyes and was designated white-peach (w^{pch}). In the germ line, the mutation reverts to the wild-type with a frequency of approximately 10^{-3} per gene per generation, and also produces null

derivatives at about the same frequency (Jacobson and Hartl, 1985). The mutation also reverts in somatic cells. When this happens in the precursor cells of the eye, flies with mosaic eyes are observed. Restriction analysis and Southern blotting hybridisation with probes from the *white* gene have shown that the w^{pch} mutation was generated by the insertion of a transposable element into the 5' un-translated region of the *white* gene. The element is 1286 bp in length with 28 bp inverted terminal repeats with four mismatches (Jacobson *et al.*, 1986). Subsequently, an autosomal factor located on chromosome 3 was identified and linked with very high rates of somatic and germline reversion of the w^{pch} mutation (Bryan *et al.*, 1987). This factor is an active copy of the transposon named *Mos1* (mosaic), that can mobilize both itself and the *peach* element (Medhora *et al.*, 1988). The cloning and sequencing of the *Mos1* copy revealed a transposon of 1286 bp in length that differs from the *peach* element at eleven nucleotide positions (Medhora *et al.*, 1991). *Mos1* is the only naturally occuring active *mariner* element known to date.

Another active element from the *mariner* family, named *Himar1*, was reconstructed from inactive copies present in the genome of the horn-fly *Haematobia irritans* (Lampe *et al.*, 1996). The consensus sequence used for the reconstruction of *Himar1* reveals a 1,293 bp transposon with 27 bp inverted terminal repeats and encoding a 348 amino acid transposase (Lampe *et al.*, 1996; Robertson and Lampe, 1995b). *Himar1* is present in approximately 17,000 copies in *H. irritans* and has been implicated in a recent

horizontal transfer between three insect species: *Haematobia irritans*, *Anopheles gambiae* and *Chrysoperla plorabunda* (Robertson and Lampe, 1995b).

1.5.2 Horizontal transfer of *mariner* transposons

Soon after the discovery of the *Mos1* element, *Mariner*-Like Elements (MLEs) were identified in various organisms. Moreover, elements identified in organisms that have diverged many millions of years ago, were sometimes almost identical. One classical example is that of *mariner* elements present in *Drosophila mauritiana* and *Zaprionus tuberculatus* (Maruyama and Hartl, 1991). The *mariner* sequence obtained from *Zaprionus tuberculatus* is 97% identical with that from *Drosophila mauritiana*, a member of the *melanogaster* species subgroup. In contrast, *mariner* sequences isolated from *Drosophila tsacasi* are only 92% identical with those from *D. mauritiana* although *D. tsacasi* and *D. mauritiana* are closely related. For comparison, the coding sequences of alcohol dehydrogenase (Adh) gene are 90% identical between *D. mauritiana* and *D. tsacasi*, as opposed to only 82% between *D. mauritiana* and *Z. tuberculatus*. This strongly suggests the presence of horizontally transmitted *mariner* elements between *D. mauritiana* and *Z. tuberculatus* (Maruyama and Hartl, 1991). Many such instances have been documented. MLEs in the *mellifera* subfamily isolated from either *Drosophila erecta* (*Demar12* and *Demar19*) or from the cat flea *Ctenocephalides felis* (*Cfmar 10.6*) are between 96 and 99% identical. The disparity with the gene for sodium-potassium transmembrane pump is remarkable: only 39% identity in this case (Hartl *et al.*, 1997).

Horizontal transmission between different orders of insects has been reported for MLEs from the horn fly *Haematobia irritans* that are more than 90% identical to elements from *Anopheles gambiae*. The two organisms diverged more than 200 million years ago (Robertson, 1993). Interphylum instances of horizontal transmission have also been reported between the planarian *Dugesia tigrina* and an ant (Garcia-Fernandez *et al.*, 1995).

How does the horizontal transmission of MLEs occur? The vectors are still a matter of speculation although a few suggestions have been made. These include, a mite (Houck *et al.*, 1991), parasites, symbionts and viruses (Hartl *et al.*, 1997). Further work is required to address this issue.

It is not known how frequently *mariner* elements pass the boundaries between organisms by horizontal transfer. It has been suggested that horizontal transmission may play a critical role in the long-term survival of *mariner* elements. Although *mariner* elements are sometimes present in a large copy number per genome, the vast majority are inactive due to mutations of various sorts (Robertson, 1993). A transposon that is not positively selected for (and probably most of them are not) will accumulate mutations by chance alone. Since transposition is a dangerous event for the host organism, active transposons are probably selected against, so when the element is still active there is a selection pressure to inactivate it. This phenomenon is known as vertical inactivation (Lohe *et al.*, 1995). Transposons can also be lost from a genome by

stochastic loss. That is, the rate of loss of elements by random genetic drift is higher than the gain of elements by transposition (Lohe *et al.*, 1995).

In the recent years, the discovery of RNA interference (RNAi) (Fire *et al.*, 1998) shed light on many aspects of cell and molecular biology including transposon regulation. In fact, transposon silencing might be the natural function of RNAi (Ketting *et al.*, 1999; Tabara *et al.*, 1999). The current model for the RNAi mechanism involves the formation of double-stranded RNAs (dsRNAi), which are cleaved into small interfering RNAs (siRNAs) of 21 – 24 nucleotides in length by the Dicer enzyme. The siRNAs then mediate the degradation of their complementary messenger RNA through the RISC complex, silencing the targeted gene (Vastenhouw *et al.*, 2004). The formation of dsRNAs is central for the initiation of RNAi. Two models have been proposed for the formation of dsRNAs in response to the presence of transposons in a genome (Ketting *et al.*, 1999; Vastenhouw *et al.*, 2004). The first mechanism postulates that when the number of elements in the genome reaches a certain level, dsRNAs are formed by a readthrough mechanism from the neighbouring regions. The second mechanism proposes that since transposons are very often flanked by ITRs, the RNA corresponding to these sequences can fold and form double-stranded structures.

In this section, three mechanisms for transposon inactivation are briefly discussed, vertical inactivation, stochastic loss and RNA interference. As discussed in the next section, other mechanisms have also been proposed. When the transposon reaches a

certain copy number, the RNAi and the transposase titration (see below) may act to suppress its activity. Vertical inactivation and stochastic loss also take their toll, so the invasion of a new, naive genome through horizontal transfer might be a way to circumvent these mechanisms and survive on an evolutionary time scale.

1.5.3 Regulation of *mariner* transposition

Transposition is a potentially deleterious event for a cell harbouring an active transposon. DNA breaks and recombination events are hallmarks of transposition. If the number of such events exceeds the cell's capacity for DNA repair, both the host and the transposon are lost. Thus, transposons have evolved means to regulate their activity. For MLEs three self-regulatory mechanisms have been proposed, overproduction inhibition, dominant negative complementation and transposase titration (Hartl *et al.*, 1997).

The phenomenon of overproduction inhibition was first observed with constructs expressing the *Mos1* transposase under the control of a fusion between *Mos1* promoter and *hsp70* promoter (Lohe *et al.*, 1996a). The assay consisted in germline excision of the *peach* element from the *white* gene, resulting in flies with wild-type eyes, as opposed to peach-coloured eyes if excision did not take place. Upon increasing the number of transposase-expressing transgenes and/or heat-shock induction of the *hsp70* promoter, the rate of peach element excision decreased by as much as 45%. *Mos1* transposase subunits have been shown to interact (Lohe *et al.*, 1996b; Zhang *et al.*,

2001). It has been proposed that at high concentrations of transposase, inactive oligomers are produced that inhibit the transposition activity (Lohe *et al.*, 1996a; Lohe *et al.*, 1996b).

The dominant-negative complementation mechanism has been observed in certain missense mutations in *Mos1* transposase (Lohe *et al.*, 1996a; Lohe *et al.*, 1996b). Mutant transposase monomers interact with wild-type monomers and thus inhibit their activity. This raises the possibility that some of these mutations might be positively selected and maintained in natural populations by virtue of their negative effect on transposition (Hartl *et al.*, 1997).

Many MLEs contain deletions, and amongst these many retain one or both of their transposase binding sites. These defective elements might act as downregulators of transposition by titrating the transposase produced by active elements (Lohe *et al.*, 1996a).

1.5.4 The molecular mechanism of *mariner* elements transposition

What is known at present about the mechanism of *mariner* elements transposition has been inferred from studies on the two active members of the family, *Mos1* and *Himar1*. As discussed above, the *Mos1* element was discovered in *Drosophila mauritiana.* It is

1286 bp in length with 28 bp inverted terminal repeats and encodes a 346 amino acids transposase (Bryan *et al.*, 1990; Bryan *et al.*, 1987).The *Himar1* element resurrected from the horn-fly *Haematobia irritans* was used to express its transposase in *E. coli* for biochemical studies (Lampe *et al.*, 1996).

Transposase is the only protein required for transposition of *mariner* elements, at least *in vitro* (Dawson and Finnegan, 2003; Lampe *et al.*, 1996; Zhang *et al.*, 2001). This finding is further supported by the very broad range of host organisms in which *mariner* elements can transpose (Coates *et al.*, 1998b; Coates *et al.*, 1997; Fadool *et al.*, 1998; Gueiros-Filho *et al.*, 1997; Sherman *et al.*, 1998; Zhang *et al.*, 1998). Members of the *mariner* family of transposons use a cut-and-paste mechanism to mobilize themselves (**Figure 1.2**) (Dawson and Finnegan, 2003; van Luenen *et al.*, 1994).

The ITRs are bound by the transposase as a first step in the transposition reaction (see **Figure 1.8** for the sequence of *Mos1* ITRs). In the case of *Mos1*, the two 28 bp ITRs differ at four nucleotide positions, and transposase binds with a higher affinity to the 5' end of the element (Auge-Gouillou *et al.*, 2001; Zhang *et al.*, 2001). Its DNA binding activity was mapped in the first N-terminal 120 amino acids (Auge-Gouillou *et al.*, 2001; Zhang *et al.*, 2001). This region contains one putative helix-turn-helix motif between amino acids 88 and 108 that plays a crucial role, since a single mutation in the second helix abolishes DNA binding (Zhang *et al.*, 2001). DNaseI footprinting experiments have shown that the transposase makes contacts with the ITR between

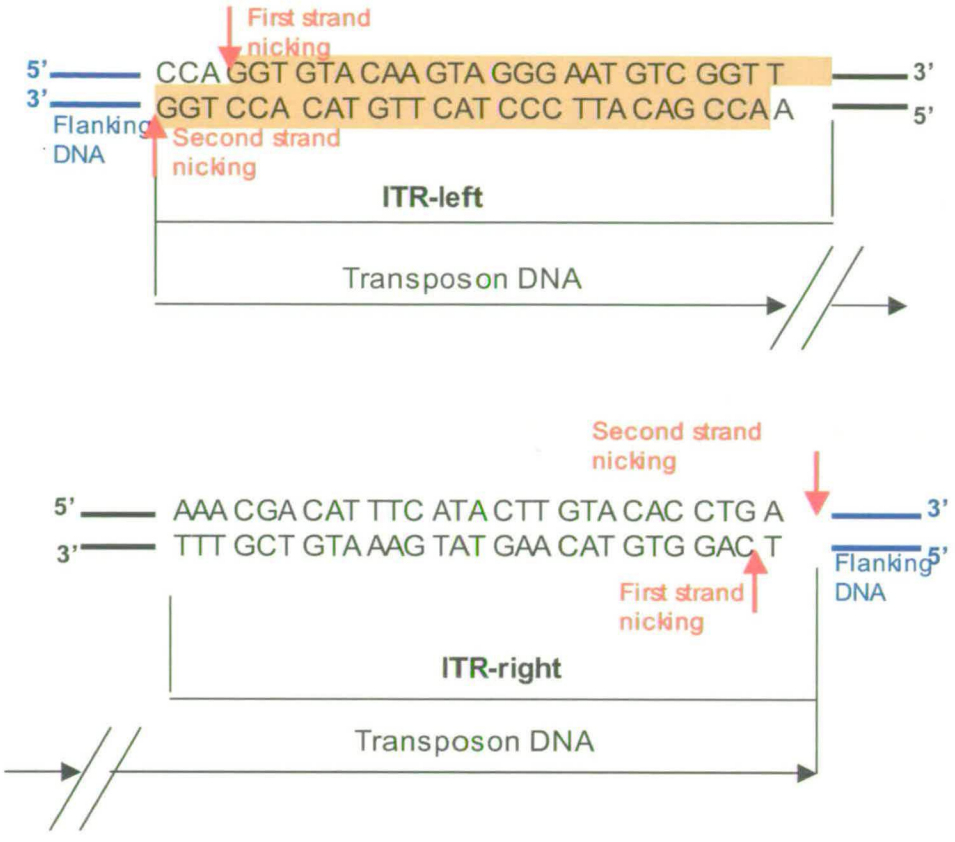

Figure 1.8. The sequence of *Mos1* ITRs and the precise position of DNA cleavages

The sequence of *Mos1* ITRs is shown. The transposon DNA is represented in black, the flanking DNA in blue. The position of DNA cleavages is marked with red arrows. *Mos1* transposase nicks the first strand three nucleotides inside the transposon DNA. The second strand is nicked precisely at the junction between the transposon and the flanking DNA. The positions protected from DNaseI digestion are highlighted for the ITR-left (Auge-Gouillou *et al.*, 2000).

positions 4 and 28 in the top strand and, between nucleotides 1 and 26 in the bottom strand (Auge-Gouillou *et al.*, 2000) (**Figure 1.8**). DNA binding does not require the presence of divalent metal ions such as Mg^{2+}, Mn^{2+} or Ca^{2+}. *Himar1* has perfect ITRs of 31 bp. These are protected from DNase I digestion between positions 4 and 32 on the top strand and 2 and 30 on the bottom strand (Lampe *et al.*, 1996).

Mos1 transposase cleaves first at the 5' end of each strand of the transposon DNA (Dawson and Finnegan, 2003) (**Figure 1.9**). Importantly, this cleavage occurs independently of synaptic complex formation (Dawson and Finnegan, 2003). The first strand cleavage occurs three bases inside the transposon DNA (Dawson and Finnegan, 2003). This is in good agreement with the footprints observed *in vivo* after *Mos1* excision, which usually consist of three bases from either the left or right end of the element flanked by the duplicated TA dinucleotides (Bryan *et al.*, 1990). The DNA cleavage activity is strictly dependent upon the presence of Mg^{2+} or Mn^{2+} ions. The ability of *Mos1* transposase to cleave at the 5' end of each strand first is in contrast with bacterial elements like Tn*10* and Tn*5* that always cleave at the 3' ends first. A reaction shown to proceed by the same pathway is the V(D)J recombination catalysed by the RAG1 and RAG2 proteins. In this case a complex of RAG1 and RAG2 proteins cleaves at the 5' end of a Recombination Signal Sequence as discussed above and the cleavage does not require synaptic complex assembly. Studies on the *Himar1* transposon have shown that cleavage at the 5' ends occurs usually two nucleotides inside the transposon DNA and cleavage at the 3' ends takes place precisely at the end of the transposon

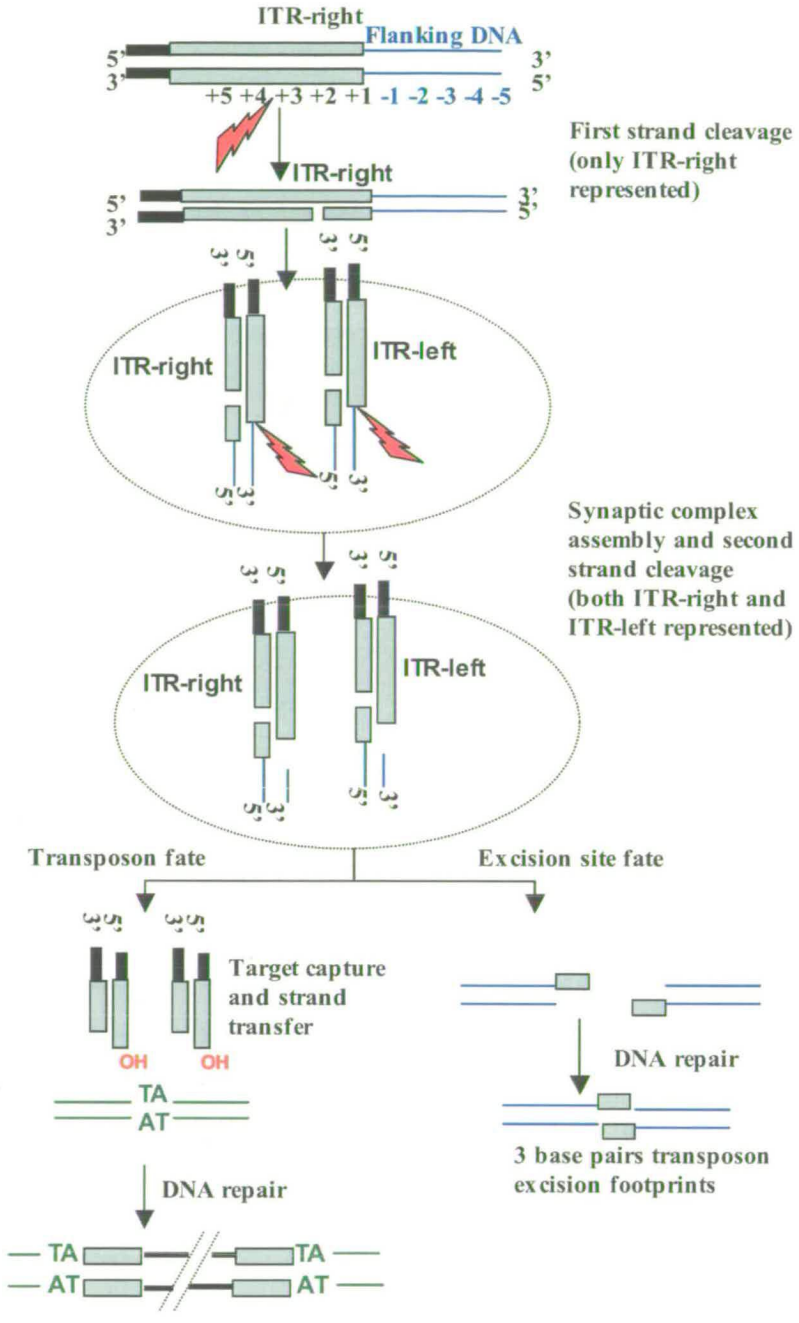

Figure 1.9. Mechanism of *Mos1* excision

Figure 1.9. Mechanism of *Mos1* excision

First strand nicking in the case of *Mos1* occurs outside the synaptic complex, on the 5' end of each strand of the transposon DNA (only the right ITR is represented for this step of the reaction). The nicking occurs 3 nt inside the transposon DNA. Second strand nicking occurs inside a synaptic complex precisely at the junction between the transposon and the flanking DNA (both transposon ends are represented for this reaction step). At the site of transposon excision, footprints are generated containing three bases from either the right or the left end of the transposon. The 3' OH groups formed upon second strand nicking are used in the strand-transfer step. *Mos1* as all members of the *IS630/Tc1/mariner* superfamily, inserts at a TA dinucleotide, which is duplicated upon insertion. Transposon DNA represented as thick black lines. Transposon ITRs are represented as grey rectangles (for the exact sequence of *Mos1* ITRs see **Figure 1.8**) Flanking DNA represented as thin blue lines. The DNA nicking steps are represented as red flashes. The target site is represented as green lines with the target TA dinucleotide.

(Lampe *et al.*, 1996; Lipkow *et al.*, 2004b). Also it appears that transposase-DNA complexes assembled on a single end are capable of nicking (Lipkow *et al.*, 2004b).

Mos1 transposase monomers have been shown to interact with each other (Lohe *et al.*, 1996b; Zhang *et al.*, 2001). A yeast two-hybrid assay has been employed to identify the regions in the transposase required for this interaction. The study found that residues located along the length of the protein are required, and there is not a distinct domain responsible for protein oligomerisation (Zhang *et al.*, 2001). This property of the transposase is extremely important. The transposase needs to cut both transposon ends to release the element from the donor DNA. The two ends need to be brought together, and this is done through an interaction between transposase subunits resulting in PEC formation. Mutant protein monomers with an impaired ability to interact are also defective in transposition (Zhang *et al.*, 2001). Subsequently, the Paired End Complex (PEC) containing the two transposon ends bridged together by transposase monomers has been visualised by Electrophoresis Mobility Shift Assay (EMSA) (Dawson and Finnegan, 2003). The formation of this complex requires Mg^{2+} or Mn^{2+} ions, but Ca^{2+} does not support its formation (Dawson and Finnegan, 2003). This is different from all other systems studied so far, in which PEC formation occurs in the absence of divalent metal ions, with Ca^{2+} stabilizing the complex but not supporting catalysis. One important characteristic of *Mos1* transposition is that the first strand cleavage occurs outside the PEC. This distinguishes *Mos1* from the bacterial elements Tn*10* and Tn*5*,

which first form a PEC and only then cleave the DNA, and relates it to the V(D)J recombination process that cleaves the first strand before PEC formation.

The cleavage at the 3'ends occurs inside the PEC and is strictly dependent on its formation. The DNA is cleaved precisely at the junction between the transposon and the flanking DNA generating a 3'OH group subsequently involved in the strand transfer reaction. This strand is referred to as the transfer strand, whereas the strand containing the 5' phosphate group is referred to as the non-transferred strand. Different transposons cleave the second strand in different ways. In the Tn*10* and Tn*5* systems, a hairpin intermediate is formed at each end of the transposon, whereas in V(D)J recombination the RAG1/RAG2 complex makes a hairpin on the flanking DNA. Indirect evidence suggests that Tam*3* transposon from *Antirrhinum majus* uses a similar mechanism for second strand cleavage as the RAG complex (Craig *et al.*, 2002). In the case of *Mos1*, since the 3'OH group generated by first strand cleavage is formed in the flanking DNA, one would expect the hairpin to be formed at this level. No hairpin intermediate could be detected. Moreover, a hairpin intermediate was ruled out by using synthetic DNA substrates assembled pre-nicked at the position of first strand cleavage, in which the deoxyadenosine was replaced with a dideoxyadenosine at the 3' end of the non-transferred strand. If the 3'OH group was required for second strand cleavage, a pre-requisite for a hairpin-forming mechanism, no second strand cleavage would be expected with the dideoxyadenosine substrate, which was not the case. The results indicate that second strand cleavage occurs in the absence of the 3'OH group (Dawson

and Finnegan, 2003). Further work is required to investigate how the second strand is cleaved in both *Himar1* and *Mos1* transposons.

Bacterial cut-and-paste transposons interact with the target DNA at defined points in the reaction, either before excision, as is the case for Tn*7*, or after excision, as for Tn*10* (Craig *et al.*, 2002; Sakai *et al.*, 1997). Studies on *Himar1* have shown that this transposon has a more "promiscuous" behaviour, interacting with the target DNA before or after transposon excision (Lipkow *et al.*, 2004a). Target interactions occur in two steps, a noncovalent interaction (also termed target capture), and a covalent interaction or strand transfer. In the above-mentioned study (Lipkow *et al.*, 2004a), *in vitro* transposition reactions were assembled with linear DNA representing the transposon ends and plasmid DNA as target. Although the reactions contained $CaCl_2$, this divalent metal ion did not support the cleavages at the two transposon ends, but catalysed the strand transfer step when the transposon ends were pre-cleaved. This feature is different from Tn*10* where none of the catalytic steps are catalysed by Ca^{2+} ions, but resembles the *Mu* and V(D)J systems where strand transfer occurs in the presence of Ca^{2+} (Hiom *et al.*, 1998c; Sakai *et al.*, 1997; Savilahti *et al.*, 1995). The different metal ion requirements at DNA cleavage and strand transfer steps suggest a different architecture of the active site at these two stages in the transposition reaction.

Thus, although the mechanism of *mariner* elements transposition only recently started to be revealed, important differences from other transposition systems are apparent.

First strand cleavage at the 5' end of each strand of the element and outside the synaptic complex and no hairpin intermediate for second strand cleavage are only some of these particularities. To further understand the mechanism of *mariner* transposition questions remain. Some of these are: what is the stoichiometry of *mariner* elements synaptic complex? What is the oligomeric state of the *mariner* transposase? What is the active form of *mariner* transposase? These questions will be addressed in the next chapters of this thesis.

1.5.5 Harnessing the *Tc1/mariner* transposable elements as genetic tools

So why study the transposition of *mariner* elements? Transposons are naturally occurring elements capable of moving from one chromosomal location to another. They have very simple requirements for their mobility, in many cases only a transposase protein and *cis*-acting DNA sequences, recognised by the transposase protein. These properties made possible their development as tools for various genetic applications ranging from sequencing and mutagenesis to vectors for genetic transformation and gene therapy. A good example is the *P* element, widely used for transformation in *Drosophila* (Rubin *et al.*, 1982; Spradling *et al.*, 1982). Attempts to use the P element in other organisms have failed, presumably because of host-specific factors required for its transposoition. The extremely broad range of organisms containing members of the *Tc1/mariner* superfamily of transposons, and the lack of host factors involved in their transposition, raised a great deal of interest towards their development as genetic tools.

The natural mutagenic potential of transposons has been widely used for genome-wide mutagenesis studies both *in vivo* and *in vitro* (Bellen *et al.*, 1989; Greenwald *et al.*, 1985; Horie *et al.*, 2003; Ivics *et al.*, 2004; Lamberg *et al.*, 2002; Moerman *et al.*, 1986; Spradling *et al.*, 1995). The advantage that transposons have over other chemical mutagens resides in the fact that mutations can be sequenced and cloned with relative ease using the known transposon sequence as a tag. In addition, desired elements can be introduced in the transposon, such as GAL4 enhancer trap lines (as with the *P* element) used to express genes in specific tissues (Bellen *et al.*, 1989). Nematodes, *Drosophila* and plants are only a few groups of organisms in which these methods have proved their advantages. Application of the transposon systems to mammals has been hampered until recently due to the lack of an efficient transposon active in mammalian cells. Members of the *Tc1/mariner* family of transposons have important characteristics, which make them especially amenable for such a role, including an extremely wide range of host organisms, and a relatively random pattern of insertion.

As discussed earlier (section 1.4.1.2), *SB* has been artificially reconstructed from transposon sequences present in salmonid fish genomes and has been shown to be highly active in vertebrates including mammalian cells (Ivics *et al.*, 1997; Ivics *et al.*, 2004; Luo *et al.*, 1998). In a genetic screen in mice, approximately 10,000 transposition sites have been estimated in the germ cells of the transformed mice. Approximately 24% of these new locations were mapped on a different chromosome from that bearing

the original insertion site. The remaining insertions were on the same chromosome, showing a preference for a region approximately 3 Mbp from the donor site. These properties make *SB* amenable both for genome-wide screens and for mutating a specific region in the donor chromosome (Horie *et al.*, 2003).

Recently, transposons have been developed as vectors for the transformation of a wide range of organisms with foreign DNA sequences. Using the *Hermes* and *Mos1* transposons as such vectors, significant advances have been made for insects with medical importance such as *Aedes aegypty* and *Anopheles gambiae*. These organisms are responsible for the transmission of major human diseases such as malaria and yellow fever. Both *Hermes* and *Mos1* mediated the germline insertion of constructs containing a luciferase gene under the control of specific promoters, opening new ways for controlling the spread of these diseases (Coates *et al.*, 1998a; Moreira *et al.*, 2000).

Gene therapy is another area in which transposons show great promise. Gene therapy approaches can be broadly divided according to the vector used for gene-delivery into viral and non-viral methods. The most widely used vectors are derived from viruses, such as adenoviruses, retroviruses, herpesviruses and adeno-associated virus (AAV) (Thomas *et al.*, 2003). In non-viral approaches, the transforming DNA is delivered either in a naked form or complexed with liposomes or polypeptides. The non-viral approaches are safe but have a very low efficiency in DNA integration and stable transgene expression. The viral vectors are superior in their efficiency of gene delivery

but raise serious issues regarding their safety (Ivics *et al.*, 2004). Since *SB* has been shown to transpose with a high efficiency in mammalian cells, it has generated considerable interest for its development as a safe and efficient vector for gene therapeutic purposes in humans. The main advantage of *SB* over viral vectors is its safety. Viral vectors can have two undesired effects, an immune response due to the administration of high titres of virus particles, or inactivation of genes through insertion (Check *et al.*, 2002; Kaiser *et al.*, 2003; Marshall *et al.*, 1999). It has been shown that repeated administration of *SB* DNA and/or transposase has no immunogenic potential even at high concentrations, and transposon integration is fairly random, being less likely to integrate into genes than HIV or AAV-based vectors (Luo *et al.*, 1998). Nevertheless *SB* needs to be actively delivered into cells, being unable to cross cellular membranes by itself. In principle, *SB* can be delivered into cells by any of the non-viral approaches used to date such as gene guns, electroporation, injection, and complex formation with liposomes, or alternatively, it can be incorporated into viral vectors for a higher efficiency of transformation. The proof of principle that *SB* can be developed as a vector for gene therapy, came from studies in which a two-plasmid system was used in mice. One plasmid, containing the gene encoding the human factor IX, flanked by the *SB* terminal repeats, was injected into the tail vein of mice with haemophilia, together with a plasmid expressing the *SB* transposase. The system allowed expression of the transgene at therapeutic levels and the partial correction of the bleeding disorder (Yant *et al.*, 2000). Since then, other genetic disorders such as Junctional epidermolysis bullosa in human keratynocites, tyrosinemia I, and haemophilia A in mice have been

treated with a very encouraging degree of success (Montini *et al.*, 2002; Morisato *et al.*, 1984; Ohlfest *et al.*, 2004a; Ohlfest *et al.*, 2004b; Ortiz-Urda *et al.*, 2003).

In conclusion, transposons have proved their potential for development as genetic tools and gene therapy vectors in a very wide variety of systems. It became clear nonetheless, that a complete fulfilment of their potential requires a much more detailed knowledge of both their intimate mechanism of transposition and their interaction with the host systems.

1.6 Aim of thesis

As described above, detailed information is available for many systems about a crucial intermediate in the transposition reaction - the synaptic complex, and the pathway leading to its assembly. The scope of my thesis was to study the synaptic complex and its formation for the soluble form of *Mos1* transposase. An assessment of the ability of the protein to catalyse the steps leading to synaptic complex assembly was required. The experiments described later were designed to address questions regarding the transposase forms present in solution, the active form of *Mos1* transposase and the stoichiometry of the *Mos1* synaptic complex.

Transposition is a dangerous event for the host cell, introducing breaks in the DNA. Little is known about the proteins involved in the repair of these DNA insults after

mariner elements transposition. An assay was designed to study this aspect in *Saccharomyces cerevisiae*, where the DNA repair processes are well understood.

Chapter 2. Materials and methods

2.1 Materials

2.1.1 Media

2.1.1.1 Bacterial media

Luria-Broth (LB)

1% (w/v) tryptone (Difco), 2.5% (w/v) yeast extract (Difco), 2.5% (w/v) NaCl. pH adjusted to 7.2 with NaOH.

Luria Broth agar (LB-agar)

1% (w/v) tryptone (Difco), 2.5% yeast extract (Difco), 2.5% (w/v) NaCl, 1.5% (w/v) agar. pH adjusted to 7.2 with NaOH.

Bacterial media were provided by the ICMB Media service.

2.1.1.2 Yeast media

YPD liquid

1% yeast extract, 2% peptone, 2% dextrose, pH adjusted to 7.5 with NaOH.

YPD agar

1% yeast extract, 2% peptone, 2% dextrose, 2% agar, pH adjusted to 7.5 with NaOH.

Amino-acids drop-out mix

The drop-out mix contains one gram of the following amino-acids:alanine, arginine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, glycine, isoleucine, lysine, methionine, phenylalanine, proline, serine, threonine, tyrosine, valine, plus two grams

of leucine. From this basal media, the desired amino-acids were left out. One gram of drop-out mix was added per 500 ml media.

Complete minimal media (YMM) (liquid)

This medium was used to grow yeast transformed with plasmids bearing auxotrophic markers. Per 500 ml it contains: 3.35 g Yeast Nitrogen Base without amino-acids (0.67% w/v) (Fluka-BioChemika), 1 g amino-acids drop-out mix (0.2% w/v), 1 pellet NaOH. The mixture was dissolved in 450 ml distilled water and sterilised by autoclaving. When the medium cooled down to approximately 60 °C, 50 ml of a sterile solution of 20% dextrose was added (2% w/v). When galactose was used as carbon source, the media was also supplemented with 25 ml of sterile solution of 20% raffinose.

Complete minimal media (YMM) (plates)

The CM medium containing agar was prepared as described above. Ten grams of agar were added to one litre of CM medium before sterilising (1% w/v). After the carbon source was added (2% w/v), the media was poured into plates.

2.1.2 Materials

2.1.2.1 Solutions

Acrylamide denaturing gel stock solutions

84 g urea (7 M final concentration) (Fisher Scientific); 20 ml 10X TBE solution (see below); 40% w/v acrylamide solution (19:1 acrylamide:bis-acrylamide) (Severn

Biotech Ltd) in the following amounts according to gel concentration: 50 ml for 10%, 40 ml for 8%, 30 ml for 6%; to 200 ml with dH_2O; stored at 4 °C in aluminium covered flasks.

4X activity buffer

24 ml 80% glycerol; 5 ml 1 M HEPES pH 7.5; 10 ml 1 M potassium acetate; to 50 ml with mqH_2O. Stored at –20 °C in 1 ml aliquots. Before use, 2 µl 0.5 M EDTA and 4 µl DTT was added to one aliquot of 1 ml.

Agarose gel loading buffer

30% glycerol; 0.25% bromophenol blue; 0.25% xylene cyanol.

Block solution

5 g Safeway dried milk in 100 ml PBS plus 200 ml Tween 20 (Polyoxyethylene 20 – sorbitan monolaurate) (BDH).

Column Buffer

500 mM NaCl, 20 mM HEPES. Sterilised by autoclaving.

Coomassie staining solution

1 g coomassie brilliant blue R-250 (Sigma); 250 ml methanol; 50 ml acetic acid; to 500 ml. Stored at room temperature.

Destaining solution

10% methanol; 10% acetic acid; 80% dH_2O.

Ethidium Bromide solution

10 mg ethidium bromide per mililiter.

Lithium-Tris solution

10 mM Tris pH 7.5; 100 mM Lithium acetate (Sigma). The solution was prepared fresh, and sterilised by filtration through a 0.2 μm minisart filter (Sartorius).

5 M NaCl

146 g NaCl (Fisher Scientific); to 500 ml. Autoclaved and stored at room temperature.

PEG solution

Two grams of PEG (MW = 3350) were dissolved in 2 ml Lithium-Tris solution in a pre-warmed water-bath and filter-sterilised through a 0.2 μm filter as above.

Salmon sperm carrier DNA

Sonicated salmon sperm DNA at a concentration of 10 mg/ml in water. The carrier DNA was boiled before each transformation to ensure complete denaturation and sterility.

SDS-PAGE boiling mix

1 ml stacking gel buffer; 0.8 ml 25% (w/v) SDS; 0.5 ml β-mercaptoethanol; 1 ml glycerol; 0.05% (w/v) bromophenol blue.

SDS-PAGE resolving gel buffer

18.15 g Tris; 0.4 g SDS; per 100 ml, adjusted to pH 8.9 with HCl.

SDS-PAGE stacking gel buffer

5.1g Tris; 0.4g SDS; per 100ml, adjusted to pH 6.7 with HCl.

STOP buffer

95% formamide; 20 mM EDTA; 0.05% bromophenol blue; 0.05% xylene cyanol; to 100 ml with mqH_2O. Aliquotted and stored at –20 °C.

10X TBE

157.5 g Tris; 46.3 g boric acid; 9.1 g EDTA; to 1 litre.

10x TGS

30 g Tris; 144 g glycine; 10 g SDS; per litre.

Transformation buffer pH6.5

2.31 g MOPS; 0.74 g $CaCl_2$; 0.12 g rubidium chloride (Sigma); adjusted to pH 6.5 with HCl; to 100 ml.

Transformation buffer pH7

0.23 g MOPS; 0.12 g rubidium chloride (Sigma); adjust to pH 7 with HCl; to 100 ml.

Transposase dilution buffer

50 mM Tris pH 7.5; 500 mM NaCl; 100 mM Triethanol amine (Sigma) (pH 7.5); 10% glycerol; 5 mM DTT.

Tris-Glycine buffer

10 g Tris, 47.5 g Glycine. To 3.3 litres with distilled water.

1M Tris pH 7.5

121 g Tris; to 1 litre.

2.1.2.2 Enzymes

All restriction enzymes were from New England Biolabs. Taq DNA polymerase was from Promega and Roche. Pfu DNA polymerase enzyme was from Promega.

2.1.2.3 Isotopes

γ-32P-dATP from Amersham (6000 μCi/mmol).

2.1.2.4 Plasmids

Plasmid	Description	Use	Reference
pGEM	Cloning vector with single T overhangs.	Cloning of PCR products	Promega manual
pEG202	Yeast two-hybrid vector containing the GAL10 promoter and the ADH terminator sequence.	Used as template to amplify the ADH terminator sequence by PCR.	(Serebriiskii *et al.*, 1999)
pMAL-C2	Vector for expression of MBP fusion proteins in *E. coli.*	Expression of MBP-*Mos1* transposase fusion in *E. coli.*	NEB manual
pBM125 (ycp50)	Yeast vector containing the GAL1-10 promoter, URA3 marker and ARS/CEN4.	Used to express *Mos1* transposase in *S. cerevisiae.*	(Rose *et al.*, 1987)
pBM-SV40	pBM 125 containing the SV40 nuclear localization sequence under the control of the GAL1-10 promoter	Used to direct the nuclear internalization of *Mos1* transposase in yeast.	This study.
pRS314	Yeast vector containing TRP1 marker and ARS/CEN4.	Used to construct the transposon donor plasmid Ytdp.	(Sikorski *et al.*, 1989)
pAA	Yeast vector containing the ADE2 gene under the control of ADH promoter.	Used as template to amplify the ADH-ADE2 cassette by PCR.	(Chang *et al.*, 1994)
p*Mos*	Plasmid containing the *Mos1* element cloned in the pBluescript vector.	Used as template to amplify the *Mos1* transposase ORF.	(Garza *et al.* 1991)
YtepE	Plasmid containing the SV40 NLS under the control of	Used to clone the *Mos1* transposase in frame with the	This study.

	the Gal1 inducible promoter. Also contains the ADH terminator sequences. Contains the URA3 marker.	SV40NLS.	
YtepM	Plasmid containing the *Mos1* ORF in frame with the SV40NLS under the control of the Gal1 promoter. Derived from YtepE.	Used to express the SV40NLS-*Mos1* transposase in the yeast *Saccharomyces cerevisiae*.	This study.
YtdpE	Plasmid containing the ADE2 gene under the control of the ADH promoter.	Used as template to clone the inactive *Mos1* element between the promoter and the ADE2 ORF.	This study.
YtdpM1	Same as YtdpE but containing an inactive *Mos1* element cloned between the ADH promoter and the ADE2 ORF. The element is in inverse orientation with respect to the ADH promoter.	Used as transposon donor plasmid.	This study.
YtdpM11	Same as YtdpM1 but the *Mos1* element cloned in direct orientation with respect to the ADH promoter.	Used as transposon donor plasmid.	This study.

2.1.2.5 *E. coli* strains

Name	**Genotype**	**Use**	**Reference**
DH5α	ϕ80dlacZΔM15,	Common strain used	(Hanahan, 1983)

	recA1, endA1, gyrA96, thi-1, hsdR17, (r_K^-, m_K^+, supE44, relA1, deoR, Δ9LACZYA-ARGF)U169	for cloning.	

2.1.2.6 *S. cerevisiae* strains

Name	Genotype	Use	Reference
KY117	MATa, ura3-52, trp1-Δ1, lys2-801^{am}, ade2-101, his3-Δ200	Yeast strain used to assay for *Mos1* transposition in *S. cerevisiae.*	(Butler, 1991)

2.1.2.7 Oligonucleotides

Name	Sequence	Comments
ADHT-sense	CGCGAATTTCTTATGATTTATG	Sense primer to amplify the ADH terminator sequences from pEG202
ADHT-anti-sense	AAAATACACACCGAGATTCATC	Anti-sense primer to amplify the ADH terminator sequences from pEG202
GEM-forward	GTTTTCCCAGTCACGACG	Sense primer for sequencing PCR products cloned into pGEM
GEM-reverse	CAGGAAACAGCTATGAC	Anti-sense primer for sequencing PCR products cloned into pGEM
BM-seq-sense	CTAATACTTTCAACATTTTCG	Sense primer for sequencing the SV40 NLS cloned into pBM125
BM-seq-anti-sense	GATATAGGCGCCAGCAACC	Anti-sense primer for sequencing the SV40 NLS cloned into pBM125
SV40-NLS-sense	TATGAGACCAAAGAAGAAGAGA	Sense strand sequence

	AAGGTTAGTG	of SV40 NLS
SV40-NLS-anti-sense	CACTAACCTTTCTCTTCTTCTTTG GTCTCATA	Anti-sense strand sequence of SV40 NLS
SDM-SV-NLS-sense	AACCCCGGATCTATGAGACCAAA GAAGAAGAGAAAGGTTAGTG	Used for site-directed mutagenesis
SDM-SV-NLS-anti-sense	CACTAACCTTTCTCTTCTTCTTTG GTCTCATAGATCCGGGGTT	Used for site-directed mutagenesis
Mos-Bam-ATG	CAGGATCCATGTCGAGTTTCGTG CCGAAT	Sense primer for *Mos1* ORF amplification from pMos
Mos-Bam-STOP-anti-sense	CAGGATCCTGTCATTTATTCAAA GTATTTGC	Anti-sense primer for *Mos1* ORF amplification from p*Mos*
p*Mos*-forward	ACCTCTAGAAACTGTATATATGCGT AAGAACG	Sense primer to amplify *Mos1* from pMos. Contains an XbaI site.
p*Mos*-reverse	ACCTCTAGAAACATAGGCATCCCAC AGTACG	Antisense primer to amplify *Mos1* from pMos. Contains an XbaI site.
F28074 – 33 mer	ACGGGCTCCAGTCGGTGAATA TAGAAACTATCA	Used to generate the S* substrate and as marker in the assay for first strand cleavage.
F28076 – 67 mer	GGTGTACAAGTATGAAATGTC GTTTTTTTTAAATCAAAAAAC ACGTTAAATTTTGTGGAAAAAG AAA	Used to generate the S* substrate.
F28078-70 mer	TTTCTTTTTCCACAAAATTTAA CGTGTTTTTTGATTTAAAAAAA ACGACATTTCATACTTGTA CACCTGA	Used as merker for second strand cleavage.
IRR100SGS	TTTCTTTTTCCACAAAATTTAA CGTGTTTTTTGATTTAAAAAAA ACGACATTTCATACTTGTA CACCTGATAGTTTCTATATTCA CCGACTGGAGCCCGT	Used as the sense strand to generate the S* substrate and the AS* substrate.
IRR100ASGS	ACGGGCTCCAGTCGGTGAATA TAGAAACTATCAGGTGTACAA GTATGAAATGTCGTTTTTTTTA AATCAAAAAACACGTTAAATT	Used as the anti-sense strand to generate the AS* substrate. Complementary to

	TTGTGGAAAAAGAAA	IRR100SGS.
IRR80SGS	AACGTGTTTTTTGATTTAAAAAAAA CGACATTTCATACTTGTACACCTGA TAGTTTCTATATTCACCGACTGGAG CCCGT	Used as sense strand to generate the 80 bp substrate
IRR80ASGS	ATAGAAACTATCAGGTGTACAAGT ATGAAATGTCGTTTTTTTTAAATCA AAAAACACGTTAAATTTTGTGGAAA AAGAAA	Used as the anti-sense strand to generate the 80 bp substrate.
pAA-sense	GCGCAACTGTTGGGAAGGGC	Forward primer used to amplify the ADH promoter-ADE2 ORF cassette from the pAA plasmid (product then cloned into the Ytdp construct).
pAA-antisense	CCTGATTCTGTGGATAACCG	Reverse primer used to amplify the ADH promoter-ADE2 ORF cassette from the pAA plasmid.
Mos-stop-forward	AACATGTCGAGTTTCTAACCG	Forward primer used to introduce a STOP codon into the *Mos1* ORF after the START site (in the YtdpM plasmid).
Mos-stop-reverse	CGGTTAGAAACTCGACATGTT	Reverse primer used to introduce a STOP codon into the *Mos1* ORF after the START site (in the YtdpM plasmid).

2.2 Methods

2.2.1 Manipulation of bacteria

2.2.1.1 Growth of *E. coli* bacterial cultures

E. coli cells were grown in LB media. Cells were streaked from frozen stocks on LB-agar plates and incubated overnight at 37 °C. One colony was inoculated into 5 ml liquid LB media and grown overnight at 37 °C. When higher culture volumes were needed, a fresh 5 ml overnight culture was diluted 1:1000 with the desired volume of LB. Cultures were grown at 37 °C with agitation in an orbital shaker (200 rpm). If plasmid containing cells were grown, the required antibiotic was added to the media. This was usually ampicillin at a concentration of 100 μg/ml.

2.2.1.2 Storage of *E. coli* bacterial cultures

One colony of bacterial cells was inoculated into 5 ml LB and grown overnight. Glycerol was added to the culture to a final concentration of 10% and 1ml aliquots were frozen at –80 °C. When needed, cells were streaked on LB plates with antibiotic added if required.

2.2.1.3 Transformation of plasmid DNA into competent cells

A fresh overnight culture was diluted 1:50 with LB. The culture was grown at 30 °C with shaking at 200 rev/min until mid-log phase (approximately three hours). Aliquots of 1.5ml were spun down at 8000 rpm for 2 minutes at 4 °C. Cells were resuspended in

500 µl ice-cold transformation buffer pH 7 and spun down as before. The pellet was resuspended in 500 µl ice-cold transformation buffer pH 6.5 and incubated on ice for one hour. Cells were collected by centrifugation and resuspended in 200 µl ice-cold transformation buffer pH 6.5. 100-200 ng DNA was added to the cell suspension in approximately 5 µl. Cells were incubated on ice for 30 minutes, mixing occasionally, and heat-shocked in a water-bath at 45 °C for 40 seconds. After the heat-shock, the cells were returned to ice; 1 ml of LB was added and the cell suspension was incubated at 37 °C with shaking at 200 rev/min for one hour to allow the expression of the antibiotic resistance gene. A 100 µl aliquot was plated on the appropriate medium

.

2.2.1.4 Electroporation of DNA into competent cells

One fresh overnight culture was diluted 1:50 with 50 ml LB and grown for approximately three hours at 37 °C with shaking at 200 rev/min. When the cell culture reached an OD of 0.5-0.7 at 600 nm, the cells were harvested by centrifugation at 4000 rpm for 20 minutes and washed three times with distilled water. The pellet was resuspended in the remaining liquid, and the cells were divided in 200 µl aliquots. The DNA was added to the cell suspension in a volume of 1 µl, mixed and the mixture was transferred to a BioRad *E. coli* Pulser Cuvette with 0.2 cm electrode gap. A BioRad Gene Pulser electroporator was used to apply a pulse of 200 Ω, 2.5 V, 25 µF. A time constant between 4.5 – 6.2 was obtained. One millilitre of LB medium pre-warmed at

37 °C, was added and the cells were incubated at 37 °C for one hour. An 100 µl aliquot was plated on the appropriate media.

2.2.1.5 Preparation of plasmid DNA from *E. coli* cells

A single colony was used to inoculate 5 ml LB media plus the appropriate antibiotic and grown over-night at 37 °C with shaking at 200 rev/min. Next day, the plasmids were extracted using a Qiagen miniprep plasmid kit according to manufacturer instructions. Briefly, the cells from 5 ml culture were collected by centrifugation at 8000 rpm and resuspended in 250 µl resuspension buffer with RNaseA added. An equal volume of lysis buffer was added, mixed and the proteins precipitated with 350 µl neutralisation buffer. The lysate was cleared by centrifugation and the supernatant was run through a Qiagen miniprep column. The column was washed with the provided buffer and the DNA was eluted using 10 mM Tris, 1 mM EDTA, pH 8.0.

2.2.2 Manipulation of yeast cells

2.2.2.1 Growth of yeast cells

Saccharomyces cerevisiae cells were grown in YPD medium or YPD-agar medium if no selection was necessary. When selection was required, CM medium lacking the corresponding amino-acid(s) was used. Yeast cells from a frozen stock were streaked on YPD-agar plates and incubated at 30 °C until colonies appeared. One colony was used to inoculate 5 ml YPD medium and grown at 30 °C with shaking at 200 rpm. The

culture stage was determined by measuring the OD at 600 nm. An OD_{600} of 1 corresponds to approximately 3 x 10^7 yeast cells ml^{-1}.

2.2.2.2 Storage of yeast cells

For short-term storage, yeast was kept at 4 °C, either on plates or liquid media, for up to three weeks. For long-term storage, frozen 15% glycerol stocks were used. A single colony was inoculated in 5 ml YPD (or CM media) and grown overnight at 30 °C with shaking at 200 rpm. One millilitre of culture was mixed with an equal amount of sterile 30% glycerol solution, and 1ml aliquots were stored at –80 °C.

2.2.2.3 Introduction of plasmid DNA into yeast cells

Five millilitres of media (YPD or CM) were inoculated with one yeast colony and grown overnight at 30 °C with shaking. The next day, the culture was diluted 1:1000 with the same medium and grown overnight to an OD_{600} of 0.5 – 0.7. Cells from 50 ml culture were harvested by centrifugation at 3500 rpm for five minutes. Cells were washed with 50 ml sterile water, centrifuged as above and resuspended in 25 ml sterile Lithium-Tris (LiT) solution. Two hundred and fifty microlitres of 1M DTT were added and the cells were incubated at room temperature for 40 minutes with gentle shaking. While the cells were incubating, the PEG solution was prepared as described above. Transformation tubes were set up containing 50 µl sterile LiT solution, 5 µl salmon sperm carrier DNA (10 mg/ml) and 1-5 µl transforming DNA (approximately 1 µg/µl).

After 40 minutes, the cells were collected by centrifugation at 3500 rpm for 5 minutes and re-suspended in 750 µl LiT and 7.5 µl DTT. One hundred microlitres of competent yeast cells were added to each transformation tube, mixed gently and incubated at room temperature with gentle shaking for 10 minutes. Three hundred microlitres of PEC solution were added and mixed gently. The cell suspension was incubated at room temperature for a further 10 minutes again with gentle shaking. Fifty microlitres of DMSO were added, mixed and the cells were heat-shocked at 42 °C for 15 minutes. Cells were collected by centrifugation for 20 seconds at 5000 rpm and resuspended in 1 ml YPD (or CM). After two hours at 30 °C in the rotative wheel, the cells were centrifuged for one minute at 3500 rpm and re-suspended in 100 µl 10 mM Tris pH 7.5. The suspension was incubated for 10 minutes at room temperature and the cells were plated on the appropriate selective media. Usually, after two to three days the colonies became visible.

2.2.2.4 Induction of *Mos1* transposase expression in *Saccharomyces cerevisiae*

The yeast strain KY117 (MATa, his3Δ200, lys2-801, ade2, trp1Δ1, ura3-52) was transformed with the Ytep containing the *Mos1* ORF (Ytep-M). As negative control, Ytep containing only the SV40-NLS and ADHter was used (Ytep-E). The plasmids were selected for on YMM-Ura plates with glucose as carbon source (see Materials and methods section 2.1.1.2). Five millilitres of liquid YMM-Ura were inoculated with one yeast colony containing Ytep-M or Ytep-E, and the yeast cells were grown overnight at

30 °C with vigorous shaking. The overnight culture was diluted 1000x with YMM-Ura liquid medium and the cells were grown to an OD_{600} of 0.7. The cells were washed twice with sterile water, and transposase expression was induced by resuspending the cells in YMM-Ura containing 2% galactose plus 1% raffinose, and growing them for 18 hours at 30 °C with shaking (100 rpm). Fifty microlitres of culture were run on an 8% SDS-PAGE gel and the transposase expression was detected using antibodies against *Mos1* transposase.

2.2.3 Protein techniques

2.2.3.1 Analysis of proteins by SDS-PAGE

An 8% resolving polyacrylamide gel was assembled by mixing: 1.2 ml resolving gel buffer (section 2.1.2.1), 1.3 ml 30% acrylamide solution (37.5:1 acrylamide/bis-acrylamide ratio – Severn Biotech Ltd), 2.5 ml water, 30 µl 25% ammonium persulfate solution and 3 µl TEMED. The stacking gel was assembled by mixing: 0.6 ml Stacking gel buffer (section 2.1.2.1), 0.4 ml 30% acrylamide, 1.4 ml water, 10 µl ammonium persulfate and 1 µl TEMED. The gels were poured using a MiniProtean Gel system (BioRad). Before loading, SDS-PAGE boiling mix solution (section 2.1.2.1) was added to the protein extract (1:6) and the mixture was boiled for 5 minutes. The gels were run at 1.4 V/cm through the stacking gel and 2.8 V/cm through the resolving gel.

2.2.3.2 Detection of proteins by coomasie staining

Proteins were separated on an SDS-PAGE gel as described above and stained with Coomasie staining solution (section 2.1.2.1) for one hour with gentle shaking at room temperature. The gel was destained overnight with at least two changes of the destaining solution (section 2.1.2.1).

2.2.3.3 Western blotting

Proteins were separated by SDS-PAGE as described above (section 2.2.3.2). The gel was equilibrated in Tris-Glycine buffer (TGS buffer - section 2.1.2.1) for at least fifteen minutes at room temperature. The hydrophobic polyvinylidene fluoride (PVDF) membrane was prepared according to the manufacturer's instructions (Millipore-Immobilon-P transfer membrane). The transfer system was assembled by placing two pieces of filter paper on top of one brillo pad provided by the manufacturer (New England Bio-Rad), then the equilibrated gel and the membrane. The membrane was covered with another two pieces of filter paper and the second brillo pad and assembled in the cassette. The transfer was done with a Bio-Rad Trans-Blot Cell at 20 V overnight at 4 °C or at 30 V for two hours at room temperature. The orientation of the nitrocellulose membrane was marked by cutting one corner and the membrane was blocked by incubating in Block solution (section 2.1.2.1) for one hour at room temperature with shaking. The membrane was incubated with the first antibody (diluted 1:1000 in block buffer) for two hours at room temperature or overnight at 4 °C. The unbound antibody and that bound non-specifically was washed off using PBS for one

hour at room temperature with shaking. The washing solution was changed four times. The HRP-conjugated secondary antibody (Amersham Biosciences) was diluted 1:1000 in Blocking buffer and the membrane was incubated for one hour at room temperature. The washing step was repeated as before. The detection system used was according to the ECL Western blotting detection reagents and analysis system (Amersham Biosciences).

2.2.3.4 Induction of MBP-*Mos1* transposase fusion protein expression

The *Mos1* transposase was purified from *E. coli* as a soluble fusion to Maltose Binding Protein (MBP) using the pMAL expression system (New England Biolabs (diGuana *et al.*, 1988, Maina *et al.*, 1988)). The pMAL-*Mos1* expression construct was a kind gift from Corrine Auge-Gouillou. The protein expression and purification protocol was as described by Auge-Gouillou (Auge-Gouillou *et al.*, 2000, Auge-Gouillou *et al.*, 2001) and according to New England Biolabs instructions (available online at www.neb.com) with minor modifications. The pMAL-c2x vector used contains the *malE* gene of *E. coli* coding for maltose binding protein under the control of a ptac promoter (Kellerman *et al.*, 1982). Fusion proteins produced with this vector contain the sequence-specific proteolysis site for the bovine factor X protease between the MBP protein and the protein of interest (Nagai *et al.*, 1984, Nagai *et al.*, 1987). In this way the MBP moiety can be cleaved off using factor X protease and removed by further purification.

E. coli DH5α cells were transformed with pMAL-*Mos1* using the $RbCl_2$ protocol as described in Materials and methods. The transformed cells were plated on LB-Amp plates and incubated over-night at 37 °C. One colony was used to inoculate 5 ml LB-Amp liquid media and incubated over-night at 37 °C with vigorous shaking. One millilitre of culture was added to 1 ml 65% sterile glycerol solution, mixed and frozen at –70 °C. This stock was used to re-streak one LB-Amp plate, and colonies were grown overnight at 37 °C. One colony was used to inoculate 5 ml LB-Amp liquid medium and grown over-night at 37 °C. This overnight culture was diluted 100-fold with 500 ml LB-Amp liquid media and incubated at 37 °C with vigorous shaking (300 rev/min) until it reached an OD_{600} of 0.7. At this point the expression of MBP-*Mos1* transposase was induced with 1.5 ml of 0.1 M IPTG (isopropyl-1-thio-β-D-galactoside), resulting in a final concentration of IPTG of 0.3 mM. The cells were further incubated at 30 °C with 200 rev/min shaking for a further two hours. Cells were collected by centrifugation at 5000 rpm, 4 °C for 15 min. The resulting pellet was resuspended in 1 ml column buffer (20 mM TRIS pH 7.5, 500 mM NaCl) and frozen over-night at –20 °C. Induction was assessed by SDS-PAGE electrophoresis and Coomassie blue staining (see **Figure 3.1**, lanes 2 and 3).

2.2.3.5 Extraction and purification of MBP-*Mos1* transposase fusion protein

The pellet from the procedure described in the previous section was allowed to thaw on ice, and the cells were lysed by sonication using a Sanyo Soniprep150 ultra-sonicator

apparatus equipped with a 9.5 mm probe. Pulses of 20 seconds and 10 μm amplitude were applied with 60 seconds breaks until the cell suspension became clear (usually 2 minutes of sonication). The cells, and subsequently the protein extracts, were kept on ice the whole time. The sonicated cell suspension was centrifuged at 8000 rpm, 4 °C, for 15 minutes. The supernatant represented the crude extract and was shown to contain the soluble MBP-*Mos1* transposase fusion protein by SDS-PAGE gel electrophoresis (see **Figure 3.1** lane 5). A volume of 500 μl amylose resin (New England Biolabs) was resuspended in 1ml column buffer containing 20 mM TRIS pH 7.5, 500 mM NaCl and centrifuged briefly to remove the supernatant. The equilibration was repeated twice, and the amylose resin was incubated with 1ml of crude extract for two hours at 4 °C with shaking on a rotating wheel. The resin was centrifuged briefly at 4 °C and washed three times with 1ml of column buffer. During each washing step the amylose resin was incubated for 15 minutes at 4 °C in the rotating wheel. The MBP-*Mos1* transposase fusion protein was eluted for one hour at 4 °C with 1 ml column buffer containing 10 mM maltose. Glycerol was added to a final concentration of 10% (v/v), the protein was aliquoted in 50 μl aliquots and, kept at –70 °C. One aliquot was defrosted immediately before each experiment and diluted to the required concentration using column buffer (20 mM Tris pH 7.5, 500 mM NaCl). Aliquots from each purification step were analysed by SDS-PAGE electrophoresis (see **Figure 3. 1**).

2.2.4 DNA techniques

2.2.4.1 Restriction-digestion of DNA

Digestion reactions were performed in the buffer supplied by the manufacturer (New England Biolabs). Reactions contained DNA, enzyme buffer to a final 1X concentration, BSA when required, 20 units enzyme and were assembled in mqH_2O. Digestion was carried out at 37 °C for 1-2 hours. When required the enzyme was inactivated either by incubation at high temperature (according to the New England Biolabs catalogue) or by running the reaction through a PCR purification column (Qiagen).

2.2.4.2 Agarose gel electrophoresis of DNA

Gels were prepared by dissolving agarose (Roche) in 1X TBE buffer to a final concentration of 0.8 – 1.5%. Ethidium bromide was added to the gel to a concentration of 0.5 μg/ml. As molecular weight markers, Hind III-digested λ DNA, 100 bp, 1 kb or 2 log markers (New England Biolabs) were loaded on the gel. Gels were run at approximately 10 V/cm in 1X TBE buffer and the DNA was visualised under UV light.

2.2.4.3 Recovery of DNA from agarose gels

The DNA was separated by agarose gel electrophoresis as described above. The band of interest was excised under UV light exposure using a sterile scalpel and the DNA was recovered using a Gel extraction kit (Qiagen). Briefly, the gel slice was melted by

incubating with the buffer provided by the manufacturer at 50 °C. The mixture was run through the column and the bound DNA was washed and eluted in 50 µl TE buffer (section 2.1.2.1).

2.2.4.4 Polymerase Chain Reaction

DNA amplification was done using a Hybaid PCR sprint temperature cycling system. PCR reactions were assembled containing 5 µl 10X Taq Polymerase buffer, 1 µl 10 mM dNTP mix (200 µM final concentration), 2 µl (5 pmol/µl) primers sense and antisense (0.2 pmol/µl final concentration), 1.5 µl 25 mM $MgCl_2$ (0.75 mM final concentration) and, mqH_2O to a final volume of 50 µl. Taq DNA polymerase (Promega or Roche) was added in a volume of 0.25 µl corresponding to 1.25 units. A typical amplification cycle was: 95 °C for 2 minutes; 95 °C for 1 minute - denaturation, 50 °C for 1 minute - annealing, 72 °C for 2 minutes - elongation (30 cycles); 72 °C for 10 minutes. The annealing temperature was modified according to the melting temperature of the primers pair used, and the elongation time to the length of the PCR product.

2.2.4.5 DNA sequencing

Sequencing reactions contained 1µl primer (5 pmol/µl), 2 µl Big Dye version 3.1 (Applied Biosystems) and 200 – 500 ng DNA. The reaction volume was adjusted to 10 µl with mqH_2O. Reactions were amplified using the following parameters: 1 cycle at 96 °C for 10 seconds; 25 cycles at 96 °C for 10 seconds, 50 °C for 15 seconds, 60 °C for 4

minutes; 1 cycle at 60 °C for 4 minutes. After amplification the reaction volumes were adjusted to 20 µl with mqH_2O and analysed using an automatic ABI-PRISM 3100 Avant Genetic Analyzer.

2.2.4.6 Filling of recessive DNA ends using the Klenow fragment of *E. coli* DNA polymerase I

Approximately 1 – 5 µg DNA cut with the appropriate restriction enzyme was incubated with 5 units Klenow fragment of DNA polymerase I (New England Biolabs) and 1 µl 10 mM dNTP mix in the appropriate buffer for 15 minutes at room temperature. The reaction was stopped by the addition of 1µl 0.5M EDTA and the enzyme inactivated at 75 °C for 20 minutes.

2.2.4.7 Ligation of DNA using *T4* DNA ligase

Ten microlitres ligation reactions contained vector DNA and insert DNA cut with the appropriate restriction enzyme(s) (when cloning of blunt ended fragments was required, the DNA ends were treated with the Klenow fragment of DNA polymerase I as in section 2.2.4.6) in a molar ratio of insert to vector of approximately 3:1. The DNA was incubated with *T4* DNA ligase in the buffer supplied by the manufacturer (New England Biolabs) at 16 °C overnight. Aliquots of the ligation reactions were used to transform *E. coli* as described above.

2.2.4.8 Dephosphorylation of DNA ends with Calf Intestinal Phosphatase

To prevent re-circularisation of vector DNA and to enhance the cloning efficiency, the 5' phosphate groups from the vector ends were removed using Calf Intestinal Phosphatase (CIP). Ten microlitres of vector DNA (approximately 1 μg) were incubated with 10 units CIP in the appropriate buffer (New England Biolabs buffer 3) at 37 °C for one hour. The enzyme was removed using a PCR purification column (Qiagen according to manufacturer instructions).

2.2.4.9 Addition of phosphate groups at the 5' ends of DNA using *T4* Poly-nucleotide kinase

When synthetic oligonucleotides were cloned into plasmid vectors, 5' phosphate groups were added using *T4* Polynucleotide Kinase (New England Biolabs). Double-stranded oligonucleotides were incubated with 10 units T4 PNK in the appropriate buffer at 37 °C for 30 minutes. The enzyme was removed by passing the reaction through a PCR purification column (Qiagen according to manufacturer instructions).

2.2.5 *In vitro* analysis of MBP-*Mos1* transposase

2.2.5.1 DNA substrates used for *in vitro* transposition studies

The DNA substrates used to study *Mos1* transposition reaction *in vitro* were as described (Dawson and Finnegan, 2003) and also see **Figures 3.2a** and **3.3a**.

2.2.5.1.1 Purification of single-stranded oligonucleotides by gel-electrophoresis and UV shadowing

Single-stranded oligonucleotides were boiled for 5 minutes in an equal volume of STOP buffer to ensure their denaturation and separated on a 6% denaturing polyacrylamide sequencing gel. The position of the DNA was determined under UV light on a TLC plate. The gel slice containing the DNA was excised and DNA was eluted overnight in 1X TE buffer at 65 °C.

2.2.5.1.2 DNA oligonucleotide substrate for first strand cleavage assays

The DNA substrate used for the first strand cleavage assay, and to assess the protein-DNA complexes formed *in vitro* by MBP-*Mos1* transposase with *Mos1* inverted terminal repeats will be referred to as the AS* substrate (**Figure 3.2a**). It is a double stranded oligonucleotide of 100 bp containing 70 bp from the right end of *Mos1* including the right hand 28 bp Inverted Terminal Repeat and 30 bp of flanking DNA. This substrate was generated by annealing the IRR100SGS oligonucleotide with its complement IRR100ASGS (section 2.1.2.7). To measure first strand cleavage, IRR100ASGS was 5' labelled with ^{32}P using Polynucleotide Kinase (section 2.2.4.9). After labeling of the bottom strand, 40 pmol of each oligonucleotide were mixed in 10 mM Tris, pH 8.0, 1 mM EDTA, 100 mM NaCl, boiled for 5 minutes and left to return to room temperature on the same water bath.

2.2.5.1.3 DNA oligonucleotide substrate for second strand cleavage assay

The DNA substrate used for the second strand cleavage assay will be referred to as the S* substrate (**Figure 3.3a**). The same substrate was used to assess the protein-DNA complexes formed *in vitro* by MBP-*Mos1* transposase on the *Mos1* right hand end that contained a nick at the position of first strand cleavage. IRR100SGS was 5' labelled with PNK. Forty pmol of labelled IRR100SGS were mixed with 40 pmol F28074, – a 33 mer, and 40 pmol F28076, – a 67 mer. NaCl was added to a final concentration of 100 mM. The mixture was boiled for 5 minutes and left to return to room temperature on the same water bath. This generated a double-stranded DNA substrate that was assembled pre-nicked at the position of first strand cleavage.

2.2.5.1.4 Purification of double-stranded DNA oligonucleotides by Polyacrylamide Gel Electrophoresis

The DNA substrates labelled with ^{32}P and assembled as described above were ethanol-precipitated and run on an 6% native polyacrylamide sequencing gel. The gel slice containing the DNA was cut and incubated overnight at 4 °C in TE buffer. The eluted DNA was ethanol-precipitated and resuspended in an appropriate volume of 10 mM Tris pH 7.5.

2.2.5.1.5 Quantification of radiolabelled oligonucleotide substrates

Oligonucleotides labelled with ^{32}P were quantified by scintillation counting. 0.5 µl of oligonucleotide solution were placed on a piece of DEAE filter paper. The filter was

introduced into a scintillation vial containing 2 ml scintillation liquid (EcoscintA – National Diagnostics), and the amount of radioactivity was measured using a Beckman scintillation counter (Beckman Coulter).

2.2.5.2 *In vitro* transposition reaction

In vitro transposition reactions were assembled essentially as described by Dawson (Dawson and Finnegan, 2003). Reactions contained 1X Activity Buffer (see section 2.1.2.1), 50 μg/ml BSA,), 20% (v/v) DMSO and mqH_2O to a final volume of 20 μl. The final concentrations per 20 μl reactions were, 25 mM HEPES-KOH (pH 7.5), 100 mM KCl, 10% glycerol, 0.05 mg/ml BSA, 0.25 mM EDTA, 1mM DTT, 20% DMSO. Mg^{2+} was added to a final concentration of 5 mM. The concentration of the DNA substrate was 7.5 fmol unless otherwise stated. Reactions were started by the addition of 6nM/µl MBP-*Mos1* transposase diluted in Column buffer on ice (section 2.1.2.1). The reactions were assembled at room temperature and incubated at 30 °C for various amounts of time depending on the reaction product under analysis. When the reaction products were analysed on denaturing gels, reactions were stopped by the addition of an equal volume of denaturing STOP buffer. When native gels were used to resolve the protein-DNA complexes by EMSA, 10 µl were loaded directly on the gel.

2.2.5.3 First strand cleavage with MBP-*Mos1* transposase

In vitro transposition reactions were assembled as described above. The AS* DNA substrate was used at a concentration of 7.5 fmol. The DNA substrate was labelled with

^{32}P at the 5' end of the bottom strand (see **Figure 3.2a**). Reactions were incubated at 30 °C for the times indicated. On the same gel a 33 nt single-stranded oligonucleotide was used as marker (M). The sequence of this oligonucleotide is identical to the cleavage product generated with the AS* substrate and is presented in section 2.1.2.7 (F28076-33mer). Before loading, the gel was pre-run at 55 W (23 mA, 2 V/cm) until the temperature of the gel reached 55 °C. The samples were boiled for 3 minutes with an equal volume of STOP buffer then loaded. The gel was run at 45 W until the samples entered the gel, then at 35 W for the rest of the run. In this way the temperature of the gel was maintained at 50 °C for the entire duration of the run. The gel was run until the bromophenol blue present in the STOP buffer reached the bottom of the gel, and the gel was dried at 80 °C using a Bio-Rad gel drier. An X ray film was exposed to the gel for approximately 16 hours (overnight) at –70 °C over an intensifying screen.

2.2.5.4 Second strand cleavage with MBP-*Mos1* transposase

The reaction conditions were essentially as described above with a few exceptions. The DNA substrate in this case (the S* substrate) was assembled pre-nicked at the position of the first strand cleavage as described and was labelled at the 5'end of the top strand (see **Figure 3.3a**). The DNA concentration was also 7.5 fmol unless otherwise stated. Reactions were incubated at 30 °C for 120 minutes and the reaction products were analysed as described above.

2.2.5.5 EMSA of MBP-*Mos1* transposase-DNA complexes

In vitro transposition reactions were assembled as described. At the desired time points, 10 μl aliquots were removed and loaded directly on a native polyacrylamide gel. Gel concentrations ranging from 4% to 10% were used. Before loading, the gels were pre-run in the cold room for one hour at 150 V. In one empty lane, only Loading buffer (section 2.1.2.1) was used to estimate the distance migrated by the protein-DNA complexes. Gels were run at 150 V for approximately five hours in the cold room, dried, and exposed to X ray films as described above.

2.2.5.6 Gel-filtration

A Superdex 200 HR column was equilibrated with 10 mM Tris, pH 7.5; 500 mM NaCl buffer, and 200 µg MBP-*Mos1* transposase (250 µl of 0.8 mg/ml) were loaded on the column. The protein was run through the column at a flow rate of 0.250 ml / minute and eluted with one column volume in 0.5 ml fractions. When the fractions resulting from gel-filtration were analysed for DNA-binding and cleavage, they were added to the reactions in less than one hour from their collection. The protein fractions were kept all the time on ice.

When the Superdex 300 column was used, it was equilibrated with a buffer containing 10 mM Tris pH 7.5 and 200 mM NaCl or 500 mM NaCl. The protein was run through the column at a flow rate of 0.5 ml/minute. Fractions of 2 ml were collected. The K_{av} values were determined using the formula $K_{av} = (V_t - V_o) / (V_c - V_o)$ where, V_t = the

elution volume, V_0 = the void volume, V_c = the column volume. This was determined by measuring the distance at which each protein peak comes off the column measuring from the start of the run. This distance was then divided by the width of one fraction on the elution profile. The resulting value represents the number of fractions until the respective protein peak comes off the column, and was multiplied by the fraction volume to give the elution volume V_t.

V_o = the void volume. This was 7.7 ml for the S200 column, and 36.7 ml for the S300 column.

V_c = the column volume. This was 22.5 ml and 120 ml for the S200 and S300 column, respectively.

The K_{av} values thus determined were plotted against logMw of the marker proteins using the Excel programe. The resulting graphs were used to determine the molecular weight of the MBP-*Mos1* peaks eluting from the S200 and S300 columns.

Chapter 3. Biochemical studies on the *Mos1* transposase

3.1 Introduction

The cut-and-paste mechanism of transposition requires both strands of the donor DNA to be cleaved at the two transposon ends. These cleavages occur in a sequential manner, with the strand being cleaved first depending on transposon and, having important consequences in the subsequent steps. For Tn*10,* the 3' end of each strand of the transposon is cleaved first, generating a free 3' OH group (Bolland and Kleckner, 1995). The first strand cleavage occurs precisely at the junction between the transposon and the flanking DNA (Benjamin and Kleckner, 1989; Benjamin and Kleckner, 1992). The same mechanism is employed by another bacterial transposon, Tn*5* (Bhasin *et al.*, 1999). In the V(D)J recombination process the RAG1 and RAG2 complex cleaves first at the 5' ends of the RSS (corresponding to the transposon DNA) (McBlane *et al.*, 1995; vanGent *et al.*, 1997). This closely resembles the first strand cleavage in the case of *Mos1,* which also occurs at the 5' end of each strand of the element (Dawson and Finnegan, 2003). Primer extension analysis has shown that *Mos1* transposase cleaves three bases inside the transposon ITR at the 5' ends (Dawson and Finnegan, 2003). This is in good agreement with the *Mos1* footprint observed *in vivo* and representing the duplicated target TA dinucleotide flanking (most frequently) three bases from either the left or right ITR (Bryan *et al.*, 1990). The excision step in the transposition reaction also requires the cleavage of the second strand. Different transposons have evolved different

means of dealing with this problem (Turlan *et al.*, 2000). The bacterial transposon Tn*10* cleaves both DNA strands in the context of a higher-order protein-DNA complex – the synaptic complex (Sakai *et al.*, 1995). The cleavage of the first strand generates a free OH group that cleaves the second DNA strand and forms a hairpin structure (Bolland and Kleckner, 1995; Kennedy *et al.*, 1998). The same hairpin mechanism for second strand cleavage is used by another bacterial transposon, Tn*5* (Bhasin *et al.*, 1999). A variant of the same mechanism is involved in the V(D)J recombination process. In this case, the hairpin is formed on the coding end (corresponding to the flanking DNA for a transposon) (McBlane *et al.*, 1995). *Mos1* has also been shown to undergo second strand cleavage inside a higher order nucleo-protein complex termed synaptic complex or Paired End Complex (PEC) (Dawson and Finnegan, 2003). The cleavage does not involve a hairpin intermediate (Dawson and Finnegan, 2003). The mechanism of *Mos1* excision was determined using the transposase purified through a denaturation-renaturation process (Dawson and Finnegan, 2003; Zhang *et al.*, 2001). This raises a few questions. Transposases are proteins that need to perform a rather precise function in cleaving and rejoining DNA. This presumably requires the amino acids forming and surrounding the active site to have well defined positions. To what extent is the precise architecture of the protein remodelled after denaturation? How does this influence the excision pathway described *in vitro*? An answer to these questions requires the purification of *Mos1* transposase in its soluble form. The decision to fuse the transposase to the Maltose Binding Protein from *E. coli* was driven by the fact that the different size of the fusion compared to the native transposase would permit me to

investigate the number of transposase subunits inside different transposase-DNA complexes. Fusions to MBP were used extensively to study the V(D)J recombination process and greatly aided in answering questions related to those posed by *Mos1* transposition.

*3.2 Purification of MBP-*Mos1 *fusion transposase*

To characterize the soluble form of *Mos1* transposase, the protein was expressed and purified from *E. coli* as a fusion to the *E. coli* Maltose Binding Protein (**Figure 3.1**). The vector expressing the fusion protein was a kind gift from Corrine Auge-Gouillou and is based on the pMAL vector from New England Biolabs. The purification procedure was as described in Materials and methods (section 2.2.3.5). Protein expression was induced with 0.3 mM IPTG for two hours at 30 °C (section 2.2.3.4). Higher temperatures and IPTG concentrations above 0.3 mM during the induction step have been found to result in a large proportion of MBP-*Mos1* transposase protein being insoluble. Usually the protein concentration in the final eluate was between 1 and 2 μg MBP-*Mos1* transposase per microliter. Taking into account the molecular weight of the fusion of 82.8 kDa (42.48 kDa MBP + 40.3 kDa *Mos1* transposase), a concentration of around 20 – 30 μM/μl MBP-*Mos1* transposase fusion protein was obtained. The protein was stored at –70 °C in Column buffer (20 mM Tris, 500 mM NaCl) supplemented with glycerol to a final concentration of 10%. Aliquots were defrosted before each

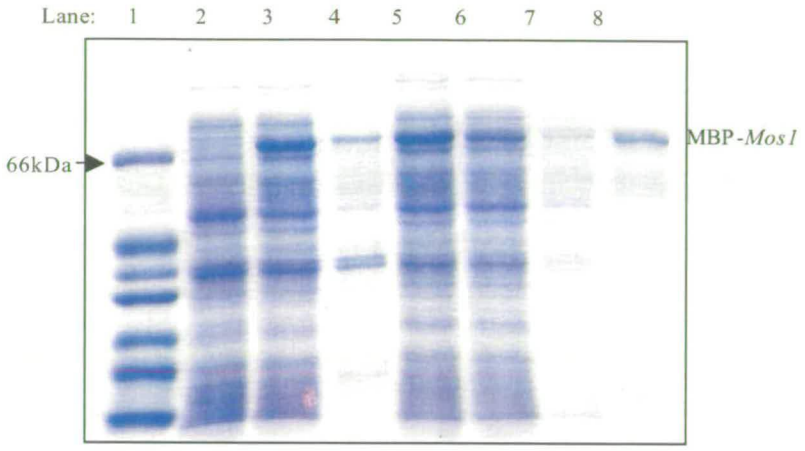

Figure 3.1. Purification of MBP-*Mos1* transposase fusion protein

Mos1 transposase was purified as a fusion to the Maltose Binding Protein from *E. coli* using the pMAL expression system (NewEngland Biolabs) (for protocol see section 2.2.3.4 and 2.2.3.5). The construct expressing the MBP-*Mos1* transposase was a gift from Corrine Auge-Gouillou. Lanes are as follows: lane 1 = molecular weight marker (Sigma VII), lane 2 = uninduced cell extract, lane 3 = induced cell extract, lane 4 = insoluble fraction, lane 5 = soluble fraction, lane 6 = flow through, lane 7 = wash, lane 8 = eluate.

experiment and diluted to the desired concentration with Column buffer on ice. When Factor X protease was used to cleave off the MBP moiety of the fusion, this resulted in the transposase becoming very unstable and quickly losing its activity.

*3.3 DNA cleavage activity of MBP-*Mos1 *transposase*

To use the MBP-*Mos1* fusion to address questions regarding the mechanism of *Mos1* PEC formation required the assessment of protein's ability to undergo all the steps involved in transposition. The steps studied were: first strand cleavage and second strand cleavage. In the next section the formation of transposase-transposon ITR complexes was assessed.

3.3.1 First strand cleavage with MBP-*Mos1* transposase fusion protein

The ability of MBP-*Mos1* fusion transposase to catalyse first strand cleavage was assessed *in vitro* by incubating the protein with ^{32}P-labelled oligonucleotide DNA substrates as previously described (Dawson and Finnegan, 2003). The DNA substrate consisted of 70 bp of *Mos1* sequence including the right ITR and 30 bp of flanking sequence. The anti-sense strand (bottom strand) was labelled with ^{32}P at the 5' end as described in Materials and methods (section 2.2.4.9). The sense and anti-sense DNA strands were annealed, generating an 100 bp double-stranded linear DNA substrate labelled at the 5' end of the anti-sense strand (**Figure 3.2a**). This substrate is referred to as AS* (anti-sense labelled substrate). The AS* substrate was incubated with 6 nM of

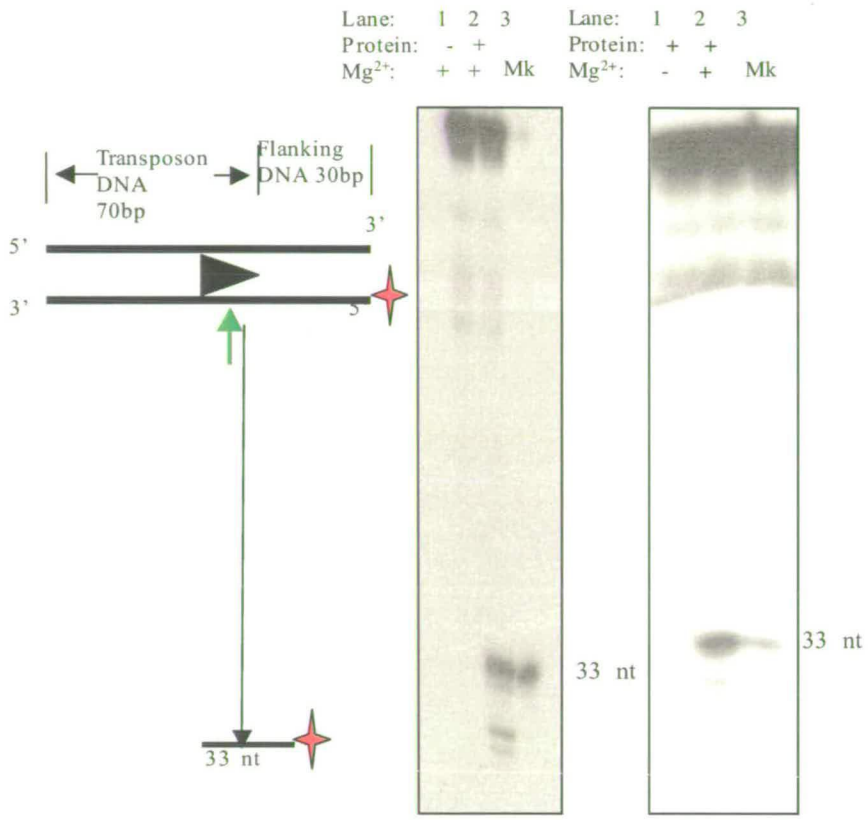

Figure 3.2a Figure 3.2b Figure 3.2c

Figure 3.2. First strand cleavage activity of MBP-*Mos1* transposase fusion protein

Figure 3.2a. DNA oligonucleotide substrate used for 1st strand cleavage studies and the size of the expected cleavage product. Red star = ^{32}P label. Green arrow = site of first strand cleavage.

Figure 3.2b. Dependence of 1st strand cleavage upon MBP-*Mos1* transposase. In lane 1 no protein was added and the reaction compared with the cleavage product generated in a reaction containing 6 nM MBP-*Mos1* transposase (lane 2). The F28074 oligonucleotide (section 2.1.2.7) representing the bottom strand of the flanking DNA plus 3 bases of *Mos1* sequence was ^{32}P labeled and used as a marker (Mk-lane 3).

Figure 3.2c. Dependence of 1st strand cleavage upon Mg^{2+}. In one reaction $MgCl_2$ was added to a final concentration of 5 mM (lane 2) and, as control in the second reaction mqH_2O replaced the $MgCl_2$ (lane 1). The same 33 nt was used as marker (Mk-lane 3).

MBP-*Mos1* transposase in a standard reaction (see Materials and methods section 2.2.5.2). The transposase was added last and reactions incubated at 30 °C for one hour (one control reaction was assembled that contained all the components except the transposase protein – **Figure 3.2b**). Reaction products were separated on an 8% denaturing PAGE gel. To determine the position of cleavage, a single-stranded DNA oligonucleotide of 33 nt (F28074-33 mer, section 2.1.2.7) was

labelled with ^{32}P and run alongside the lane containing the cleavage reaction. This marker was identical with the cleavage product generated if MBP-*Mos1* cleaves its substrate at the 5' end, 3 bases inside the ITR. The results are presented in **Figure 3.2b and 3.2c**. The cleavage activity is dependent on the presence of the transposase protein. When transposase is added, in the absence of Mg^{2+} ions, no cleavage product is detected. When $MgCl_2$ was added to a final concentration of 5 mM, cleavage products accumulate: a predominant species which comigrates with the 33 nt marker oligonucleotide and a number of cleavage products with higher mobility.

3.3.2 Second strand cleavage with MBP-*Mos1* transposase

I have assessed the ability of MBP-*Mos1* transposase to cleave the second strand (top strand) of an oligonucleotide DNA substrate *in vitro*. For this, oligonucleotide substrates were used as described (Dawson and Finnegan, 2003). In order to assess the second strand cleavage independent of first strand cleavage the substrate was assembled

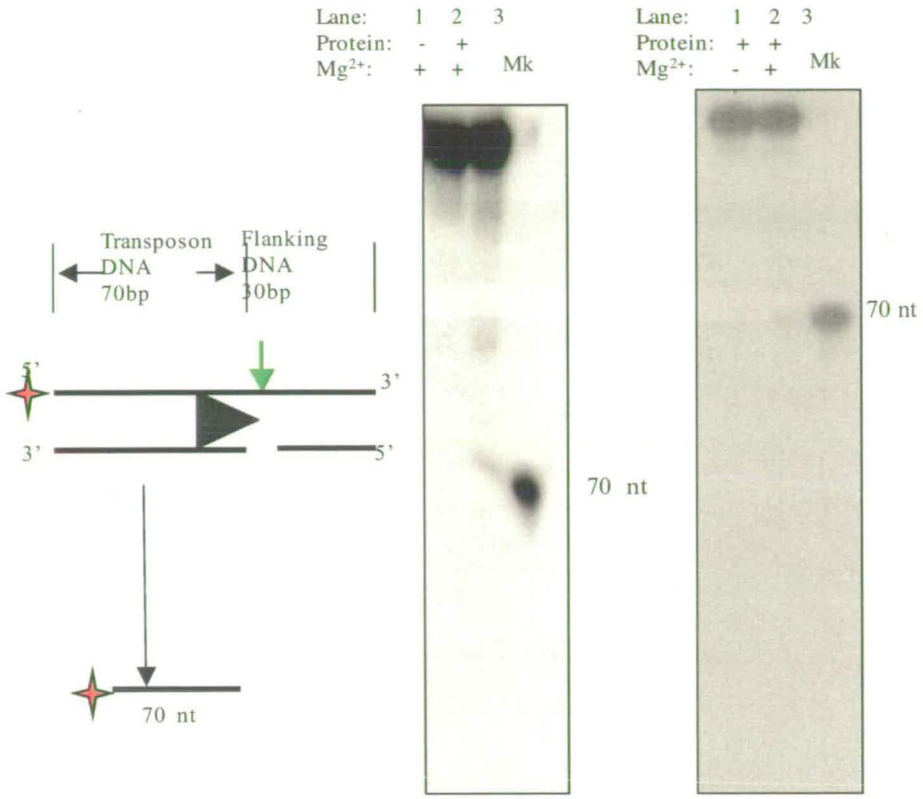

Figure 3.3a **Figure 3.3b** **Figure 3.3c**

Figure 3.3: Second strand cleavage activity of MBP-*Mos1* transposase fusion protein

Figure 3.3a. Schematic representation of the DNA substrate used to study the second strand cleavage *in vitro* and the expected 70 bases cleavage product. Red star = ^{32}P. Green arrow = the site of second strand cleavage.

Figure 3.3b. Dependence of cleavage on MBP-Mos1 transposase. The cleavage was compared between a reaction containing 6 nM MBP-*Mos1* transposase (lane 2) and a reaction where the transposase was replaced only with column buffer (see section 2.1.2.2) (lane 1). A 70 nt was used as marker (Mk-lane 3; for sequence see F28078 in section 2.1.2.7).

Figure 3.3c Dependence of second strand cleavage on Mg^{2+} ions. The cleavage was compared between a reaction containing 5 mM $MgCl_2$ (lane 2) and, a reaction with no $MgCl_2$ added (lane 1). The same 70 nt marker was used (Mk-lane 3).

pre-nicked at the position of the first strand cleavage as described in Materials and methods (Substrate S*) (**Figure 3.3a**). If second strand cleavage occurs precisely at the boundary between the transposon and flanking DNA, as previously shown for *Mos1* transposase purified by renaturation, a cleavage product of 70 nt should be generated (Dawson and Finnegan, 2003). To test this, the DNA substrate was incubated with 6nM MBP-*Mos1* transposase in a standard transposition reaction at 30 °C for 120 minutes. Reactions were stopped with an equal volume of STOP buffer (see Materials and methods section 2.1.2.1) and the reaction products were loaded on a 6% denaturing polyacrylamide gel. To determine the cleavage position, a ^{32}P labelled oligonucleotide of 70 nt with the sequence identical to the expected cleavage product was run alongside the lanes containing the reactions (F28078-70 mer, section 2.1.2.7). The results are presented in **Figures 3.3b and 3.3c**. Controls were also included in which no Mg^{2+} was added, or no transposase. In the presence of 6 nM transposase and 5 mM Mg^{2+}, a cleavage product can be separated by PAGE that runs with the same mobility as the 70 nt marker.

*3.4 Protein-DNA complexes formed by MBP-*Mos1 *transposase*

Mos1 transposase has been shown to bind sequence-specifically to *Mos1* ITRs (Auge-Gouillou *et al.*, 2000; Zhang *et al.*, 2001). Upon binding a series of protein-DNA complexes are formed, complexes that can be analysed by mobility shift assays on

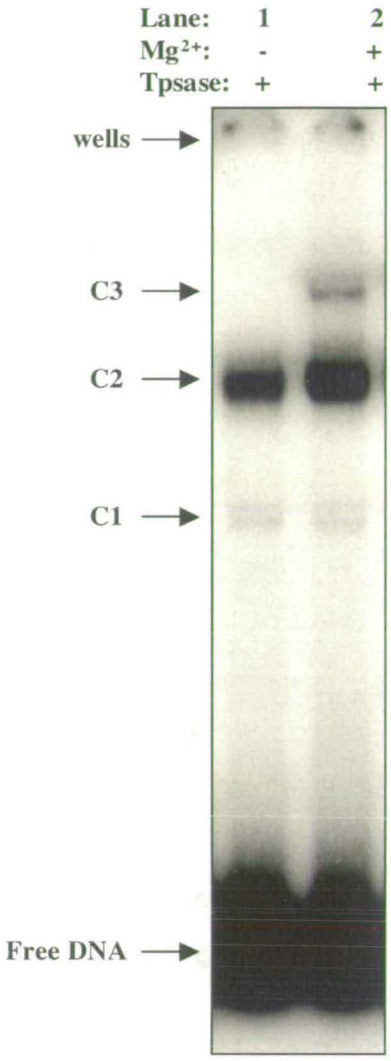

Figure 3.4: MBP-*Mos1* transposase protein-DNA complexes in the presence and absence of Mg^{2+} ions.

MBP-*Mos1* transposase was incubated with the AS* DNA substrate in the presence of 5 mM $MgCl_2$ (lane 2), or in the absence of $MgCl_2$ (lane 1). In the absence of Mg^{2+} ions, two transposase-DNA complexes are formed denoted **C1** and **C2**. In the presence of Mg^{2+} ions, apart from **C1** and **C2**, a third complex with a slower mobility is formed, denoted **C3**.

native polyacrylamide gels. To assess the protein-DNA complexes formed by MBP-*Mos1* transposase with the transposon Inverted Terminal Repeats (ITRs), the protein was incubated with substrates AS* or S* in the absence or presence of Mg^{2+} ions. Aliquots were analysed by native polyacrylamide electrophoresis. The reactions were assembled at room temperature. In one reaction, Mg^{2+} was added from a 100 mM stock of $MgCl_2$ solution to give a final concentration of 5 mM. In the other reaction only mqH_2O was added. The reactions were incubated at 30 °C for 120 minutes. Ten µl aliquots were loaded on a 6% native polyacrylamide gel with no loading dye added. In the absence of Mg^{2+}, two protein DNA complexes are visible under the experimental conditions used here: a less abundant complex with a higher mobility (**C1**) and, a more abundant complex with a slower mobility (**C2**). In the presence of Mg^{2+} ions, in addition to the same two complexes (**C1** and **C2**), a third complex with a slower mobility is visible (**C3**) (**Figure3.4**).

3.4.1 What are the MBP-*Mos1* protein-DNA complexes?

To study in more detail the complexes formed by MBP-*Mos1* transposase with the transposon ITR, the molecular weight of **C1**, **C2** and **C3** was determined by Ferguson analysis (Ferguson, 1964; Orchard and May, 1993). On native polyacrylamide gels, proteins and protein-DNA complexes migrate according to their size, shape and electrical charge. By running the protein-DNA complexes on a series of gels with increasing polyacrylamide concentrations, the influence of electrical charge on

migration can be eliminated from further calculations. In this way, the change in migration is due only to the mass and shape of the complex. The complexes are run together with protein markers of known molecular weight on a series of polyacrylamide gels and, their mobility relative to the bromophenol blue dye is plotted against the polyacrylamide concentration. The slope of each line is the retardation coefficient (K_r), which is inversely proportional to the mass of the complex. The K_r values are plotted against the molecular weight of the markers and the graph is then used to determine the molecular weight of the protein-DNA complexes (Orchard, 1993; Tavakoli, 1997).

A series of polyacrylamide gels with 5, 6, 7, 8, 9 and 10% acrylamide were prepared and kept overnight at 4 °C. Prior to loading, the gels were pre-run for 60 minutes at 150 V in the cold room. An *in vitro* transposition reaction was assembled at room temperature containing 1X Activity Buffer (see Materials and Methods, section 2.2.5.2), 50 μg/ml BSA, 900 fmol DNA substrate (substrate AS*), 20% (v/v) DMSO and mqH_2O to a final volume of 120 μl. The mixture was divided in two and, to one half, $MgCl_2$ was added to a concentration of 5 mM. Reactions were started by adding 6 nM MBP-*Mos1* transposase and incubated at 30 °C for 2 hours. Ten µl aliquots were loaded on the polyacrylamide gels. A series of non-denatured protein molecular weight markers (Sigma) were run on the same gels (**Table 2** in Appendix). The distance migrated by the bromophenol blue dye was measured. The part of the gels containing the marker proteins was cut and stained with Coomassie Blue. The 5% polyacrylamide gel is presented in **Figure 3.5**. For the marker protein gels, the distance migrated by

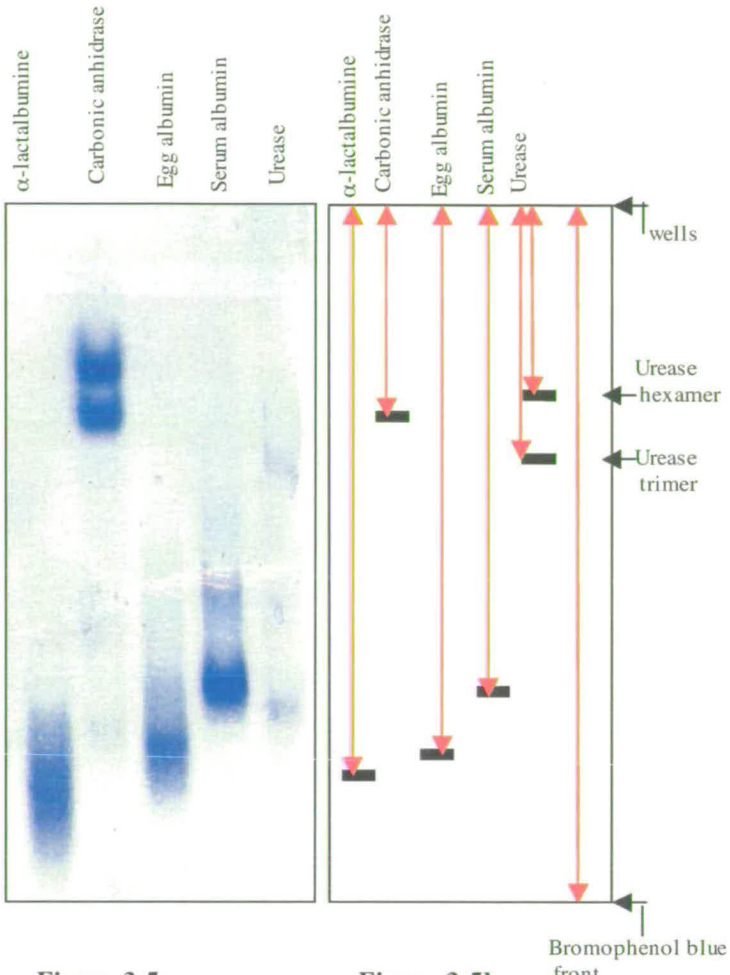

Figure 3.5a **Figure 3.5b**

Figure 3.5: Ferguson analysis - protein marker gels.

Figure 3.5a. Marker protein gel: an example of a 5% gel is shown. The proteins used are: α-lactalbumin (14, 200 Da), Carbonic anhydrase (29, 000 Da), Chicken egg albumin (45, 000 Da), Bovine serum albumin (monomer 66, 000 Da and dimer 132, 000 Da), Jack bean urease (trimer 272,000 Da and hexamer 545,000 Da).
Figure 3.5b. Diagram representing the distances migrated by the marker proteins and the bromophenol blue on the 5% gel (red lanes). These distances were used to calculate the Rf coefficients for each percentage of polyacrylamide (see text).

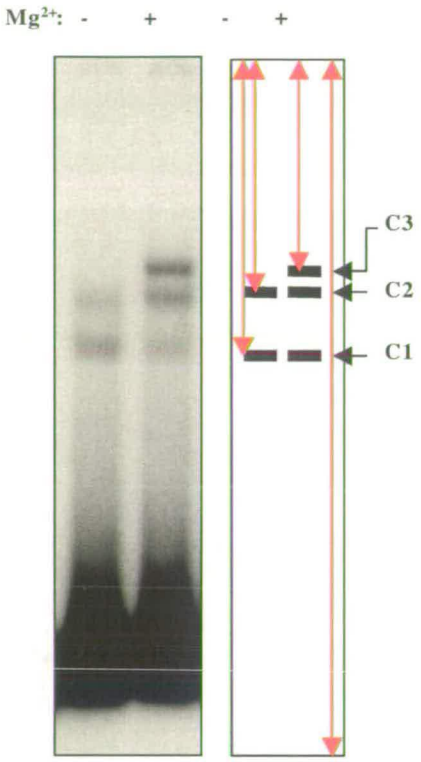

Figure 3.6a **Figure 3.6b**

Figure 3.6: Ferguson analysis – protein-DNA complexes gels.

Figure 3.6a. The same 5% polyacrylamide gel as in **Figure 3.5** showing the separation of **C1**, **C2** and **C3**. **Figure 3.6b.** diagram showing the distances migrated by **C1**, **C2**, **C3** used to calculate the Rf coefficients for MBP-*Mos1*-DNA complexes. The protein-DNA complexes were assembled *in vitro* using the AS* DNA substrate (section 2.2.5.1).

each marker protein and the bromophenol blue dye was measured (**Table 3** in Appendix). The mobility of each marker protein relative to the mobility of bromophenol blue dye (R_f) was determined by dividing the distance migrated by each marker protein to the distance of bromophenol blue migration (**Table 4** in Appendix). For the R_f values for each protein marker, logarithmic values were calculated according to the manufacturer's instructions with the formula, 100[log(Rfx100)] (**Table 5** in Appendix). The gels containing the MBP-*Mos1* transposase-DNA complexes were dried and exposed to X-ray films before measuring the distance migrated by each complex (**Figure 3.6**). To account for the change in the size of the gel due to drying, a series of correction coefficients were calculated (Table 6). The migration distance of each transposase-DNA complex was multiplied by the correction coefficient (Table 7). The corrected values were used to calculate the relative mobility of each complex (**Table 8** in Appendix) and the 100x[log(Rfx100)] (**Table 9** in Appendix). The 100x[log(Rfx100)] values were plotted against the gel concentration (**Figure 3.7**) and the slope calculated for each marker protein, and each transposase-DNA complex (Table 10). The slopes were plotted against the known molecular weights of marker proteins. The equation corresponding to the linear plot was determined using the Calculate slope command from the Excel programme, y = 18944x - 29088, where y = molecular weight and x = slope. This equation was used to determine the molecular weight of each transposase - DNA complex (**Figure 3.8**).

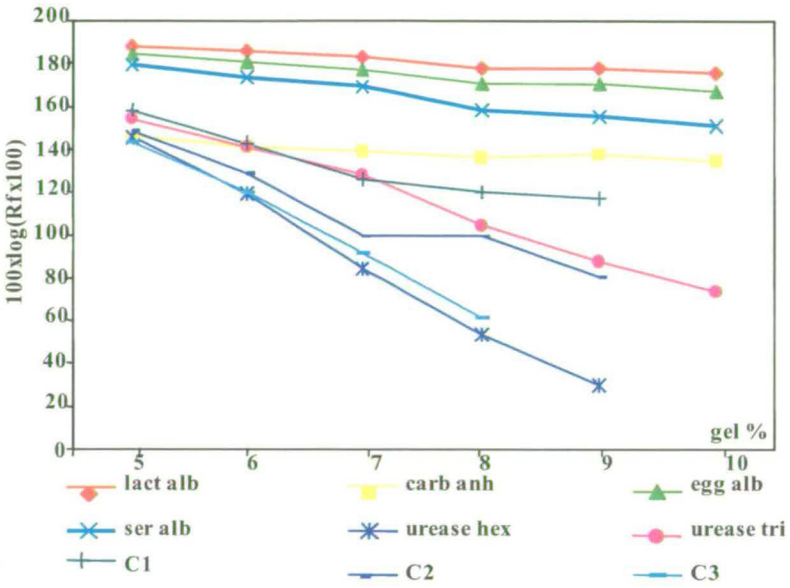

Figure 3.7: Standard slopes plot. The 100xlog(Rfx100) values from the **Tables 5** and **9** were plotted against the acrylamide gel percentages. The slope was calculated for each marker protein and protein-DNA complex.

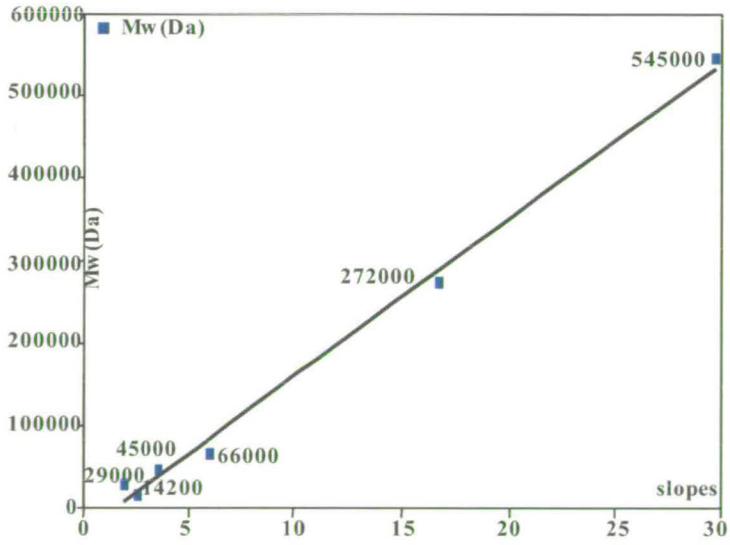

Figure 3.8: Molecular weight plot. For each marker protein, the slope was plotted against the molecular weight. The graph was used to determine the molecular weight of MBP-*Mos1* transposase-DNA complexes.

For the higher mobility complex (**C1** in **Figures 3.4** and **3.6**), MW = 18944 (10.4) - 29088 = 167930 Da. If on one hand, we consider that **C1** contains only one transposon end, by subtracting the molecular weight of the DNA substrate (649 Da x 100bp = 64900 Da), a molecular weight of 103030 Da is obtained for the protein part of C1. This value corresponds to 1.2 monomers of MBP-*Mos1* in **C1**. If on the other hand, two transposon ends are present in **C1**, 167930 Da - 2(64900 Da) = 38130 Da. For the complex with the intermediate mobility (**C2** in **Figures 3.4** and **3.6**), MW = 18944 (16.7) - 29088 = 287277 Da. A molecular weight of 222377 Da is obtained for the protein part of **C2** after subtracting the DNA contribution (64900 Da). This value corresponds to 2.7 monomers of MBP-*Mos1* transposase in **C2**. By subtracting a double value for the DNA part, 222377 Da – 129800 Da = 92577 Da. This corresponds to 1.1 monomers of MBP-*Mos1* transposase bound to two transposon ends. The **C3** complex which is formed in the presence of Mg^{2+} was assessed in the same way: MW = 18944(27.4) - 29088 = 489978 Da. This complex behaves as a Paired End Complex (as discussed later), and has been shown to contain two transposon ends (**Figure 4.1**). By eliminating the DNA contribution to the molecular weight of **C3**, 489978 Da – 129800 Da = 360178 Da. This corresponds to 4.3 monomers of MBP-*Mos1* transposase.

3.5 Influence of protein and Mg^{2+} concentrations on the transposition reaction

First I have assessed the effect of increasing transposase concentrations on all the steps studied above. This was done initially to find the optimum concentration of MBP-*Mos1* transposase to use in experiments. However, a few observations can be made regarding the effect of transposase concentrations on the transposition reaction (see also Discussions). *In vitro* transposition reactions were assembled as described in Materials and methods (section 2.2.5.2). Reactions were started by the addition of transposase to the final concentrations indicated in **Figure 3.9**. When the influence of protein concentration on first strand cleavage was investigated, the AS* substrate (section 2.2.5.1) was used and the reactions were incubated for one hour at 30 °C. Aliquots were analysed on an 8% denaturing polyacrylamide gel or on a 6% native polyacrylamide gel to assess for transposase-DNA complexes - **Figure 3.9 a** and **b**. For second strand cleavage, the S* substrate (section 2.2.5.2) was used and reactions were analysed as above - **Figure 3.9 c** and **d**. When the transposase assembles on the AS* substrate, for which the transposase has to cleave the first strand, **C3** formation increases up to 12 nM. Higher transposase concentrations are inhibitory for **C3** formation. The **C2** complex continues to increase further up to 24 nM. Above this concentration, the protein tends to aggregate in the wells. When the first strand cleavage was assessed, the 33 nt product increases up to 18 nM. As the cleavage at this position is inhibited, cleavage at additional sites is increasing generating products with a faster

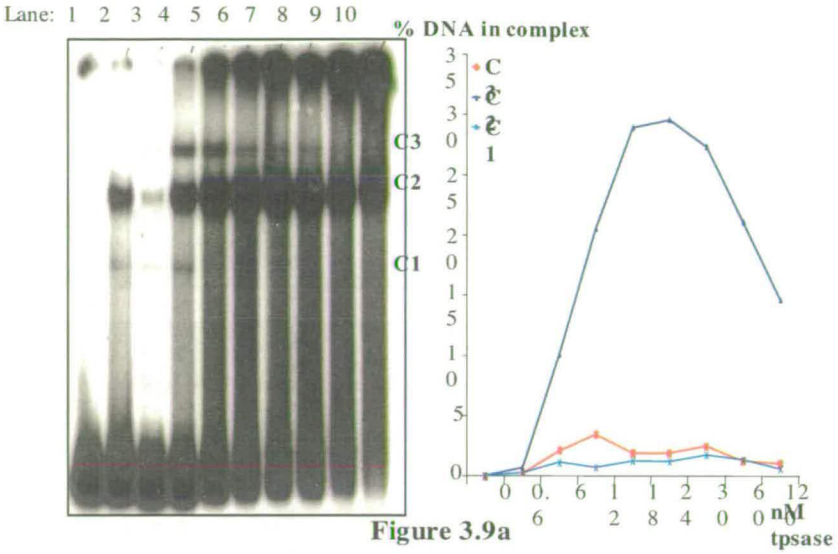

Figure 3.9a

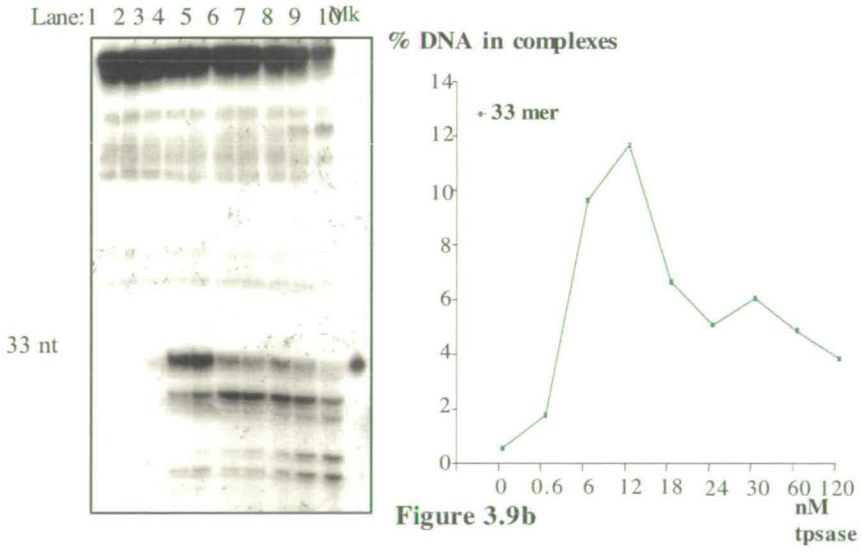

Figure 3.9b

Figure 3.9: Influence of transposase concentration on excision *in vitro* (I).

Figure 3.9: Influence of transposase concentration on excision *in vitro* (I).

Figure 3.9a. Influence of transposase concentration on transposase-transposon ITR complexes. The protein DNA complexes were assembled on the AS* DNA substrate (section 2.2.5.1) as described in section 2.2.5.2 and separated on a 6% native polyacrylamide gel. The amount of DNA substrate present in each protein-DNA complex was quantified using a STORM Phosphorimager (Molecular Dinamics) and plotted against the transposase concentration used in each lane. Lane 1 = no transposase, Lane 2 = no Mg^{2+}; Transposase concentrations: Lane 3 = 0.6 nM, Lane 4 = 6 nM, Lane 5 = 12 nM, Lane 6 = 18 nM, Lane 7 = 24 nM, Lane 8 = 30 nM, Lane 9 = 60 nM, Lane 10 = 120 nM.

Figure 3.9b. Influence of transposase concentration on first strand cleavage. Aliquots from the reactions assembled as described above were analysed on an 8% denaturing polyacrylamide gel (section 2.2.5.3). The F28074 oligonucleotide (section 2.1.2.7) was used as marker. The amount of 33 nt cleavage product was quantified as described above and plotted against the transposase concentration present in each lane. Lane 1 = no transposase, Lane 2 = no Mg^{2+}; Transposase concentrations: Lane 3 = 0.6 nM, Lane 4 = 6 nM, Lane 5 = 12 nM, Lane 6 = 18 nM, Lane 7 = 24 nM, Lane 8 = 30 nM, Lane 9 = 60 nM, Lane 10 = 120 nM.

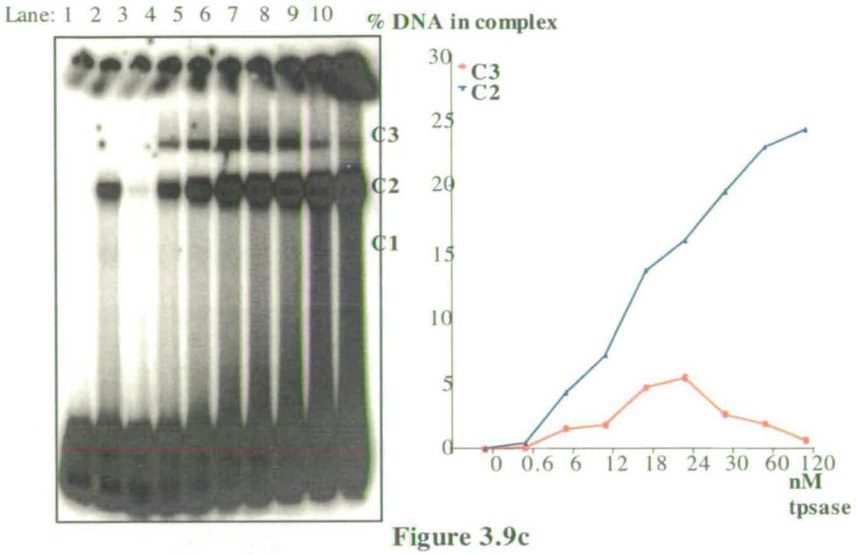

Figure 3.9c

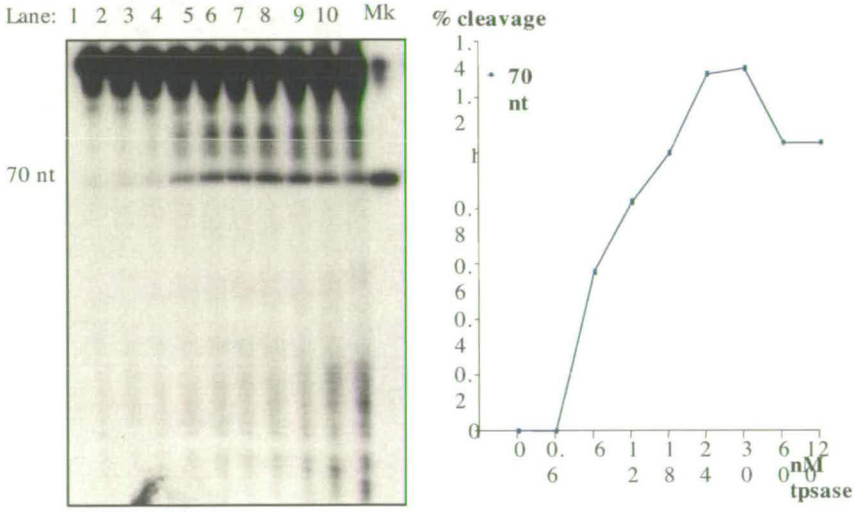

Figure 3.9d

Figure 3.9: Influence of transposase concentration on excision *in vitro* (II)

Figure 3.9: Influence of transposase concentration on excision *in vitro* (II)

Figure 3.9c. Influence of transposase concentration on transposase-transposon ITR complexes (S* DNA substrate-section 2.2.5.1). *In vitro* transposition reactions using the S* DNA substrate were assembled as described in section 2.2.5.2 and aliquots were analysed on a 6% native polyacrylamide gel. The amount of DNA substrate present in each protein-DNA complex was quantified using a STORM Phosphor imager (Molecular Dinamics) and plotted against the transposase concentration present in each lane. Lane 1 = no transposase, Lane 2 = no Mg^{2+}; Transposase concentrations: Lane 3 = 0.6 nM, Lane 4 = 6 nM, Lane 5 = 12 nM, Lane 6 = 18 nM, Lane 7 = 24 nM, Lane 8 = 30 nM, Lane 9 = 60 nM, Lane 10 = 120 nM.

Figure 3.9d. Influence of transposase concentration on second strand cleavage (S* substrate-section 2.2.5.1). Aliquots from the reactions assembled as described above were analysed on a 6% denaturing polyacrylamide gel (section 2.2.5.4). The F28078 oligonucleotide was labelled with ^{32}P and used as marker (Mk lane). The amount of 70 nt cleavage product was quantified using a STORM Phosphorimager and plotted against the transposase concentration present in each lane. Lane 1 = no transposase, Lane 2 = no Mg^{2+}; Transposase concentrations: Lane 3 = 0.6 nM, Lane 4 = 6 nM, Lane 5 = 12 nM, Lane 6 = 18 nM, Lane 7 = 24 nM, Lane 8 = 30 nM, Lane 9 = 60 nM, Lane 10 = 120 nM.

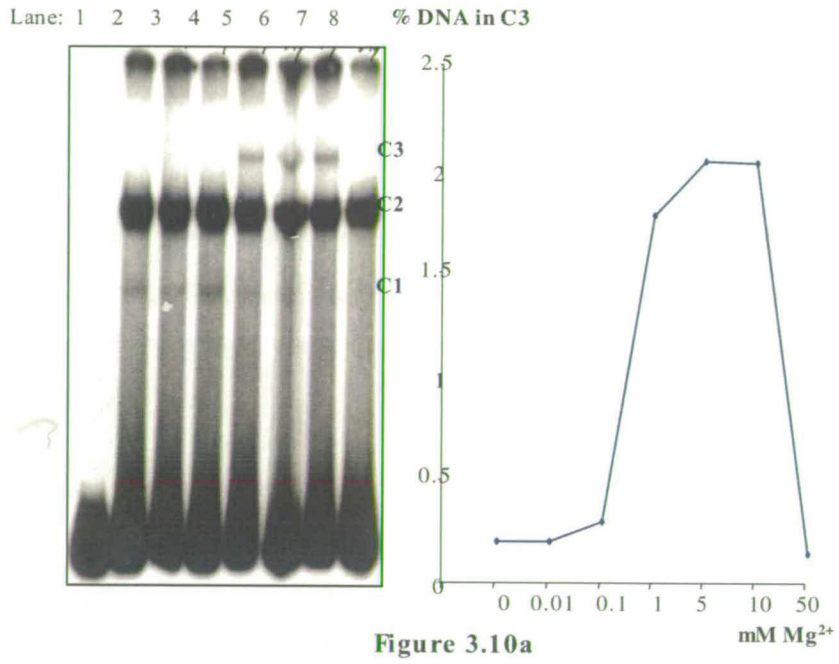

Figure 3.10a

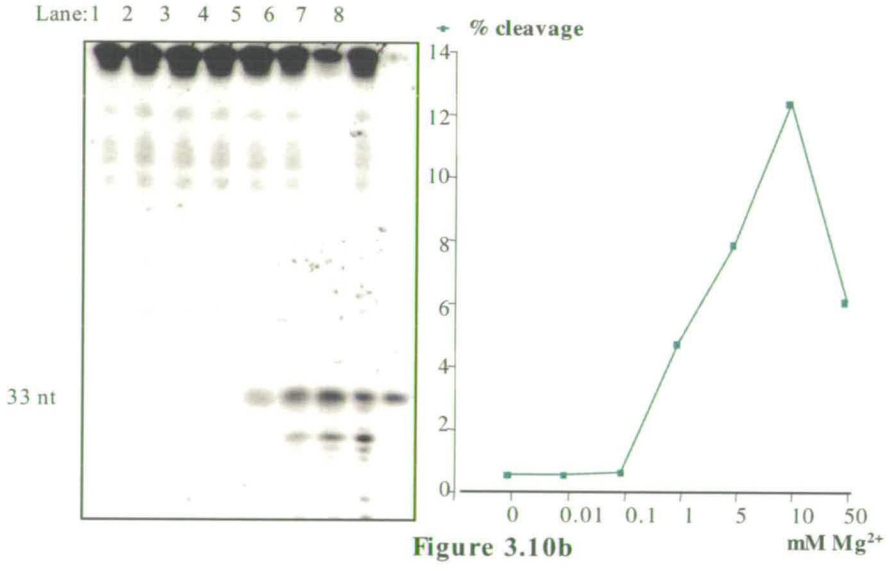

Figure 3.10b

Figure 3.10: Influence of Mg^{2+} concentration on excision *in vitro* (I)

Figure 3.10: Influence of Mg^{2+} concentration on excision *in vitro* (I)

Figure 3.10a. Influence of Mg^{2+} concentration on C3 formation (AS* DNA substrate). *In vitro* transposition reactions were assembled as described in section 2.2.5.2. Aliquots were analysed on a native 6% polyacrylamide gel (section 2.2.5.5). The amount of labelled DNA present in the C3 complex was quantified using a Storm Phosphor imager (Molecular Dinamics) and plotted against the Mg^{2+} concentrations used in each lane. Lane 1 = no transposase. $MgCl_2$ concentrations: Lane 2 = 0 mM, Lane 3 = 0.01 mM, Lane 4 = 0.1 mM, Lane 5 = 1 mM, Lane 6 = 5 mM, Lane 7 = 10 mM, Lane 8 = 50 mM.

Figure 3.10b. Influence of Mg^{2+} concentration on first strand cleavage (AS* DNA substrate). Aliquots from the above transposition reactions were analysed on an 8% denaturing polyacrylamide gel (section 2.2.5.4). The F28074 oligonucleotide (section 2.1.2.7) was labelled with ^{32}P and used as a marker (Mk lane). The amount of cleavage was quantified as above and plotted against Mg^{2+} concentrations present in each lane. Lane 1 = no transposase. $MgCl_2$ concentrations: Lane 2 = 0 mM, Lane 3 = 0.01 mM, Lane 4 = 0.1 mM, Lane 5 = 1 mM, Lane 6 = 5 mM, Lane 7 = 10 mM, Lane 8 = 50 mM.

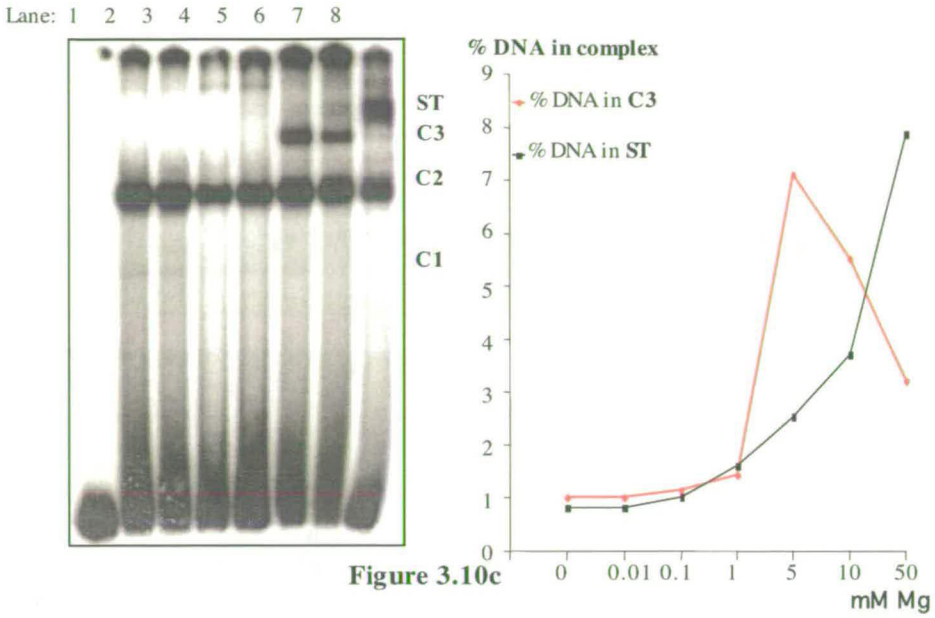

Figure 3.10c

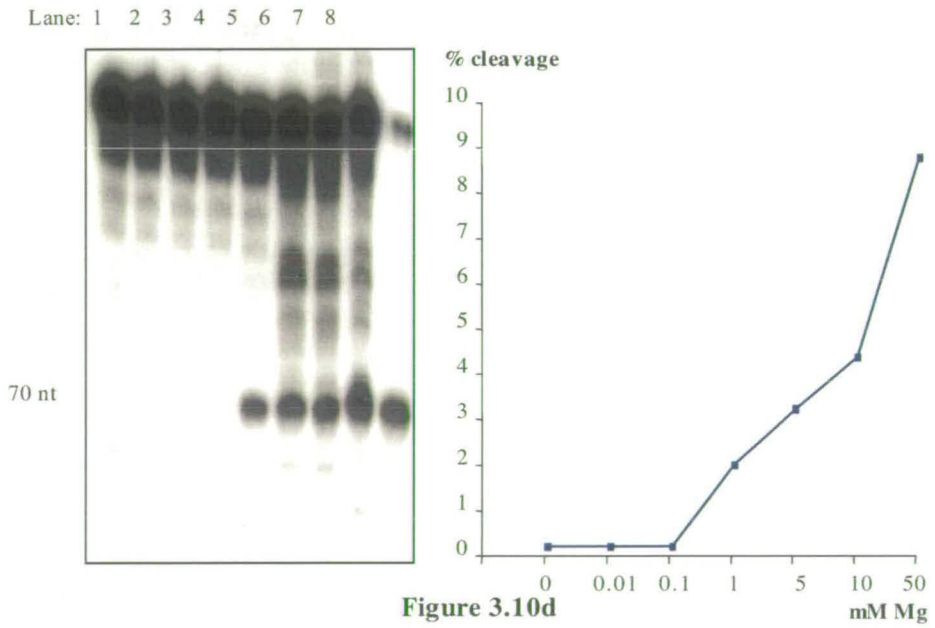

Figure 3.10d

Figure 3.10: Influence of Mg^{2+} concentration on excision *in vitro* (II)

Figure 3.10: Influence of Mg^{2+} concentration on excision *in vitro* (II)

Figure 3.10c. Influence of Mg^{2+} concentration on C3 (Paired End Complex - PEC) formation (S* DNA substrate). *In vitro* transposition reactions were assembled as described in section 2.2.5.2 and aliquots were analysed on a 6% native polyacrylamide gel. The amount of IDNA substrate present in the **C3** and Strand Transfer (**ST**) complexes (see Discussions for explanation of the **ST** complex) was quantified using a STORM Phosphorimager (Molecular Dinamics) and plotted against the Mg^{2+} concentrations present in each lane. Lane 1 = no transposase. $MgCl_2$ concentrations: Lane 2 = 0 mM, Lane 3 = 0.01 mM, Lane 4 = 0.1 mM, Lane 5 = 1 mM, Lane 6 = 5 mM, Lane 7 = 10 mM, Lane 8 = 50 mM.

Figure 3.10d. Influence of Mg^{2+} concentration on second strand cleavage (S* DNA substrate). Aliquots from the reactions assembled as described above were analysed on a 6% denaturing polyacrylamide gel (section 2.2.5.4). The F28078 oligonucleotide labeled with ^{32}P (section 2.1.2.7) was used as marker (Mk lane). The amount of cleaved product was quantified as above and plotted against Mg^{2+} concentration. Lane 1 = no transposase. $MgCl_2$ concentrations: Lane 2 = 0 mM, Lane 3 = 0.01 mM, Lane 4 = 0.1 mM, Lane 5 = 1 mM, Lane 6 = 5 mM, Lane 7 = 10 mM, Lane 8 = 50 mM.

mobility in the gel. This indicates that the cleavage position shifts towards the flanking DNA and away from the transposon DNA. When the transposase assembles on the S* substrate where the first strand cleavage is bypassed, the **C3** complex continues to accumulate up to 24 nM transposase and the **C2** complex accumulates up to the highest concentration point tested –120 nM. The second strand cleavage product of 70 nt accumulates up to 60 nM transposase.

The optimum Mg^{2+} concentration was determined in the same manner. $MgCl_2$ concentrations of 0; 0.01; 0.1; 1; 5; 10 and 50 mM were tested in *in vitro* transposition reactions assembled as above. When the transposase complexes assemble on the AS* DNA substrate, the **C3** complex formation is inhibited by 50 mM $MgCl_2$. On the S* DNA substrate however, the **C3** complex becomes detectable at 5 mM $MgCl_2$. At higher Mg^{2+} concentrations, another retarded band is formed, termed here **C4**. **C4** appears to contain more than one complex and is formed on the expense of **C3**. The optimum Mg^{2+} concentration for first strand cleavage is around 10 mM $MgCl_2$ with higher concentrations being inhibitory. In contrast, second strand cleavage products continue to accumulate at 50 mM $MgCl_2$, the last concentration tested (**Figures 3.10 a, b, c** and **d**).

3.6 Discussions

The *Mos1* transposase was purified in a soluble form as a fusion to the Maltose Binding Protein from *E. coli*. The use of MBP fusions has several advantages. The fusion proteins are easy to purify in fairly large amounts, are more likely to be soluble and, the size is increased substantially by the presence of the MBP tag. This latter property is especially useful allowing the tagging of the protein for protein mixing experiments (discussed in more detail in the next chapter). The MBP-*Mos1* transposase was assessed biochemically for its ability to perform the transposition reaction *in vitro* using linear oligonucleotide substrates. In the presence of Mg^{2+} ions, MBP-*Mos1* transposase was able to nick the anti-sense strand of a linear DNA substrate containing the transposon ITR, three bases inside the transposon DNA, as previously shown for the *Mos1* transposase purified by renaturation. A series of cleavage products with higher mobility were also formed. The second strand cleavage was also dependent on the presence of Mg^{2+} ions and occurred precisely at the transposon end. The cleavage reactions were also assessed under native conditions to detect the complexes formed between MBP-*Mos1* transposase and the transposon ITR. In the absence of Mg^{2+} ions two bands with retarded mobility were observed. In the presence of Mg^{2+}, a third band with a slower mobility formed. The bands are denoted **C1**, **C2** and **C3,** with **C1** being the fastest migrating complex. In an attempt to identify these complexes, a Ferguson analysis was done to determine their molecular weights. According to this analysis, the fastest migrating complex has a molecular weight corresponding to one monomer of MBP-

5'-TTT CTT TTT CCA CAA AAT T**TA** ACG TGT TTT TTG ATT

TAA AAA AAA CGA CAT TTC A**TA** CTT G**TA** CAC CTG A**TA**

GTT TCT A**TA** TTC ACC GAC TGG AGC CCG T-3'

Figure 3.11. Location of possible TA dinucleotide insertion sites located in the IRR100SGS oligonucleotide DNA substrate.

The TA dinucleotides present in these substrates that might act as target sites for strand transfer (**ST**) are represented in bold (**TA**).

Mos1 transposase bound to one transposon end. The molecular weight of the second complex formed in the absence of Mg^{2+} suggests a dimer of MBP-*Mos1* bound to one transposon end. Under the experimental conditions used in this study the relative abundance of these two complexes was very constant with the dimer being present in a much higher amount than the monomer. In the presence of Mg^{2+} a band with a slower mobility than the dimer is formed. The molecular weight deduced from the Ferguson analysis suggests that the most likely protein to DNA ratio in this complex (**C3**) corresponds to a tetramer of transposase bound to two transposon ends. The complexes observed in these experiments correlate well with *in vitro* studies using *Mos1* transposase purified by renaturation (Dawson and Finnegan, 2003).

To determine the optimum reaction conditions, experiments using increasing transposase concentrations were done. These experiments set the standard transposase concentration used in all future experiments to around 6 nM. Higher transposase concentrations favour first strand nicking at aberrant positions generating cleavage products smaller than the expected 33 nt. This indicates that the aberrant cleavages occur towards the flanking DNA. The optimum Mg^{2+} concentration was determined to be around 5 mM with higher concentrations having an inhibitory effect on first strand cleavage and **C3** formation. In contrast, the second strand cleavage was stimulated by higher Mg^{2+} concentrations.

In the presence of increasing Mg^{2+} concentrations, a fourth band (**ST**) with a retarded mobility slower than **C3** is formed. This band seems to contain more than one protein-DNA complex having a more "smeary" appearance. As the **ST** amount increases with increasing Mg^{2+}, the **C3** complex decreases. This implies that **C3** is a precursor for the **ST** complex. Although I did not studied the **ST** complex in any detail due to time limitations, it is possible it represents strand transfer (ST) of PECs into target sites present in the free DNA substrate. The DNA substrate used in the studies presented here represents 70 bp from the right hand end of *Mos1* (the 3' end of the transposon) plus 30 bp of flanking DNA. This substrate is always present in reactions in excess over the protein and contains six TA dinucleotides (**Figure 3.11**). *Mariner* elements insert at DNA sites containing a TA dinucleotide, which is duplicated upon insertion (Bryan *et al.*, 1990). In principle any of these TA sites could act as target for strand transfer. This possibility can be assessed by purifying the **ST** complexes from native gels and analysing them by Maxam-Gilbet sequencing. In this way it can be distinguished if the integration in the target DNA took place at the predicted TA dinucleotide sites.

Chapter 4. Synaptic complex formation with MBP-*Mos1* transposase

4.1 Introduction

All transposition systems and site-specific recombination processes studied so far in greater depth, involve the formation of higher-order nucleo-protein complexes termed synaptic complexes: Tn*5* (Bhasin *et al.*, 2000; Davies *et al.*, 1999), Tn*10* (Sakai *et al.*, 1995; Sakai *et al.*, 2000), bacteriophage *Mu* (Craigie *et al.*, 1985b; Yang *et al.*, 1996; Yang *et al.*, 1995a), *IS911* (Haren *et al.*, 1998; Normand *et al.*, 2001), RAG1 and RAG2 mediated V(D)J recombination (Hiom and Gellert, 1998a; Hiom *et al.*, 1998b), bacteriophage λ (Segall, 1998), retroviral integrases (Wei *et al.*, 1997). The best-characterised synaptic complex involved in a transposition system is that of Tn*5* for which the crystal structure has been determined (Davies *et al.*, 2000). *Mos1* is the first and, at the moment the only eukaryotic transposable element for which a Paired End Complex has been identified *in vitro* (Dawson and Finnegan, 2003). This is consistent with previous findings that cleavage at *Mos1* ends is coordinated (Zhang *et al.*, 2001), and *Mos1* transposase monomers interact with each other (Lohe *et al.*, 1997; Zhang *et al.*, 2001). It is probable that other eukaryotic transposons also use a synaptic complex during their transposition. For example, *P* element excision requires the presence of both left and right ends of the element (Beall and Rio, 1997). Another member of the *mariner* family, the *Himar 1* element from the horn flies *Haematobia irritans* and the

lacewing *Chrysoperla plorabunda* has been proposed to involve the formation of a PEC, but this complex could not be directly visualised to date by EMSA (Lipkow *et al.*, 2004b). For *Mos1* second strand cleavage has been shown to be dependent on PEC formation. Indeed the PEC is detectable by EMSA using the renatured form of *Mos1* transposase, and the second strand cleavage takes place only under conditions that permit PEC formation (Dawson and Finnegan, 2003).

The dependence of **C3** formation on Mg^{2+} ions and its slower mobility suggested that **C3** might correspond to a PEC formed by the MBP-*Mos1* transposase. To address this possibility, the number of transposon ends in **C3**, the time-course of **C3** formation and the stability of **C3** were examined.

4.2 How many transposon ends are contained within C3?

Mos1 ITRs differ at three nucleotide positions. Inside a synaptic complex both ends of the transposon are present. However, the *in vitro* transposition assay used here contains only one transposon end (the oligonucleotide corresponding to the right-hand end of *Mos1*), so the PEC would contain two right-hand ends of *Mos1*. It has been shown that a *Mos1* transposon containing two right Inverted Terminal Repeats (ITR's) is able to transpose in bacteria (Auge-Gouillou *et al.*, 2001). To determine the number of transposon ends present in **C3**, a mixing experiment was used as previously described (Bhasin *et al.*, 2000; Dawson and Finnegan, 2003; Sakai *et al.*, 1995). By using two

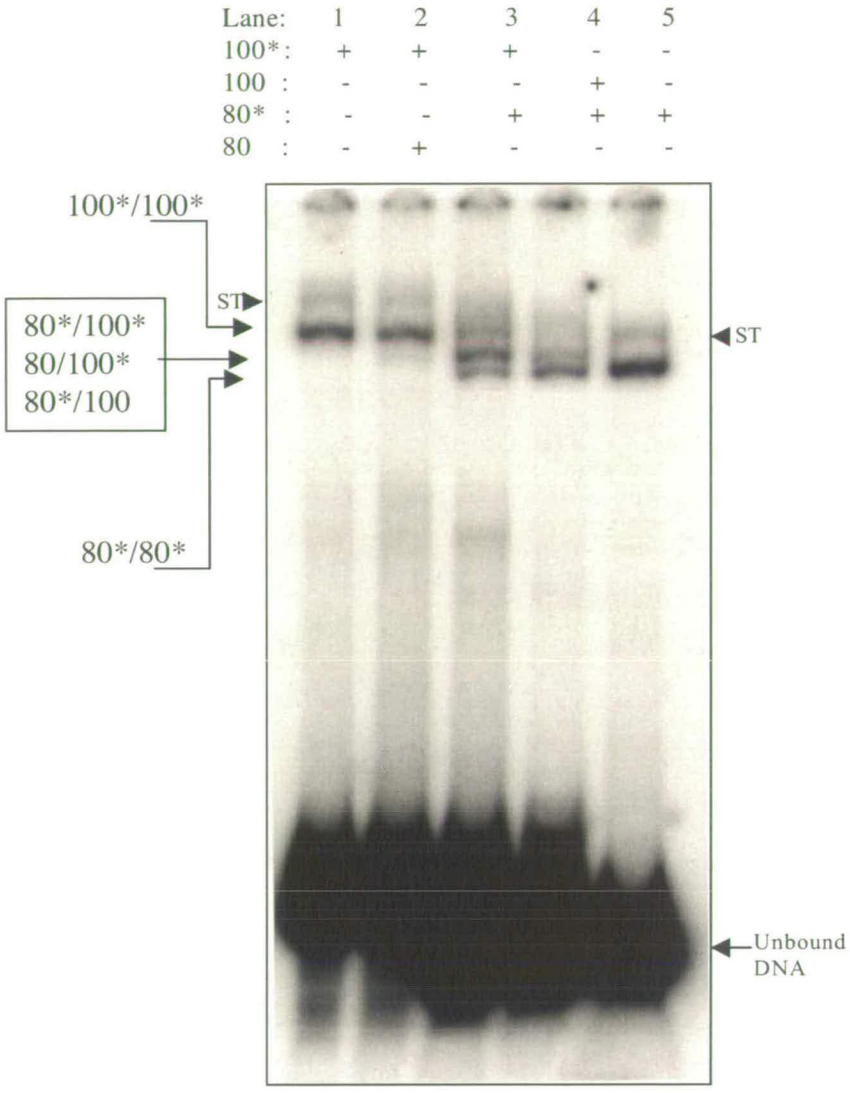

Figure 4.1: Synaptic complex formation with MBP-*Mos1* transposase

Figure 4.1: Synaptic complex formation with MBP-*Mos1* transposase

In vitro transposition reactions were assembled as described in section 2.2.5.2. Two different sized DNA substrates (100bp and 80bp) were incubated with MBP-*Mos1* under conditions promoting **C3** complex formation (5mM $MgCl_2$). An additional incubation at 45 ℃ was used to destabilize the monomer and dimer complexes (**C1** and **C2**) (see section 4.4). Aliquots from each reaction were analysed on a 6% native polyacrylamide gel (section 2.2.5.5). A number of n-1 intermediate sized bands are expected, (where n = number of DNA molecules in the complex) when both DNA substrates are labelled. The * represents the ^{32}P label. The strand transfer complexes (**ST**) are visible above each PEC complex as bands somehow less defined denoted **ST**. These probably represent strand transfer events into the unbound DNA present in the reactions. For a more detailed discussion on the **ST** products see sections 3.5 and 3.6.

DNA substrates with different sizes, n-1 (where n = number of transposon ends in the complex) intermediate complexes are expected to form. Five *in vitro* transposition reactions were assembled, as in Materials and methods (section 2.2.5.2). In the first reaction (**Figure 4.1** lane 1), 150 fmol of ^{32}P labelled 100mer oligonucleotide substrate (substrate AS*) was added. In the second reaction (**Figure 4.1** lane 2), 150 fmol of ^{32}P labelled 100 mer and 150 fmol of unlabelled 80 mer substrate were mixed. The 80 mer substrate is identical with the 100 mer substrate but lacks 20 bp of transposon DNA (section 2.1.2.7). In the third reaction (**Figure 4.1** lane 3), both 100 mer and 80 mer were labelled with ^{32}P and, 150 fmol of each were added to the reaction. The fourth reaction (**Figure 4.1** lane 4) contained 150 fmol of unlabelled 100 mer and 150 fmol of ^{32}P labelled 80 mer. The last reaction (**Figure 4.1** lane 5) contained only 150 fmol of ^{32}P labelled 80 mer DNA substrate. The reactions were incubated at 30 °C for 2 hours. In order to eliminate the monomer and the dimer complexes, that might obscure the mixed **C3** complexes formed, the reactions were incubated for a further hour at 45 °C . This incubation has been shown to destabilise the monomer and dimer complexes but not the **C3** or the **ST** complex (see section 3.6 for a discussion on the **ST** complex and section 4.4 for the effect of the 45 °C incubation on **C1**, **C2**, **C3** and **ST** complexes). The protein-DNA complexes were resolved on a 6% native polyacrylamide gel. When both 100 mer and 80 mer are present in the reaction and both are ^{32}P labelled, three **C3** complexes are formed (**Figure 4.1**, lane 3), a **C3** complex formed with two 100* DNA substrates, one **C3** complex with intermediate mobility containing one 100* and one 80* DNA substrates and, a **C3** complex with a higher mobility containing two 80*

substrates. The complexes corresponding to strand transfer events into the free DNA present in reaction are elso visible(marked **ST**). When both DNAs are present but only one of them is labelled, only the **ST** complexes and two of the three **C3** complexes are visible (**Figure 4.1**, lanes 2 and 4): the complex containing both the labelled and the unlabelled substrates and, the complex formed with the labelled DNA substrate. The complex formed only with the unlabelled substrate cannot be detected. When only one DNA substrate is present one complex is formed corresponding to the previously described **C3** and the **ST** complexes (**Figure 4.1** lanes 1 and 5). In conclusion, **C3** most likely contains two transposon ends, being a Paired End Complex formed with the soluble form of *Mos1* transposase.

4.3 When is the C3 complex formed in the course of the transposition reaction?

Three MBP-*Mos1* transposase-DNA complexes were detected (**C1**, **C2** and, **C3**). **C1** corresponds to a monomer of transposase bound to one transposon end, and **C2** corresponds to a dimer of transposase bound to one transposon end. **C3** contains two transposon ends bridged together by a number of transposase monomers, corresponding to a PEC. PEC formation is expected to occur in the later stages of the reaction after DNA binding and first strand cleavage (Dawson and Finnegan, 2003). To verify this hypothesis, the time course of **C3** formation was analysed by EMSA. An *in vitro* transposition reaction containing 1X Activity Buffer, 50 μg/ml BSA, 150 fmol DNA

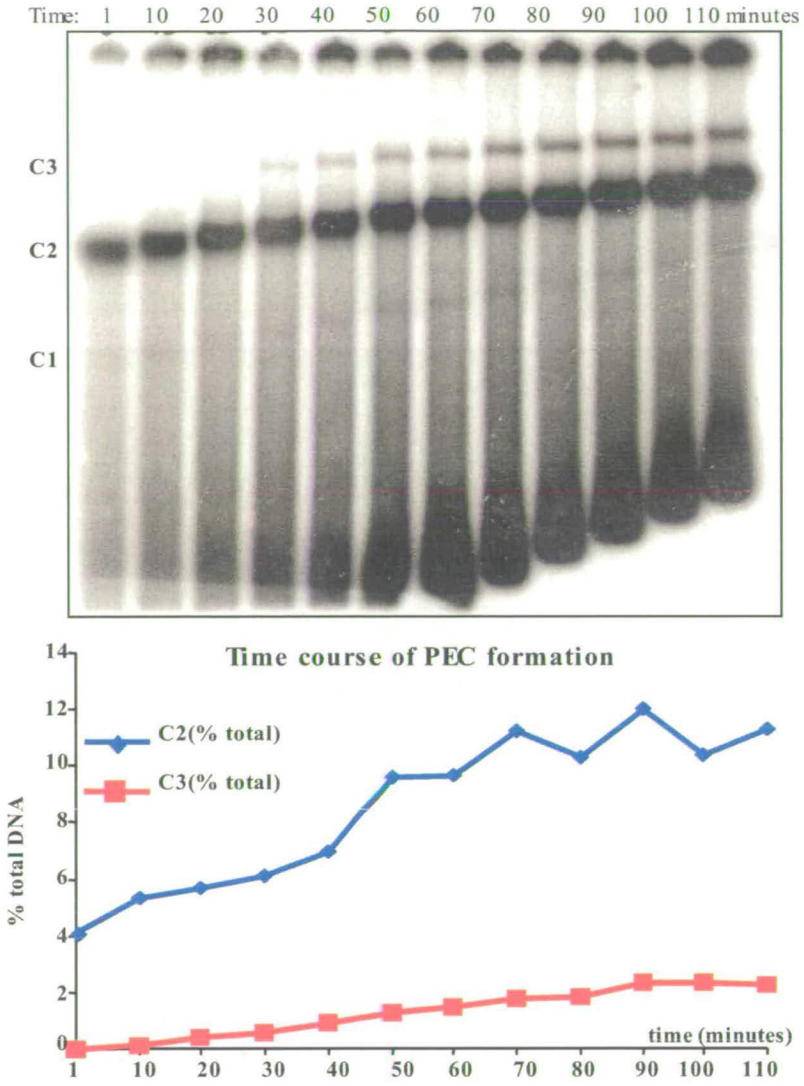

Figure 4.2: Time course of C3 complex formation with MBP-*Mos1* transposase

In vitro transposition reactions were assembled as described in text. At the indicated time-points 10 µl aliquots were taken and the reaction analysed by EMSA on a 6% native gel. The transposase-DNA complexes were quantified using ImageQuant and expressed as percentage from the total DNA present in the last lane (time-point: 110 minutes). **C1** denotes the monomer of transposase bound to one transposon end. **C2** denotes the dimer of transposase bound to one transposon end. **C3** represents the synaptic complex formed by MBP-*Mos1* transposase with two transposon ends. The **C1** complex is not represented on the graph being too faint to give reliable measurements by this method.

substrate (substrate AS*), 20% (v/v) DMSO and 5 mM $MgCl_2$ was assembled in a final volume of 150 µl. The reaction was incubated at 30 °C and 10 µl aliquots were loaded immediately on a 6% native polyacrylamide gel. The following time points were analysed, 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110 minutes (**Figure 4.2**). The **C2** and **C3** complexes were quantified using a STORM phosphor imager and the ImageQuant programe (**Figure 4.2**). Due to its weak signal and the high background in that region of the gel, the **C1** complex could not be reliably quantified. The **C3** complex is formed later in the reaction compared with **C2** and **C1**, consistent with it being a later intermediate in the transposition reaction, as expected for a PEC.

4.4 Stability of the C3 complex at higher temperatures

Synaptic complexes and PECs have been shown to be very stable (Dawson and Finnegan, 2003; Lamberg *et al.*, 2002; Savilahti *et al.*, 1995). This stability has enabled their use as mutagenic tools in various species of prokaryotes (Lamberg *et al.*, 2002). Synaptic complexes withstand higher temperatures (Dawson and Finnegan, 2003), and heparin treatments (Savilahti *et al.*, 1995). *Mos1* PEC has been shown to be stable for 1 hour at 55 °C (Dawson and Finnegan, 2003). To assess the stability of the **C3** complex, *in vitro* transposition reactions were assembled and incubated at increasing temperatures. The transposase-DNA complexes were resolved by EMSA on native 6%

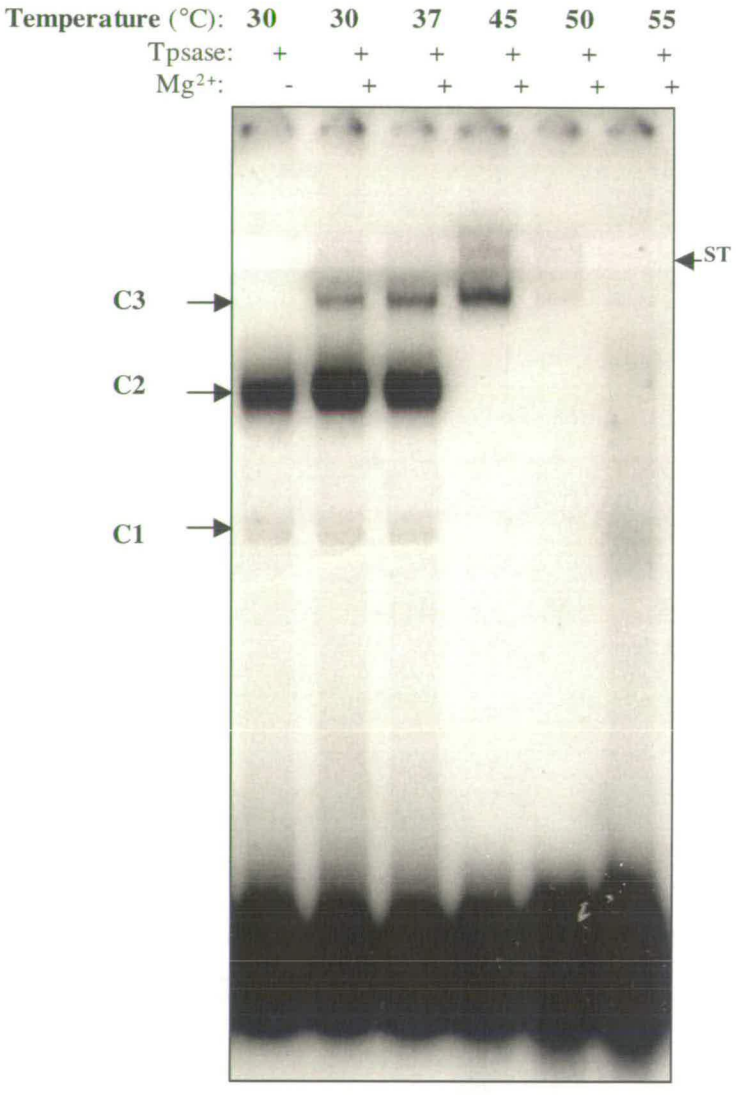

Figure 4.3: Temperature stability of C3 complex

The stability of **C3** complex was assessed by incubating the reactions at increasing temperatures for 60 minutes and analysing the transposase-DNA complexes by native polyacrylamide gel electrophoresis (PAGE). It can be seen that the **C3** complex is more stable than **C1** and **C2,** withstanding one hour at 45°C, conditions under which **C1** and **C2** are destabilised. This is consistent with **C3** being a synaptic complex. ST denotes strand transfer (see sections 3.5 and 3.6).

polyacrylamide gels. A six fold volume was assembled. One aliquot of 20 µl was removed and used as control from which $MgCl_2$ was omitted. 5 mM $MgCl_2$ was added to the rest of the reaction, and the reaction was started by the addition of 6nM MBP-*Mos1* transposase. Reactions were incubated for 2 hours at 30 °C. The reaction containing $MgCl_2$ was aliquoted in 20 µl aliquots and incubated at the following temperatures: 30, 37, 45, 55, and 65 °C for a further one hour. From each reaction, 10 µl aliquots were loaded on a 6% native polyacrylamide gel (**Figure 4.3**). In the absence of Mg^{2+} ions no **C3** is formed. In the presence of Mg^{2+} the **C3** complex is formed which is stable for one hour at 45 °C. The monomer (**C1**) and the dimer (**C2**) complexes withstand one hour at 37 °C but dissociate from the DNA at 45 °C. The **ST** complex becomes visible after the 45 °C incubation, being very stable. In conclusion, the **C3** complex behaves as a stable protein-DNA complex, withstanding one hour at 45 °C, a temperature at which the monomer and dimer complexes dissociate. This is consistent with the behaviour of a synaptic complex.

4.5 What is the stoichiometry of Mos1 *Paired End Complex?*

As described above, **C3** represents the Paired End Complex formed by MBP-*Mos1* transposase. The Ferguson analysis predicts that this complex contains a tetramer of transposase bound to two transposon ends. This prediction is based on the relative mobility of the complex on polyacrylamide gels. The migration of a protein-DNA

complex on a polyacrylamide gel is influenced by its mass, electrical charge and shape. By running the complex on different percentages of polyacrylamide gels, the influence of electrical charge on mobility is eliminated, so as the complexes migrate solely according to their mass and shape. Thus, if a conformational change occurs at the transition from monomer or dimer to the **C3** complex, this could alter its mobility in the EMSA assay employed by the Ferguson analysis. Another drawback of the Ferguson analysis is that as the complex migrates slower, the precision of the method decreases. This is due to the fact that on gels with high polyacrylamide concentrations these complexes resolve poorly, hence the migrated distance is difficult to measure accurately. So, caution is required when the stoichiometry of **C3** is obtained using this method. As a second method to investigate the stoichiometry of *Mos1* PEC complex, I took advantage of the different size of the MBP-*Mos1* transposase fusion compared with *Mos1* transposase. MBP-*Mos1* fusion transposase has a molecular weight of 82.8 kDa (42.48 kDa MBP plus 40.3 kDa the *Mos1* transposase), approximately double the size of the *Mos1* transposase alone. This difference in size between the two forms of the protein can be used to address the title question in a more direct fashion than that employed by the Ferguson analysis. By mixing the two forms of *Mos1* transposase (MBP-*Mos1* and *Mos1* transposase) two protein-DNA complexes are formed by the two individual proteins, and a number of complexes with intermediate size, hence migration, corresponding to mixtures of the two proteins in different ratios. The number of transposase monomers can be then determined from the formula $N = n + 1$, where N represents the number of transposase monomers inside the PEC, and n designates the

number of protein-DNA complexes with intermediate mobility formed. The method was extensively used to address related questions regarding the V(D)J recombination reaction (Bailin *et al.*, 1999; Swanson, 2002a; Swanson, 2002b).

Factor Xa protease was used to cleave off the MBP part of the fusion generating the *Mos1* transposase. First, conditions were sought to determine the time and the amount of protease required to cleave approximately half of the fusion protein, thus leaving half still fused to MBP. This generates the two different forms of *Mos1* transposase: MBP-*Mos1* of approximately 80 kDa and, *Mos1* transposase of circa 40 kDa. Empirically, it was determined that by incubating approximately 120 μg fusion with 1 μg factor Xa for 16 hours at 4 °C resulted in roughly half the protein being cleaved. These conditions were used to generate the two forms of *Mos1* transposase and the mixture was added in a standard reaction as described in Materials and methods (section 2.2.5.2). A dimer of MBP-*Mos1* transposase bound to one molecule of DNA substrate has a molecular weight of approximately 230 kDa and might obscure the intermediate PEC formed by one fusion and one cleaved transposase monomer (complex which would have approximately 250 kDa). The reactions were incubated at 30 °C for two hours to allow the PECs to form. At this point, the reactions were divided in half. One half was incubated at 30 °C for another hour (**Figure 4.4** lanes1-4), and the second half was incubated for one hour at 45 °C to destabilise the dimer complex (**Figure 4.4** lanes 5-6). The protein-DNA complexes were separated on a 6% native polyacrylamide gel. Two

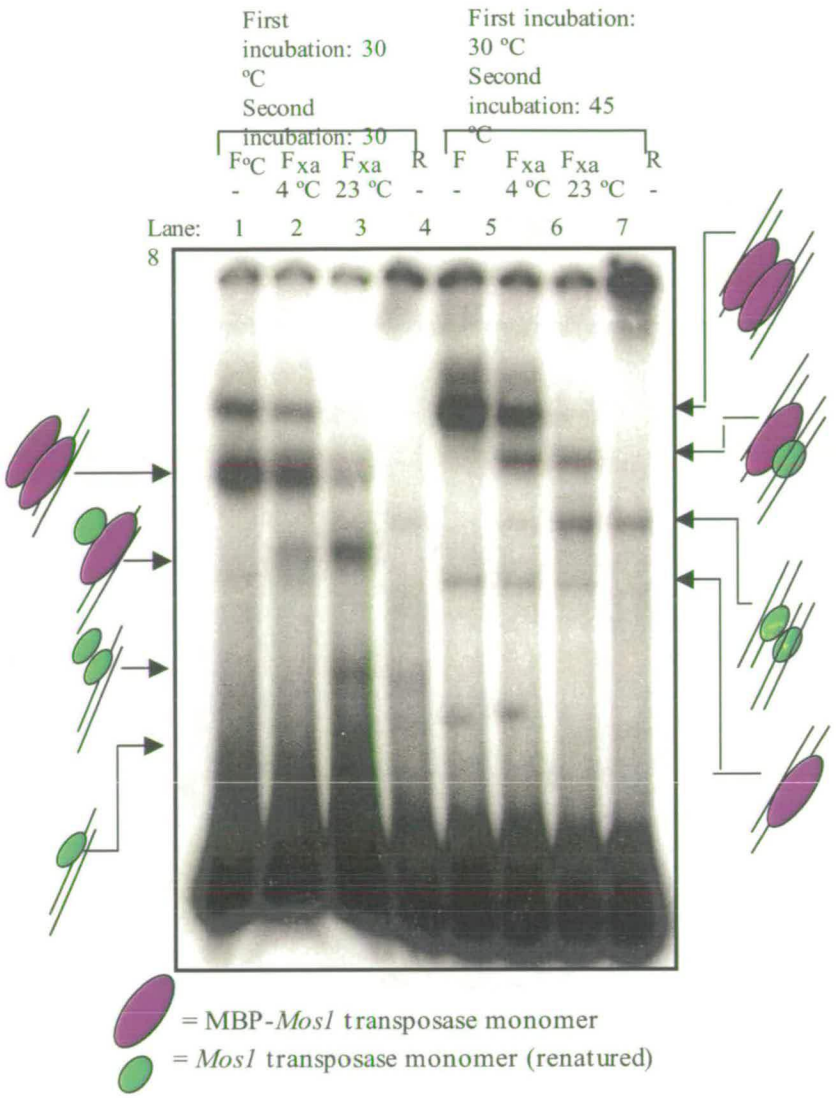

Figure 4.4 The stoichiometry of *Mos1* synaptic complex

Figure 4.4: The stoichiometry of *Mos1* synaptic complex

The *in vitro* transposition reactions were assembled using the AS* DNA substrate (sections 2.2.5.1 and 2.2.5.2). The MBP-*Mos1* transposase was cleaved with Factor Xa protease from bovine serum for 16 hours at 4 °C and 23 °C. The cleaved fusion transposase was used in gel-shift assays Two incubation temperatures were used: first incubation was at 30 °C. The second incubation was at 30 °C for 1 hour for lanes 1 - 4. In lanes 5 - 8, the second incubation was at 45 °C to destabilise the dimer complexes and favour synaptic complex formation (section 4.4). F = MBP-*Mos1* fusion transposase uncleaved. F_{xa} = MBP-*Mos1* fusion transposase cleaved with factor Xa protease at 4 °C or 23 °C. R = *Mos1* transposase purified through renaturation. Fusion transposase is represented by purple ovals. The transposase is represented as green ovals. **ST** denotes strand transfer complexes (see sections 3.5 and 3.6).

different conditions were used for factor Xa cleavage: 16 hours at 23 °C and 16 hours at 4 °C as described in **Figure 4.4**. As controls, MBP-*Mos1* fusion and *Mos1* transposase purified by renaturation were used (lanes 1, 5 and 4, 8, respectively). When the MBP-*Mos1* is partially cleaved with factor Xa, an intermediate complex is formed which withstands the one-hour incubation at 45 °C (lanes 6 and 7). This complex is seen only when both forms of the *Mos1* transposase are present. It does not appear in the control lanes when MBP-*Mos1* and *Mos1* proteins are present separately. The stability of this complex at higher temperatures suggests that it is a PEC. Again, after incubation at 45 °C the **ST** complex becomes visible (**Figure 4.4**, lanes 5 and 6).

An alternative approach was used to test if the number of intermediate complexes is the same if the factor Xa protease cleaves a PEC preformed with the MBP-*Mos1* fusion protein. Standard reactions were assembled and the PEC was formed with the fusion protein in the presence of 5 mM $MgCl_2$ for 2 hours at 30 °C. Then 0.5 μg factor Xa was added to one reaction. The reactions were further incubated at 30 °C for one hour and switched to 45 °C for one hour. As controls, reactions containing fusion protein without factor Xa and *Mos1* transposase purified by renaturation were used. The same results were obtained, with only one complex with intermediate mobility being formed (**Figure 4.5**).

Factor Xa cleaves after the arginine residue in its preferred cleavage site Ile-(Glu or Asp)-Gly-Arg. It will also sometimes cleave at other basic residues depending on the conformation of the protein substrate (Kiyoshi *et al.*, 1985). When the factor Xa cleavage products were analysed by SDS-PAGE and probed using antibodies against *Mos1*, one band with an intermediate size is detected (**Figure 4.6 b**). This is cause for concern because if this cleavage product retains its ability to form a PEC, then the intermediate PEC observed might not be actually formed through the mixing of the two forms of *Mos1* transposase. A few lines of evidence argue against this possibility. The most common non-specific cleavage site favoured by factor Xa is a Gly residue followed by an Arg (New England Biolabs pMAL protocol). This sequence of amino acids is present in *Mos1* transposase at the amino acids 117 and 118. Two studies also reported non-specific cleavage with factor Xa protease after Lys residues (Nagai and Thogersen, 1984; Wearne, 1990). The study of Nagai and Thogersen identified that Factor Xa cleaved a cIIFX β-globin fusion protein C-terminal of the sequence Arg – Lys. *Mos1* transposase contains three such putative cleavage sites: Site I corresponding to amino acids 132 and 133, Site II corresponding to amino acids 145 and 146 and Site III corresponding to amino acids 167 and 168. Should the factor Xa protease cleave at these positions, it will generate peptide products with the molecular weights listed in **Table 11**. The molecular weights of the cleavage products were determined using the Peptide Mass software (http://us.expasy.org/tools/peptide-mass.pl) (Wilkins *et al.*, 1997). The products N-terminal of the non-specific cleavage sites contain the MBP moiety fused to the N terminal part of *Mos1* transposase. These polypeptides cannot

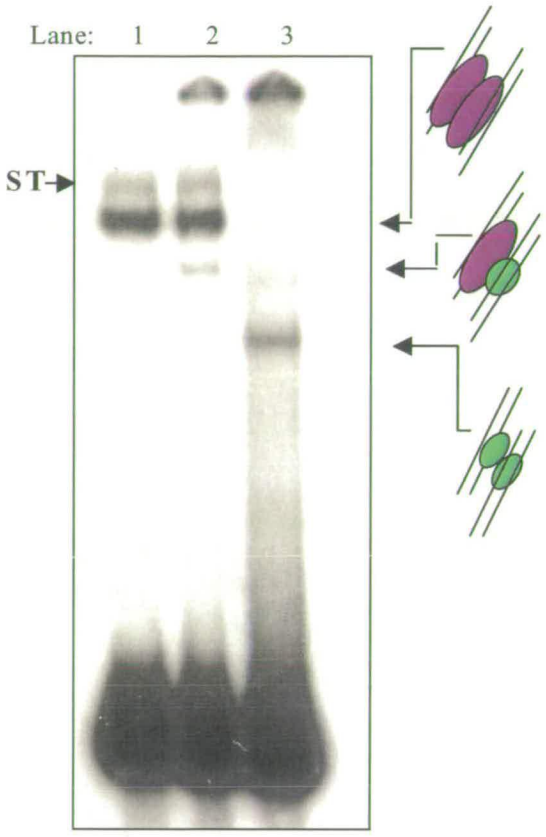

Figure 4.5: Factor Xa cleavage of a pre-assembled PEC

Figure 4.5: Factor Xa cleavage of a pre-assembled PEC

PECs were assembled as described in section 2.2.5.2 using the AS* substrate (section 2.2.5.1) and factor Xa was added to the reactions for one hour at 4 °C. A 45 °C incubation was used to destabilize the monomer and dimer complexes (see section 4.4). Aliquots from each reaction were analysed on a 6% native polyacrylamide gel (section 2.2.5.5). In lane 1 the reaction contained only the MBP-*Mos1* fusion protein, with no factor Xa added to the reaction, in lane 2 factor Xa was added to the reaction as described in text (section 4.5). In lane 3 the reaction contained transposase protein purified through a renaturation process as described (Dawson and Finnegan, 2003). Monomers of MBP-*Mos1* fusion are represented as purple ovals. Monomers of *Mos1* transposase are represented as green ovals. **ST** represents the strand transfer complex (see sections 3.5 and 3.6).

.

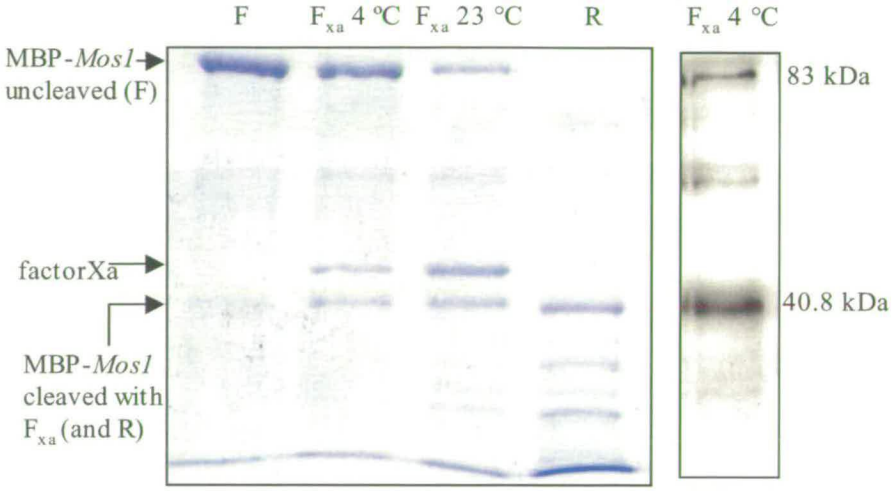

Figure 4.6a

Figure 4.6b

Figure 4.6. Factor Xa cleavage of MBP-*Mos1* fusion protein

Figure 4.6a. SDS-PAGE gel showing the cleavage products resulting after MBP-*Mos1* fusion protein cleavage with factor Xa at 4 °C and 23 °C. F represents the MBP-*Mos1* fusion protein. F_{xa} represents the MBP-*Mos1* cleaved with factor Xa at 4 °C and 23 °C (F_{xa}4 °C and F_{xa}23 °C, respectively). R denotes the *Mos1* transposase purified through renaturation (Dawson and Finnegan, 2003). The expected cleavage product runs with the same mobility as the *Mos1* transposase purified by renaturation.

Figure 4.6b. Western blot of cleavage products after factor Xa cleavage at 4 °C. From the same gel, a lane containing the MBP-*Mos1* cleaved with factor Xa at 4°C was analysed by western blotting as described in section 2.2.3.3. The molecular weights of the MBP-*Mos1* fusion (83 kDa) and the *Mos1* transposase from which the MBP tag was cleaved off (40.8 kDa) are shown. For a discussion regarding the cleavage product with intermediate mobility see text (section 4.5).

Factor Xa cleavage site	N-terminal cleavage product (kDa)	C-terminal cleavage product (kDa)
Gly-Arg	55.7	27.5
Arg-Lys (site I)	58	25
Arg-Lys (site II)	60	23.5
Arg-Lys (site III)	55.7	27.5

Table 11. Predicted mass of putative cleavage products expected if Factor Xa cleaves MBP-*Mos1* at the indicated sites.

form a PEC completely lacking the catalytic domain of the protein. Correspondingly, the products C-terminal of these cleavage sites completely lack the DNA binding domain so cannot interact with the DNA. The location of these putative cleavage residues seems particularly attractive since they are located at the boundary between the N – terminal region and the C-terminal region of the protein. This region is expected to have a more relaxed structure and factor Xa cleavage has been shown to cleave in partially unfolded regions of proteins. Thus, the intermediate band observed after factor Xa cleavage of the MBP-*Mos1* transposase, might represent the MBP fused to the N-terminus of *Mos1* transposase.

4.6 Discussion

The assembly of higher order nucleo-protein complexes is essential in all site-specific recombination reactions. These complexes termed synaptic complexes serve to bring the two recombining sites into close proximity, and as important regulatory mechanisms ensuring that the reaction follows a proper pathway (Grindley, 2002; Savilahti and Mizuuchi, 1996; Savilahti *et al.*, 1995; Surette and Chaconas, 1992; Yang *et al.*, 1996).

These complexes have an intricate structure in all the models studied so far in greater depth. A paradigm for this complexity is represented by the synaptosome of the $\gamma\delta$ and Tn*3* resolvases (Burke *et al.*, 2004; Grindley, 2002; Rice and Steitz, 1994a; Sarkis *et al.*, 2001). These complexes contain 12 monomers of resolvase, bridging the two

recombining DNA segments. In this context, the DNA is cleaved, the four strands exchanged in a process proposed to involve a dramatic conformational change, and the DNA religated in the new configuration. Transposases also involve similar protein-DNA intermediates, as it has been shown for *Mu*, Tn*5*, Tn*10* amongst the prokaryotic elements, and the eukaryotic transposon *Mos1*. The prokaryotic complexes have a simpler structure, with the *Mu* transpososome containing four transposase monomers, and the Tn*5* and Tn*10* structures containing two transposase monomers. *Mos1* transposase purified by renaturation was the first eukaryotic transposase for which a PEC was determined *in vitro*. The experiments presented above examine the ability of a soluble form of *Mos1* transposase to form a PEC. In the previous chapter, three protein-DNA complexes were determined by EMSA and their molecular weight was estimated by Fergusson analysis. Two complexes formed in the absence of Mg^{2+} and were shown to represent a monomer of transposase bound to one transposon end (**C1**), and a dimer also bound to one transposon end (**C2**). The complex formed in the presence of Mg^{2+} ions (**C3**) was identified as representing the PEC formed by the soluble *Mos1* transposase. By using a mixture of DNA substrates of different lengths, the complex was shown to contain two transposon ends. The monomer and dimer complexes precede the **C3** complex formation as expected, since the PEC is formed in later stages of the reaction. In other systems, as the protein-DNA complexes advance in the transposition process, they become more committed and tend to become more stable. This was also found for **C3**, which is stable under conditions in which the monomer and dimer complexes dissociate. Under the experimental conditions used, the PEC formed

with the MBP- *Mos1* transposase is less stable than the PEC formed with the renatured form of the protein (Dawson and Finnegan, 2003). One possible explanation for this property is the presence of the MBP part of the fusion, resulting in a more "relaxed" architecture of the PEC. Under certain conditions (after one hour incubation at 45 °C - see section 4.4 and at increased Mg^{2+}concentrations - see sections 3.5 and 3.6), a complex with a more 'smeary' appearance is formed (**ST**). Although I have not analysed this complex due to time limitations, it is likely that it represents strand transfer events into the free DNA substrate present in reactions. The DNA substrate used contains six TA dinucleotides, all of which could potentially act as targets for strand transfer. Further experiments are needed to characterise this complex.

The Ferguson analysis predicted that the *Mos1* PEC contains a tetramer of transposase. The accuracy of the Ferguson analysis decreases however, as the size of the complexes increases. The method is also sensitive to the shape of the protein-DNA complex. Since the PEC contains two DNA substrates of 100 bp each, the size and shape of the DNA might influence its migration in polyacrylamide gels. In this chapter, the stoichiometry of *Mos1* PEC was determined in a more direct way, by using two different sizes of *Mos1* transposase. The experiments showed that only two molecules of transposase are present in the PEC. This raises a few interesting questions. First strand cleavage has been shown to occur independently of PEC formation (Dawson and Finnegan 2003). How does *Mos1* cleave the second strand of the DNA using two monomers of transposase? One strategy is employed by the bacterial elements Tn*5* and Tn*10*,

involving the formation of a hairpin intermediate. This enables one active site to cleave both strands of a DNA duplex, and to further catalyse the strand transfer reaction. *Mos1* cannot use the same pathway, since no hairpin intermediate is involved (Dawson and Finnegan, 2003). One enzyme that has been suggested to cleave two DNA strands with one active site is the intron-encoded endonuclease I-TevI from bacteriophage T4. The suggested mechanism involves the introduction of a DNA bend that would allow the second strand cleavage (Mueller *et al.*, 1995). Do the DNA cleavages occur *in cis* or *in trans*? How does the active site accommodate these two cleavages with opposite polarities?

Chapter 5 MBP-*Mos1* transposase in solution

5.1 Introduction

A question with implications for the mechanism of transposition is, what form of *Mos1* transposase initiates the transposition reaction? Binding of the transposase to the transposon ends represents the first step in the reaction. Different transposases perform this step in different oligomeric states, a fact that influences the subsequent steps required for transposon excision: first strand cleavage, synaptic complex formation and second strand cleavage (not necessarily in this sequence). For example, studies have shown that the V(D)J recombination protein RAG1 is present in solution as a dimer, and binds to Recombination Signal Sequences (RSS) as such (Rodgers *et al.*, 1999). Another eukaryotic transposase, that of the *Sleeping Beauty*, a member of the *Tc1* family of transposons, is present in solution as a monomer and forms tetramers in the presence of DNA (Izsvak *et al.*, 2002).

Mos1 transposase monomers have been shown to interact (Zhang *et al.*, 2001). This property is common amongst transposases since transposition requires the formation of higher order nucleo-protein complexes termed synaptic complexes (Craigie and Mizuuchi, 1985a; Haren *et al.*, 1998; Lohe and Hartl, 1996a; Lohe *et al.*, 1996b). The formation of a synaptic complex requires interaction between the constitutive protein subunits. The ability of *Mos1* transposase to form oligomers has been proposed to be

also involved in the autoregulation of the element, through a process named overproduction inhibition (Lohe and Hartl, 1996a). This process involves the formation of higher-order oligomers of transposase that are inactive in an as yet unidentified step in the transposition reaction. The question addressed below is what oligomeric form(s) of *Mos1* transposase is / are present in solution. This will lead to the question in the next chapter, what form of *Mos1* transposase binds to the transposon ITRs?

5.2 The Mos1 *transposase oligomers in solution*

To determine the form of *Mos1* transposase in solution, the MBP-*Mos1* transposase was analysed by gel-filtration chromatography on Superdex 200 (S200) and Superdex 300 (S300) columns. The two collumns can separate complexes with different molecular weights. The S200 column was used to separate protein complexes with a molecular weight of up to 200 kDa. The S300 column was used to determine the presence of possible oligomeric forms higher than dimers, such as tetramers, which would have a molecular weight of approximately 320 kDa. The protein was analysed in the column buffer described in Chapter 2 section 2.1.2.1 with two different ionic strengths, namely a high ionic strength, 500 mM NaCl, and a low ionic strength of 200 mM NaCl. The elution profiles were determined by measuring the absorbance at 280 nm (**Figures 5.1** and **5.3**) and compared with the elution profile of a series of marker proteins. The elution volume for each marker and transposase peak was used to calculate the K_{av} values as described in Materials and methods section 2.2.5.6. The K_{av} values for the

marker proteins were plotted against logMw (**Figures 5.2** and **5.4**). The plot was used to determine the molecular weights of MBP-*Mos1* transposase peaks as discussed below.

When the transposase was run on the S300 column in buffer of high ionic strength (500 mM NaCl), a series of three peaks was observed, denoted peak 1, 2, and 3 (**Figure 5.1**). When the molecular weight of the protein in the peaks was determined, the peak 1 eluted in the void volume and peak 2 eluted as a complex with a molecular weight of approximately 196 kDa. The peak 3 showed a molecular weight of approximately 50 kDa. The presence of the MBP-*Mos1* transposase in the peak fractions was confirmed by Western blotting using antibodies against *Mos1* transposase (**Figure 5.5**).

Reasoning that the formation of possible higher order oligomers of transposase might be inhibited by the relatively high NaCl concentration, the experiment was repeated with the MBP-*Mos1* transposase purified in buffer containing 200 mM NaCl. The same peaks are observed as in 500 mM NaCl (**Figure 5.3**), peaks that elute in the void volume (peak 1), as an 170 kDa complex (peak 2), as a 60 kDa complex (peak 3) and, peak 4 of approximately 25 kDa (**Figure 5.4**).

In an attempt to increase the resolution of the experiment, and extend the range of the marker proteins used, the MBP-*Mos1* transposase was also analysed on an S200 column. The elution profile is presented in **Figure 5.6**. The first peak (peak 1) elutes in

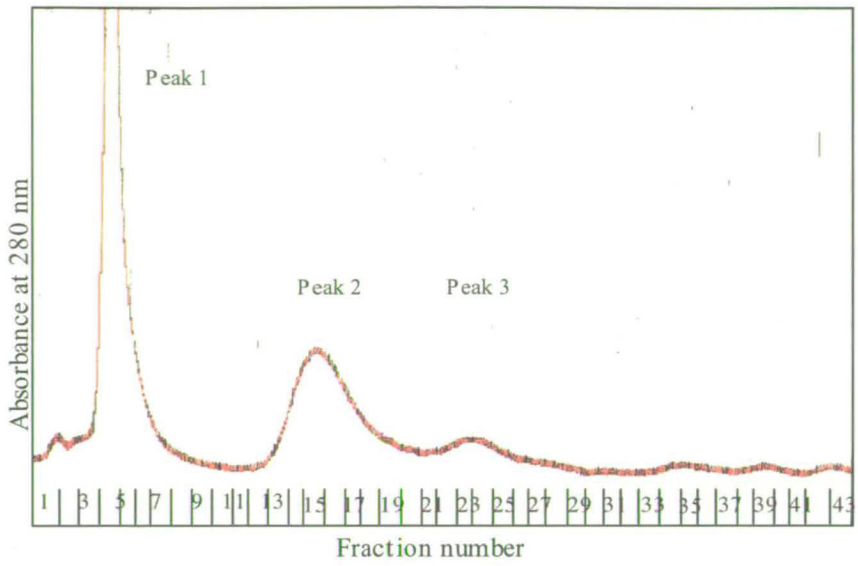

Figure 5.1. Gel filtration profile of MBP-*Mos1* in high salt (500mM NaCl) on the S300 column

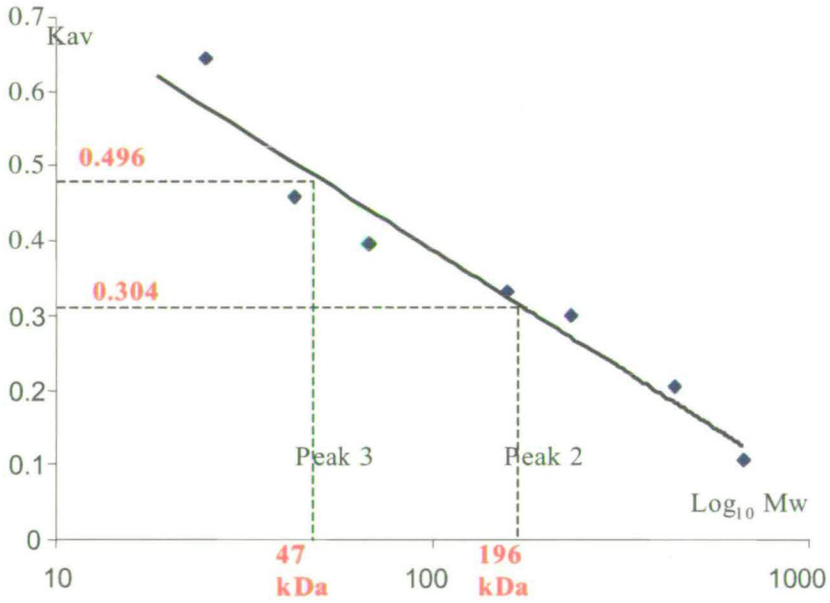

Figure 5.2. Molecular weight plot of peaks for MBP-*Mos1* in high salt (500 mM NaCl). Peak 1 has a molecular weight bigger than 670 kDa running in the void volume on the S300 column.

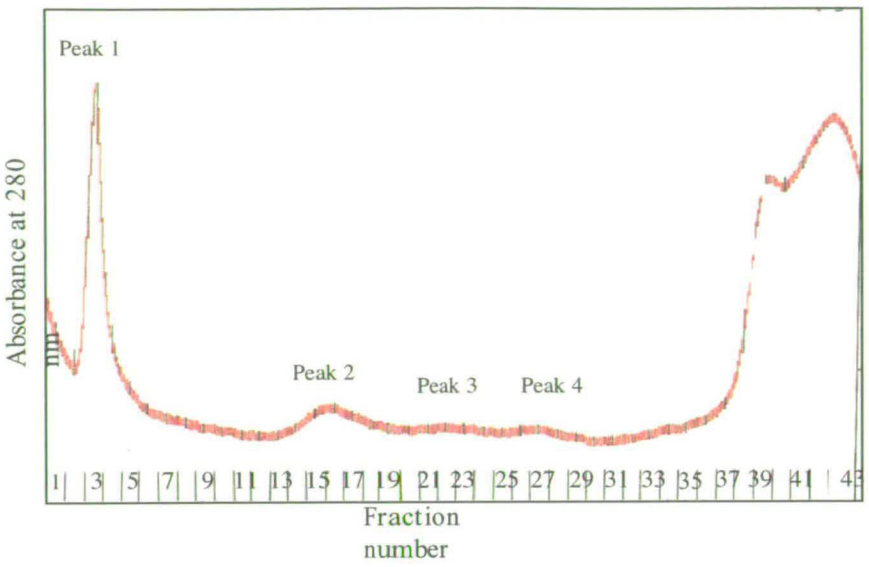

Figure 5.3. Gel-filtration profile of MBP-*Mos1* in low salt (200 mM NaCl) on the S300 column

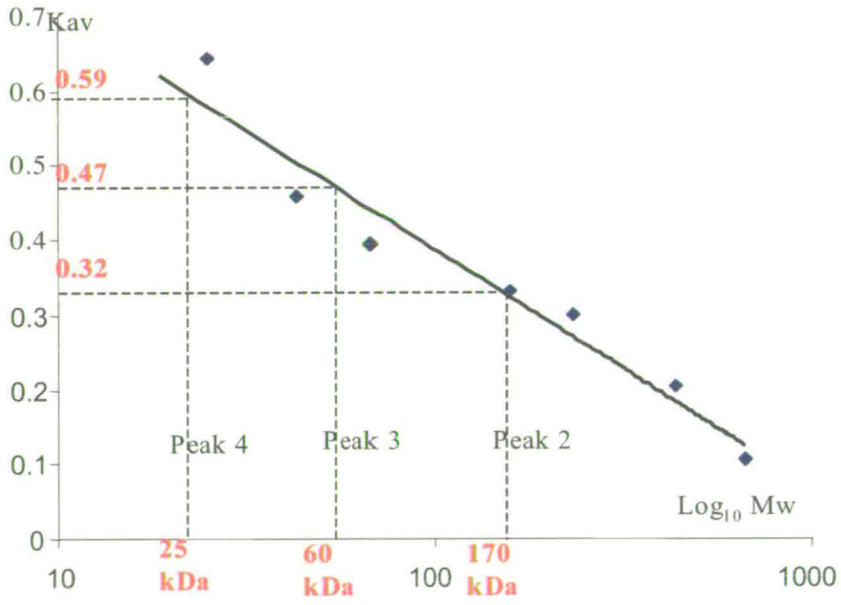

Figure 5.4. Molecular weight plot of peaks for MBP-*Mos1* in low salt. Peak 1 has a molecular weight bigger than 670 kDa so it runs in the void volume on the S300 column.

Figure 5.5 Analysis of fractions resulting from MBP-*Mos1* gel-filtration on the S300 column in 500 mM NaCl

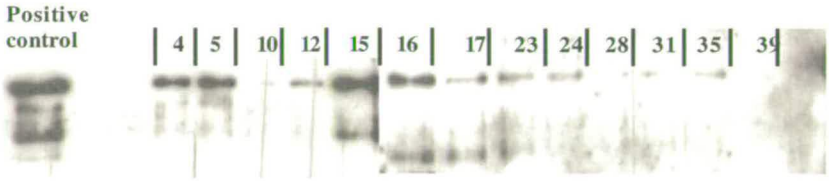

Figure 5.5a

Figure 5.5a. Western analysis of fractions after gel-filtration on S300

The fractions collected after the MBP-*Mos1* was run on the S300 column in 500 mM NaCl were analysed by western blotting (section 2.2.3.3).

Fraction numbers are as in **Figure 5.1**.

Figure 5.5b

Figure 5.5b. First strand cleavage with fractions after gel filtration on S300

The fractions collected after the MBP-*Mos1* was run on the S300 column in 500 mM NaCl were tested for their ability to catalyse first strand cleavage (section 2.2.5.3) using the AS* DNA substrate (section 2.2.5.1). Fractions are numbered as in **Figure 5.1**.

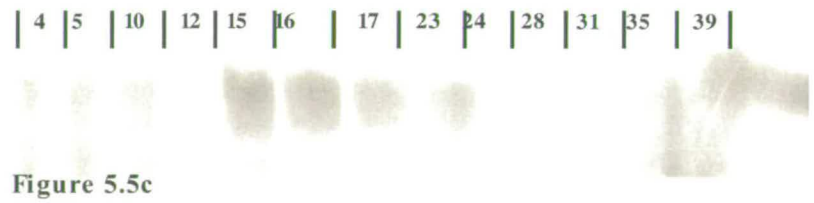

Figure 5.5c

Figure 5.5c. Second strand cleavage with fractions after gel filtration on S300

The fractions collected after the MBP-*Mos1* was run on the S300 column in 500 mM NaCl were tested for their ability to catalyse second strand cleavage (section 2.2.5.4) using the AS* DNA substrate (section 2.2.5.1). Fractions are numbered as in **Figure 5.1**.

the void volume, meaning that the protein aggregate(s) found in this peak are more than 200 kDa. The second peak (peak 2) runs as approximately 196 kDa. The last two peaks (peaks 3 and 4) run at about 20 and 10 kDa, respectively (**Figure 5.7**). The presence of the MBP-*Mos1* transposase in the peak fractions was determined by Western blotting (**Figure 5.8**).

5.3 Is the dimer of Mos1 transposase stable in solution?

Previous studies have shown that the dimer of RAG1 protein in solution after overnight incubation at 4 °C remains stably associated as a dimer (Rodgers *et al.*, 1999). To address the same question for *Mos1* transposase, the fraction resuled from the S200 column and containing the dimer was incubated at 4 °C for two hours and then re-run on the S200 gel-filtration column as described in Materials and methods (section 2.2.5.6). The resulting profile is presented in **Figure 5.9**. After two hours incubation the dimer partially dissociates to the constituent monomers also forming aggregated protein.

5.4 Discussion

The behaviour of transposases in solution has been analysed in a series of studies. The MuA protein exists principally as a monomer (Kuo *et al.*, 1991). Size exclusion

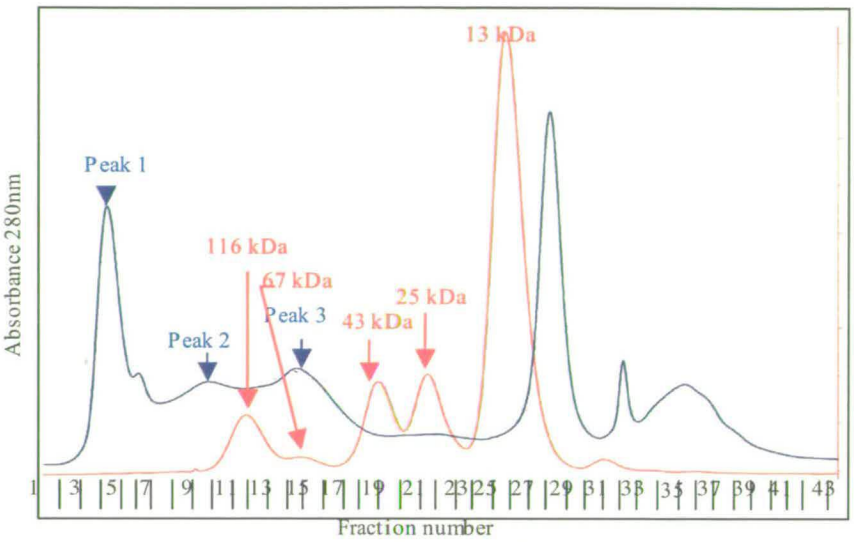

Figure 5.6. Gel filtration profile of MBP-*Mos1* in high salt (500 mM NaCl) on the S200 column.

The MBP-*Mos1* profile is represented by the blue line. The red line represents the molecular weight markers.

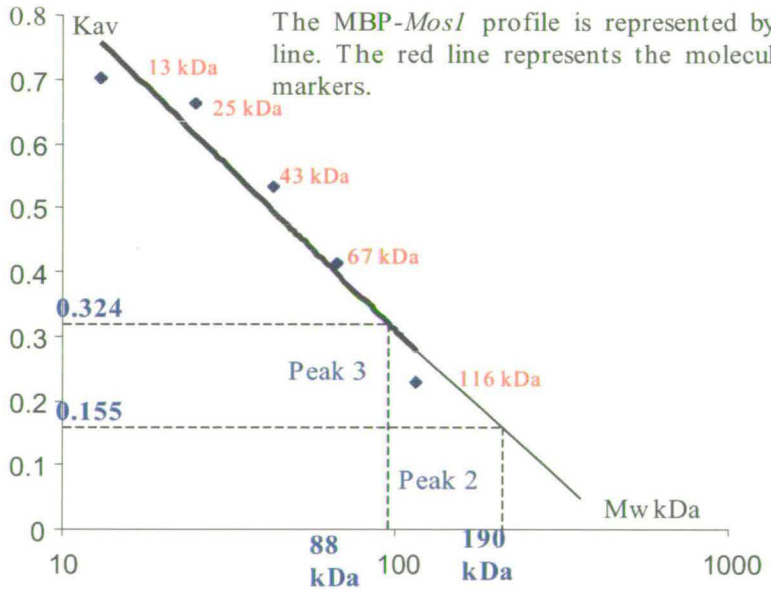

Figure 5.7. Molecular weight plot of peaks for MBP-*Mos1* on the S200 column

Peak 2 has a molecular weight higher than the highest marker. Its molecular weight was extrapolated from the linear plot generated by the marker proteins. Peak 1 runs in the void volume on the S200 column.

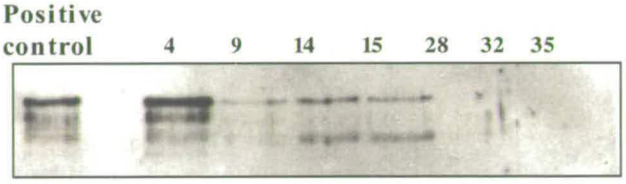

Figure 5.8. Western analysis of peak fractions from Figure 5.6.

chromatography studies showed that in solution RAG1 exists as a dimer in a buffer containing 500mM NaCl (Rodgers *et al.*, 1999). At lower NaCl concentrations, RAG1 is present as a mixture of dimers and tetramers (Godderz *et al.*, 2003). Sedimentation equilibrium was used to show that the N-terminal domain of the *SB* transposase is principally monomeric in solution (Izsvak *et al.*, 2002).

In this chapter gel filtration experiments were used to study the behaviour of MBP-*Mos1* transposase in solution. Two main forms of the transposase were detected: a monomeric form, which elutes as a peak of approximately 88 kDa on the S200 column, and a dimeric form that elutes as a peak of 190 kDa on both the S200 and S300 columns. Importantly, no transposase oligomers bigger than the dimer have been detected in either high or low ionic strength buffers under the experimental conditions used. Some smaller degradation products of approximately 10 and 20 kDa were also observed. These could be due to a proteolytic activity present in the protein preparation.

Rechromatography of the protein that eluted as a dimer yielded again the same three peaks suggesting that there is a dinamic equilibrium between the three forms of the protein. The protein monomers are not stably associated in the dimers, dissociating after two hours incubation at 4 °C. This is in contrast with the RAG1 protein, for which the dimers are very stable (Rodgers *et al.*, 1999)

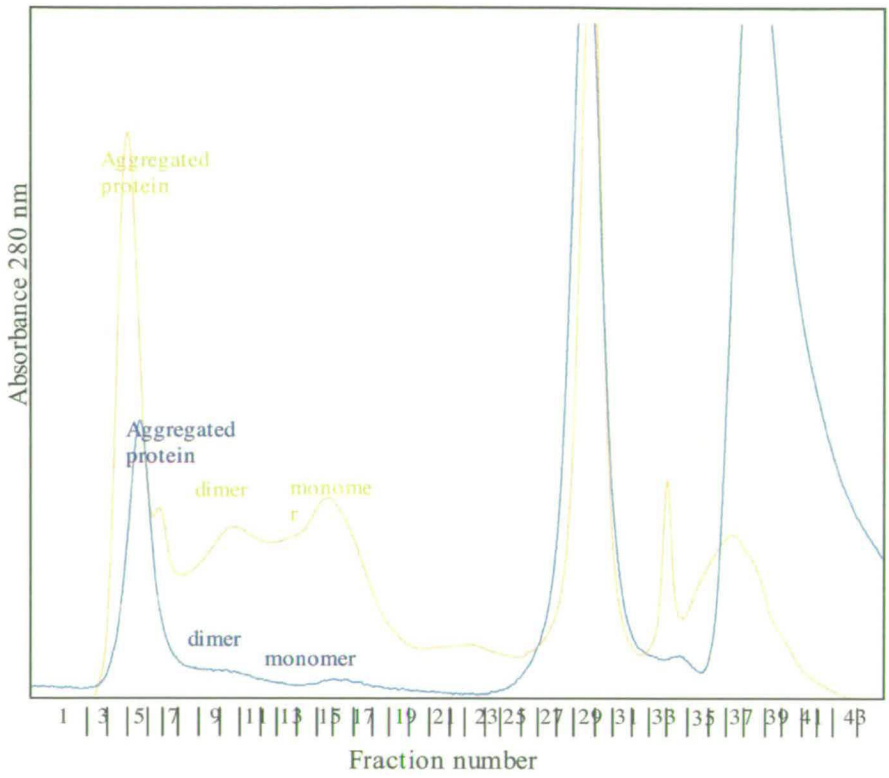

Figure 5.9. Stability of dimer complexes formed by MBP-*Mos1* transposase

The dimer formed by MBP-*Mos1* transposase was separated by gel filtration on an S200 column and re-run on the same column to assess its stability. Aggregated protein, dimers and monomers are all recovered, suggesting that the dimer is able to dissociate. The yellow line represents the first load of transposase. The blue line represents the dimer re-run (dr in **Figures 6.2** and **6.4**).

From these studies we conclude that the MBP-*Mos1* transposase exists in solution as a mixture of mainly monomers and dimers. These two forms of the protein seem to be in equilibrium, with the dimers dissociating into the constituent monomers.

Chapter 6. What is the active form of *Mos1* transposase?

6.1 Introduction

The monomeric and dimeric forms of MP-*Mos1* transposase were separated by gel-filtration and analysed for their ability to catalyse first strand cleavage, second strand cleavage and to bind to the transposon ITRs. *In vitro* transposition reactions were assembled as described (Materials and methods section 2.2.5.2). When the MBP-*Mos1* fractions were collected from the gel-filtration column, 1µl from each of the fractions was added and the reactions were incubated for 120 minutes at 30 °C (fractions are numbered as in **Figure 5.6**).

*6.2 What form of MBP-*Mos1 *transposase is active for first strand cleavage?*

For analysis of first strand cleavage the AS* DNA substrate was used in the transposition reactions described above. Following incubation under standard conditions, 10 µl of each reaction were mixed with an equal volume of STOP Buffer (see Materials and methods section 2.1.2.1) and boiled for 5 minutes. Ten µl were loaded on an 8% denaturing polyacrylamide sequencing gel and run at 50 °C until the bromophenol blue dye reached the bottom of the gel. The gels were dried and exposed

Figure 6.1a. First strand cleavage with MBP-*Mos1* fractions resulting from gel-filtration.

After gel-filtration, the resulting fractions (numbered as in Figure 5.6) were analysed for first strand cleavage (see 2.2.5.3) using the AS* DNA substrate (see 2.2.5.1).

Figure 6.1a

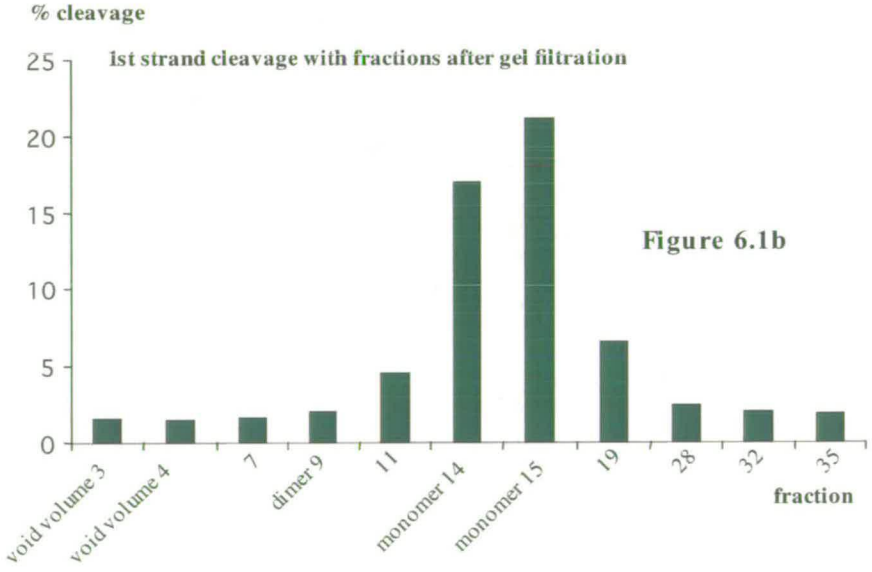

Figure 6.1b

Figure 6.1b. First strand cleavage with MBP-*Mos1* fractions resulting from gel-filtration.

The gel from **Figure 6.1a** was quantified, and the ability of each fraction to nick the first DNA strand was represented as percentage from the total amount of DNA present per reaction.

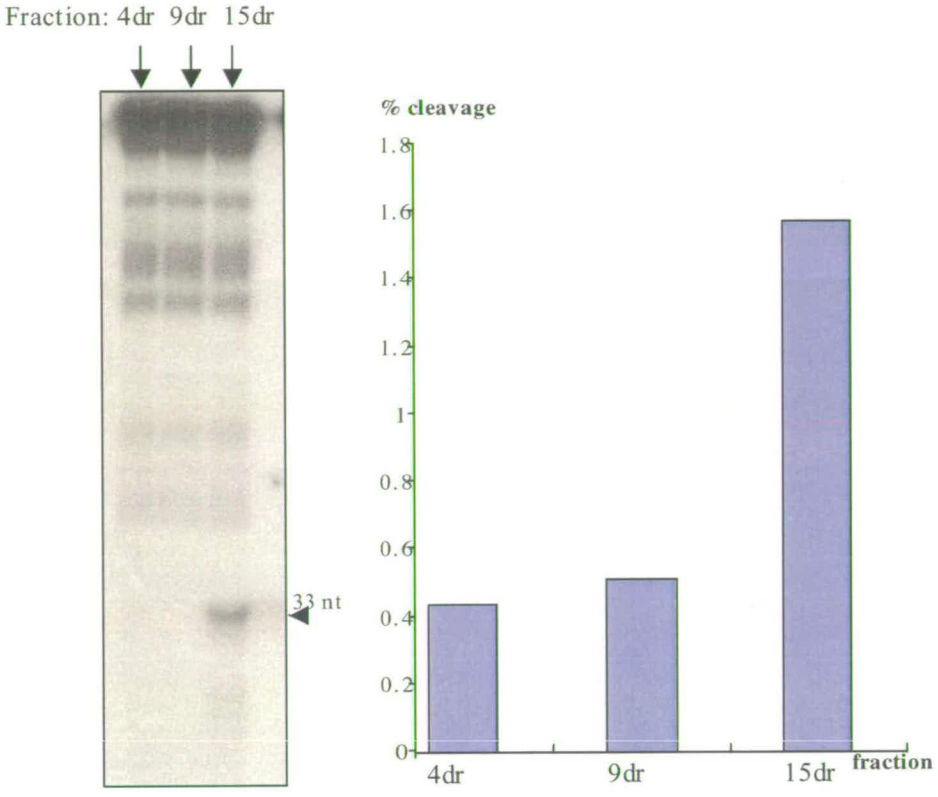

Figure 6.2a

Figure 6.2b

Figure 6.2. First strand cleavage with the fractions resulting from MBP-*Mos1* dimer dissociation.

Figure 6.2a. The ability of the monomer, dimer and aggregated protein fractions resulting from dimer dissociation to cleave the first DNA strand was assessed (see 2.2.5.3). The AS* substrate was used (section 2.2.5.1). The fraction numbers correspond to aggregated protein (4dr), dimer (9dr) and monomer (15dr). Dr = dimer rerun. Fractions are numbered as in **Figure 5.9** (blue line) The monomer is again the active form of the transposase. **Figure 6.2b.** The cleavage activity was quantified by phosphor-imaging and represented as percentage cleavage from total DNA present per reaction.

to an X-ray film for approximately sixteen hours (**Figure 6.1a**). The transposase fractions were added to reactions immediately after gel-filtration to avoid the possible re-assortment of monomers or dimer dissociation. The protein concentration in each fraction was below the limit of detection with the Bradford method, so the intensities of Western blot bands were used to compare the amount of protein found in each fraction. Purified MBP-*Mos1* transposase was used as control (6 nM). Only the fractions F14 and F15 gave a detectable first strand cleavage product of 33 nt (**Figure 6.1a**). According to their elution profile, these fractions correspond to monomers of MBP-*Mos1* transposase. The gels were scanned using a STORM PhosphorImager (Molecular Dynamics) and the 33 nt cleavage product was quantified with the ImageQuant software. The cleavage activity of each fraction was expressed as percentage cleavage from the total DNA substrate present in reaction. (**Figure 6.1b**).

The fractions resulting after dimer dissociation (monomer, dimer and aggregated protein – see **Figure 5.9**) were tested for first strand cleavage as well. The monomeric fraction regains its ability to cleave the first strand after dimer dissociation, showing that the inhibition is only transitory (**Figure 6.2**). When the fractions separated on the S300 column, only one peak is present and active for first strand cleavage (**Figure 5.5b**). This peak elutes from the column as a dimer. It might be that the S300 column can not separate well the monomer and dimer complexes, both eluting in the same peak on this column. An alternative explanation would be that as discussed above, the dimer dissociates reforming the monomer.

6.3 What form of MBP-Mos1 transposase is active for second strand cleavage?

For analysis of second strand cleavage the S* DNA substrate was used in the transposition reactions described above. The S* substrate was assembled pre-nicked at the position of first strand cleavage. This allows second strand cleavage to be assessed directly without it being dependent upon first strand cleavage. Reactions were assembled as described, and the transposase fractions were added to reactions immediately after gel-filtration as above. The protein concentration in each fraction was determined as described for first strand cleavage. Again, only fractions F14 and F15 gave a detectable second strand cleavage product of 70 nt (**Figure 6.3a**). According to their elution profile, these fractions correspond to monomer of MBP-*Mos1* transposase. The gels were scanned using a STORM PhosphorImager (Molecular Dynamics) and the 70 nt cleavage product was quantified with the ImageQuant software. The cleavage activity of each fraction was expressed as percentage cleavage from the total DNA substrate present in reaction (**Figure 6.3b**). From the fractions resulting after dimer dissociation only the monomer is active for second strand cleavage (**Figures 6.4a** and **6.4b**). The S300 column separates only one peak running approximately in the range of a dimer, fraction which is able to catalyse second strand cleavage (**Figure 5.5c**). As discussed above, it might be that the S300 column can not separate well the monomer and dimer complexes, both eluting in the same peak on this column. An alternative explanation would be that as discussed above, the dimer dissociates reforming the

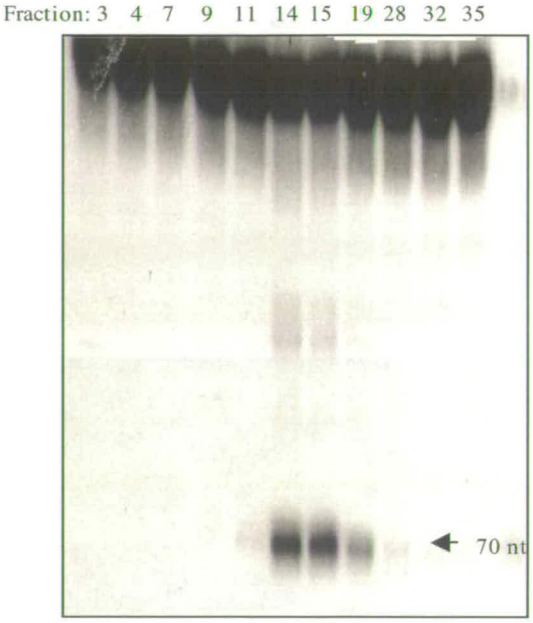

Figure 6.3a. Second strand cleavage with M B P - *M o s 1* fractions resulting from gel-filtration.

After gel-filtration, the resulting fractions noted as in Figure 5.6 were analysed for second strand cleavage (see 2.2.5.4) using the S* substrate (see 2.2.5.1).

Figure 6.3a

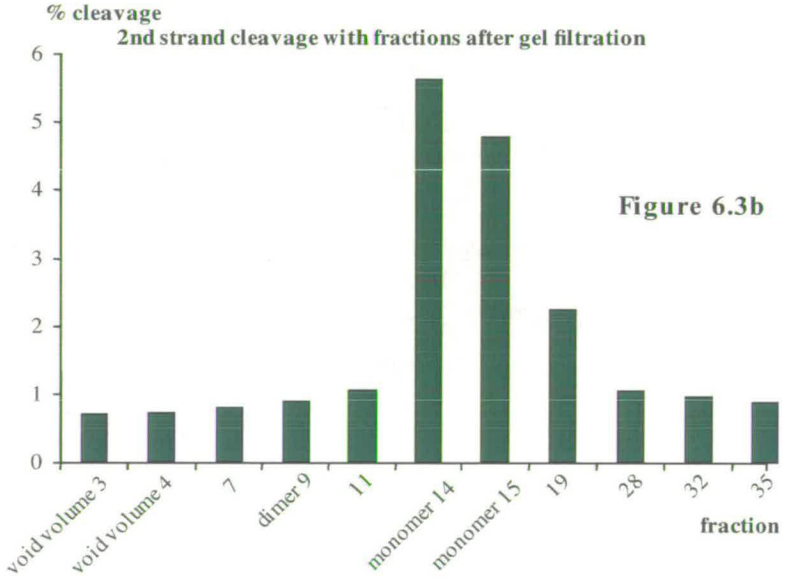

Figure 6.3b

Figure 6.3b. Second strand cleavage with MBP-*Mos1* fractions resulting from gel-filtration.

The gel from **Figure 6.2a** was quantified, and the ability of each fraction to nick the second DNA strand was represented as percentage from the total amount of DNA present per reaction.

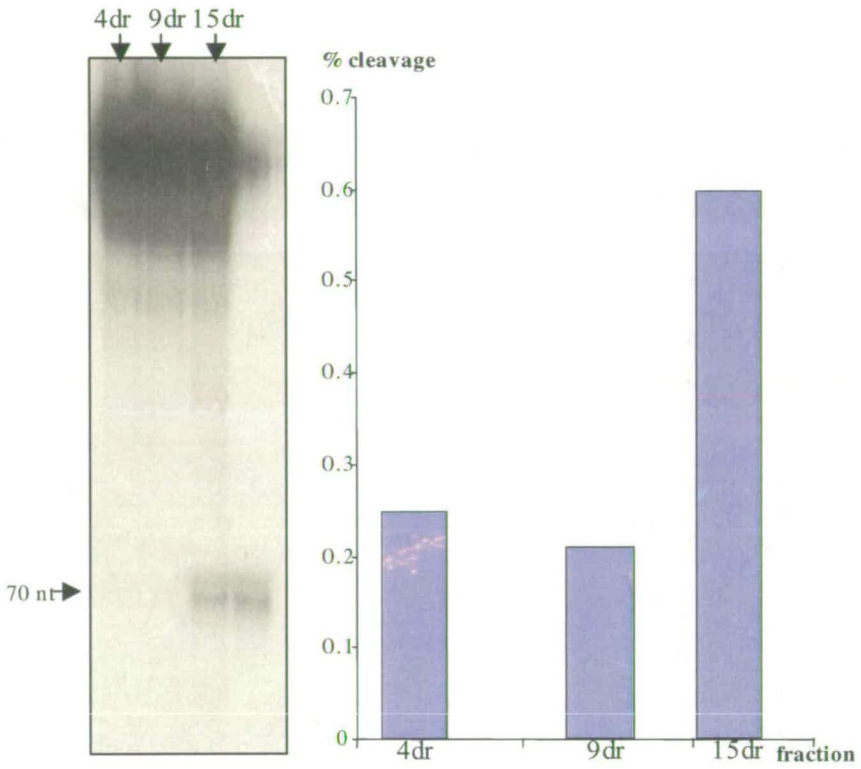

Figure 6.4a **Figure 6.4b**

Figure 6.4. Second strand cleavage with the fractions resulting from MBP-*Mos1* dimer dissociation.

Figure 6.4a. The ability of the monomer, dimer and aggregated protein fractions resulting from dimer dissociation to cleave the second DNA strand was assessed (see section 2.2.5.4). The S* substrate was used (section 2.2.5.1). The fraction numbers correspond to aggregated protein (4dr), dimer (9dr) and monomer (15dr). dr = dimer rerun. Fractions are numbered as in **Figure 5.9** (blue line) The monomer is again the active form of the transposase.
Figure 6.4b. The cleavage activity was quantified by phosphor-imaging and represented as percentage cleavage from total DNA present per reaction.

monomer, which is active at least at the level of DNA binding and maybe first strand cleavage. In this way, the monomer of MBP-*Mos1* transposase is required to initiate the transposition reaction under the conditions used in this study.

*6.4 What form of MBP-*Mos1 *transposase binds the transposon ITRs?*

Binding of the transposase to the transposon's ITRs represents the first step in the transposition reaction. What form of transposase performs this step has important consequences for the subsequent pathway of transposition. Transposase-DNA complexes formed with both the AS* substrate (**Figure 6.5**) and with the S* substrate (**Figure 6.6**) were assessed as described in Materials and methods (section 2.2.5.5). Only the monomer fractions (F14 and F15) can bind DNA forming the monomer complex, dimer complex, Paired End Complex and, Strand Transfer complexes. The final product of the transposition reaction *in vitro* in the absence of required DNA repair factors is the strand transfer product, which accumulates with time. So, an increase in the transposition efficiency is best reflected by the accumulation of these strand transfer products. When the monomer binds to a DNA substrate that is un-nicked at the position of first strand cleavage (substrate AS*), approximately 4.5% of the labelled DNA present in the reaction is found in Strand Transfer complexes. When the monomer binds a DNA substrate that was assembled pre-nicked at the site of first strand cleavage (substrate S*), a higher percentage of the DNA substrate has been integrated into a

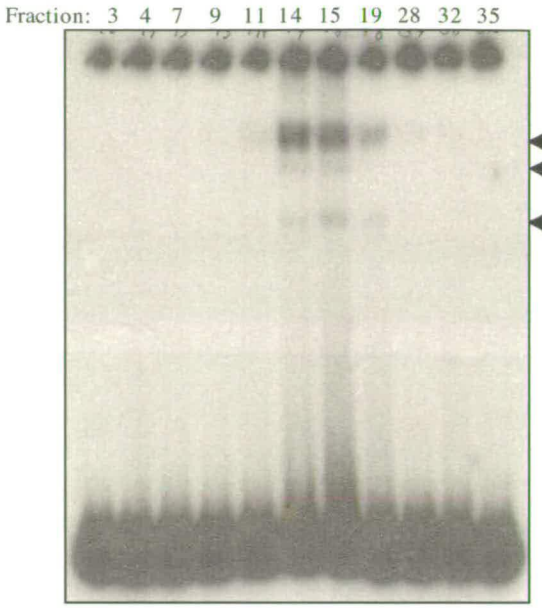

Figure 6.5a

Figure 6.5a. DNA binding with MBP-*Mos1* fractions resulting from gel-filtration.

After gel-filtration, the resulting fractions noted as in Figure 5.6 were analysed for their ability to form the transposase-DNA complexes with the AS* DNA substrate (see text for details). **ST** denotes the strand transfer complex (sections 3.5 and 3.6)

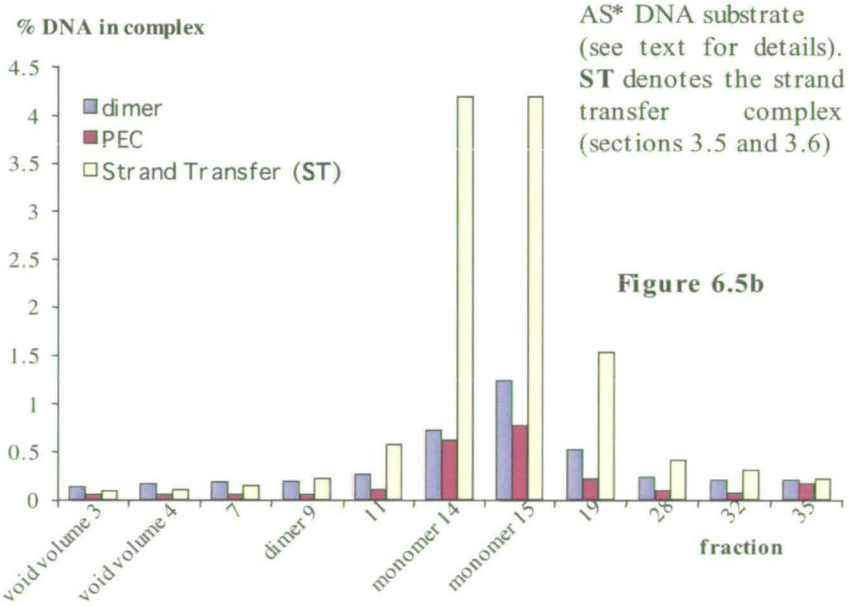

Figure 6.5b

Figure 6.5b. DNA binding with MBP-*Mos1* fractions resulting from gel-filtration.

The gel from **Figure 6.5a** was quantified, and the amount of DNA present in each transposase-DNA complex was represented as percentage from the total amount of DNA present per reaction.

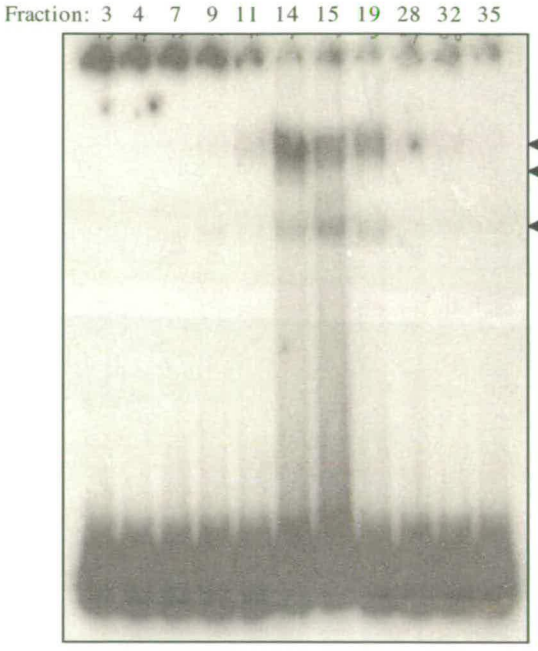

Figure 6.6a. DNA binding with MBP-*Mos1* fractions resulting from gel-filtration.

After gel-filtration, the resulting fractions noted as in Figure 5.6 were analysed for their ability to form the transposase-DNA complexes with the S* DNA substrate (see text for details). St denoptes the strand transfer complex (sections 3.5 and 3.6)

Figure 6.6a

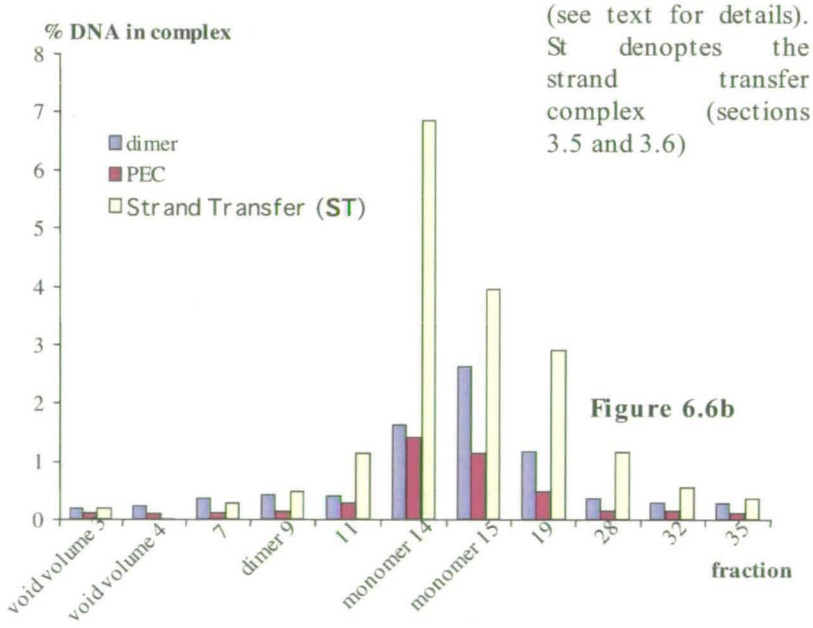

Figure 6.6b

Figure 6.6b. DNA binding with MBP-*Mos1* fractions resulting from gel-filtration.

The gel from **Figure 6.6a** was quantified, and the amount of DNA present in each transposase-DNA complex was represented as percentage from the total amount of DNA present per reaction.

target DNA molecule – approximately 7%. The higher proportion of Strand Transfer complexes when the S* substrate is used most probably reflects the fact that the requirement for first strand cleavage is by-passed.

6.5 Discussions

The ability of monomers and dimers of *Mos1* transposase to undergo DNA binding, first and second strand cleavage has been assessed. The two oligomeric forms of MBP-*Mos1* transposase have been isolated by gel-filtration as described above and used in *in vitro* transposition reactions. Aliquots from these reactions were analysed on native gels by EMSA to detect transposase-DNA complexes and, on denaturing gels to detect the accumulation of cleavage products (33 nt for first strand cleavage and 70 nt for second strand cleavage). Only the fractions containing monomers were able to catalyse all the steps required for transposition to proceed until accumulation of Strand Transfer products. These steps are: binding of the transposase to the transposon ITRs, cleavage of the first strand, cleavage of the second strand, and the strand transfer reaction. These results point towards a series of differences between *Mos1* transposition and other related eukaryotic systems such as V(D)J recombination. RAG1 protein in solution adopts a dimeric form (Rodgers *et al.*, 1999). This is the form that is active for DNA binding. *Mos1* transposase exists in solution as a mixture of monomers, dimers and higher order complexes (probably aggregated protein). The monomer is the form of

Mos1 transposase that can bind the transposon ITRs. The fact that *Mos1* transposase can form oligomeric complexes such as dimers and aggregates of protein and that these oligomeric forms are inactive for transposition is in good agreement with the proposal that at high concentrations the transposase is down-regulating its activity by forming inactive oligomers - a phenomenon termed overproduction inhibition (Lohe and Hartl, 1996a; Lohe *et al.*, 1996b). A few questions are still open. What form of *Mos1* transposase does the first strand cleavage? The experiments described above show that *Mos1* transposase has to be a monomer to bind to *Mos1* transposon ITRs. This does not mean however, that the monomer has to actually catalyse the first strand cleavage reaction (this possibility has been depicted in **Figure 6.7a**). Another possibility would be that one monomer binds to the transposon end. Once bound to the DNA it forms a dimer bound to one transposon end, and it is this dimer that cleaves the first strand (**Figure 6.7b**). Another question regards the pathway(s) involved in synaptic complex formation. A few possible pathways are depicted in **Figure 6.8**. One pathway involves the binding of one monomer of transposase to each of the two transposon ITRs. The monomers perform first strand cleavage, and through an interaction between the two monomers, the synaptic complex is formed (**Figure 6.8a**). An alternative pathway requires the formation of a dimer of transposase bound to one transposon end. The second transposon end enters the reaction as naked DNA forming the synaptic complex (**Figure 6.8b**). The pathway for PEC formation involving the recruitment of a naked ITR raises a series of questions. If a dimer from solution cannot bind DNA, how can a dimer assembled on a single end? How is the first strand cleaved in the ITR that enters

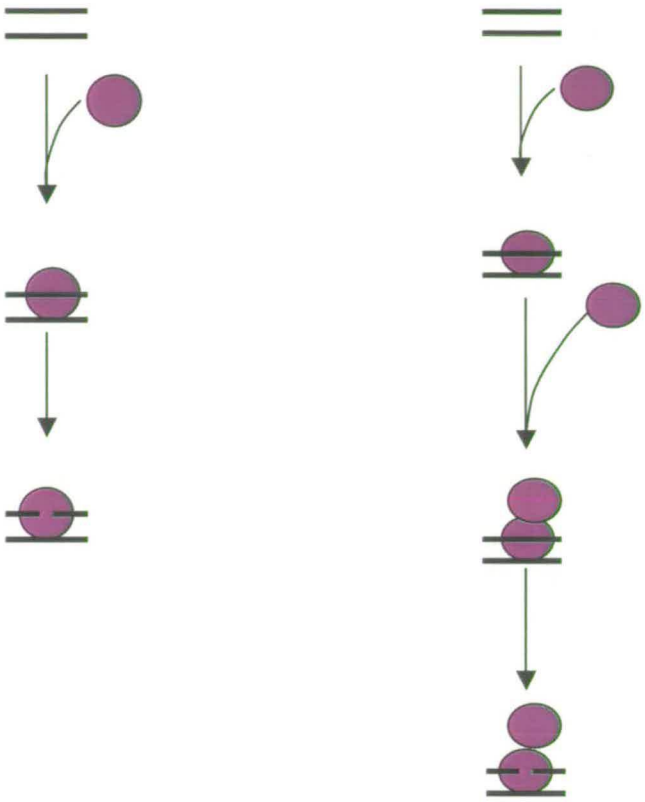

Figure 6.7a

Figure 6.7b

Figure 6.7. Models for first strand cleavage by *Mos1* transposase

Figure 6.7a. In this model, a monomer of transposase binds to each transposon end and nicks the first DNA strand
Figure 6.7b. In this model, the transposase binds the transposon ITRs as a monomer but first strand nicking is done by a dimer of transposase. In both figures the transposase monomers are represented as purple ovals. dsDNA corresponding to one transposon ITR is represented as black lines. Only one ITR is shown.

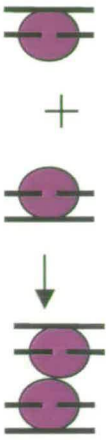

Figure 6.8a

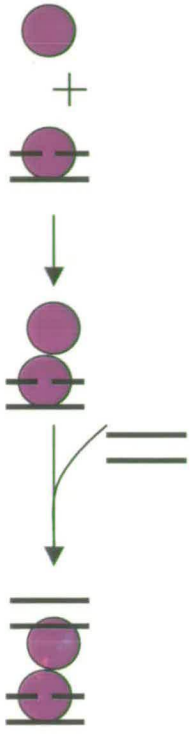

Figure 6.8b

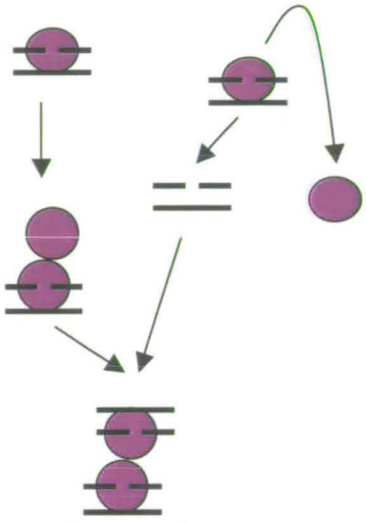

Figure 6.8c

Figure 6.8. Models for *Mos1* synaptic complex assembly

Figure 6.8. Models for *Mos1* synaptic complex assembly

Figure 6.8a. The synaptic complex might be formed through an interaction between two monomers already bound to DNA. Note that in this case both first strand nickings occur outside the synaptic complex.
Figure 6.8b. A second possibility would be that on one transposon end a monomer does first strand nicking, it dimerises, and the second transposon end enters the synaptic complex as naked DNA. In this case the two first strand nicking steps have to occur in an assymetric fashion, with one inside the synaptic complex and the other before the synaptic complex assembly.
Figure 6.8c. First strand nicking done by monomers. One monomer comes off the DNA at one transposon end and the corresponding DNA enters the synaptic complex as pre-nicked naked DNA.

the PEC as naked DNA? The answer to the first problem might involve a conformational change in the protein dimer after assembly on its cognate DNA substrate. The two ITRs of *Mos1* differ at four nucleotide positions, and it has been shown that the transposase binds with a higher affinity to the right ITR (Zhang *et al.*, 2001). This suggests that the transposase can bind to the left ITR and release the DNA after first strand cleavage, producing a nicked ITR to be recruited by the dimer complex assembled on the right ITR (**Figure 6.8c**). These hypotheses remain to be tested.

Chapter 7. Transposition of *Mos1* in *Saccharomyces cerevisiae*

7.1 Introduction

Transposition is a dangerous event for the host organism. Transposable elements can inactivate genes through insertion into ORFs, or influence their expression by transposing into their regulatory regions. In addition, transposition generates lesions in the DNA of the host organism, which in the case of a cut-and-paste transposon are double-strand breaks upon excision and single-stranded gaps on integration.

Broken chromosomes pose a serious threat to cell survival. The presence of an unrepaired double-strand break (DSB) will trigger the DNA-damage response systems, which will cause the cell to arrest its progression through the cell cycle and sometimes to cause apoptotic cell death. Even if a cell with an unrepaired DSB will continue to divide, the broken chromosome fragments can mis-segregate and be lost, degraded, or produce aneuploidy.

In response to this threat, cells have elaborated an entire arsenal of DNA repair pathways that can be classified into two general types of repair: homologous recombination (HR), and non-homologous end joining (NHEJ). A cell has also multiple choices within each of these two main groups of mechanisms (Haber , 2000; Jeggo, 1998; Paques and Haber, 1999).

7.1.1 Homologous recombination mechanisms

In lower organisms, HR represents the major mechanism for the repair of DSBs and requires extensive regions of homology between DNA ends and the repair template. Szostak *et al.* in 1983 and Resnick and Martin in 1976 have proposed a model for the mechanism of HR (Resnick and Martin, 1976; Szostak *et al.*, 1983). During HR, the ends of the DSB are resected by exonuclease activities to produce long, 3'-end single-stranded regions derived from the original break, which will then invade the intact homologous DNA, creating two heteroduplex DNA regions. The gaps are filled by DNA synthesis using the intact DNA as template. Cross-over or non-cross over products are generated, depending on the way in which the resulting Holliday junctions are resolved. In *Saccharomyces cerevisiae*, HR requires the products of the *RAD52* epistasis group, represented by genes numbered from *RAD50* to *RAD58* (Jeggo, 1998; Paques and Haber, 1999)

An alternative repair pathway named single-strand annealing (SSA) has been described, which uses short direct repeat sequences. This process also involves the exonucleolytic resection of broken DNA ends, resulting in ss DNA regions. The ss regions however, are annealed at short regions of homology located originally at both sides of the original DSB. The regions of homology can be as short as 30 nt, but with longer regions the process is much more efficient (Sugawara *et al.*, 2000). The protein products of *RAD1*

and *RAD10* have been shown to resect the resulting single-stranded tails (Fishmann-Lobell and Haber, 1992). Mammalian homologues of *RAD1* and *RAD10* have also been identified, namely *XPF* and *ERCC1* (Brookman *et al.*, 1996; Sijbers *et al.*,, 1996; van Duin *et al.*, 1986). The SSA pathway depends on the *RAD52* and *RAD59* gene products (Davis *et al.*, 2001).

A process termed break-induced replication takes place when only one end of a DSB is engaged in homologous recombination. If the region of homology is located on a different chromosome, or in a different region of the same chromosome, a non-reciprocal translocation occurs (Bosco *et al.*, 1998).

7.1.2 Non-homologous mechanisms

In comparison with yeast, mammalian cells have a much greater ability to join DNA ends with no sequence homology (Roth *et al.*, 1985). The non-homologous end-joining pathway(s) of DNA repair require a set of proteins at least partially conserved between mammalian and yeast cells. In *Saccharomyces cerevisiae* the *YKU70* (*HDF1*) and *YKU80* (*HDF2*) gene products (Boulton and Jackson, 1996a; Boulton and Jackson, 1996b; Milne *et al.*, 1996), have been shown to be required for at least part of the NHEJ repair mechanisms. *LIG4* has also been involved (Schar *et al.*, 1997; Wilson *et al.*, 1997), which has homology to the fourth mammalian DNA ligase (Teo and Jackson, 1997). *SIR2*, *SIR3*, and *SIR4* are required for religation of linearised plasmids with

either 3' or 5' overhangs (Tsukamoto *et al.*, 1997), together with the Rad50/Mre11/Xrs2 protein complex that has endo-and exonuclease activity (Haber, 1998; Milne *et al.*, 1996; Moore and Haber, 1996; Tsukamoto *et al.*, 1996). Other genes involved in NHEJ in yeast, *LIF1* and *NEJ1* have been discovered more recently (Frank-Vaillant and Marcand, 2001; Kegel *et al.*, 2001; Ooi *et al.*, 2001; Valencia *et al.*, 2001).

In mammalian cells, the same core proteins are required for NHEJ, namely Ku70, Ku80, (Featherstone *et al.*, 1999), DNA ligase 4 and Xrcc4 (Grawunder *et al.*, 1997; Li *et al.*, 1995) (the mammalian homolog of Lif1 protein (Herrmann *et al.*, 1998), and the catalytic subunit of DNA phospho-kinase (DNA-PKcs), the latter having no obvious homologue in yeast (Carter *et al.*, 1990; Jackson *et al.*, 1990). In the higher eukaryotes studied, and in *Schizosaccharomyces pombe*, the Rad50/Mre11/Xrs2 complex does not seem to play a key role, as it does in *Saccharomyces cerevisiae* (Haber, 1998).

Sensing of the DNA breaks in NHEJ requires the binding of the Ku70/Ku80 heterodimer to the broken DNA ends (Bradbury and Jackson, 2003). This recruits the catalytic subunit of the DNA-PK (DNA-PKcs) to the site of DNA repair, and also it might be responsible for directing the repair away from HR (Pierce *et al.*, 2001; Smith and Jackson, 1999). DNA-PK, a member of the phosphatidyl 3-kinase-like family of kinases (PIKKs) then recruits and activates other factors involved in NHEJ.

Studies in *Saccharomyces cerevisiae* have shown that the efficiency of NHEJ decreases around 10 to 20 fold in Ku70, Ku80, DNA ligase IV, and sir2, 3, and 4 mutant backgrounds (Boulton and Jackson, 1996a; Boulton and Jackson, 1996b; Teo and Jackson, 1997; Tsukamoto *et al.*, 1997; Wilson *et al.*, 1997). Some rejoining activity still remains, however, suggesting the presence of a third, Ku-independent rejoining process. In wild-type backgrounds, the ends are joined accurately, whereas in the Ku mutants the breaks suffer deletions of various sizes. The junctions in the latter case are frequently at sites of short direct repeat sequences (Moore and Haber, 1996; Yu *et al.*, 2004). This feature resembles the SSA mechanism of homologous recombination described above. In *Saccharomyces cerevisiae*, SSA requires a minimum of 60 to 90 bp of perfect homology in the regions flanking the DSB (Sugawara *et al.*, 2000). During NHEJ, only between one to five bp of homology are sufficient for repair (Mezard *et al.*, 1992).

7.1.3 DNA repair during V(D)J recombination

V(D)J recombination is the process that generates the diversity of Ig light chain and T cell receptor genes, as discussed above. The two lymphoid specific enzymes, RAG1 and RAG2 generate DSBs during lymphocyte development creating two types of DNA ends, blunt signal ends (SE), and hairpin coding ends (CE). The NHEJ arsenal of proteins is used to repair these breaks.

The blunt signal ends are joined precisely, without loss or addition of nucleotides, generating signal joints (SJ). At least four proteins are involved, namely Ku70, Ku80, Xrcc4 and DNA ligase IV. The coding ends are processed in a different manner, generating coding joints (CJ). Apart from the four proteins required for SJ formation, two other proteins are involved in CE processing and CJ formation, Artemis and the catalytic subunit of DNA-phospho kinase (DNA-PKcs) (Gellert, 2002). Artemis, a member of the metallo-β-lactamase superfamily of proteins (Moshous *et al.*, 2001) has been shown to be required for hairpin opening in CE processing (Ma *et al.*, 2002). The proposed model involves the activation of Artemis by DNA-PK through phosphorylation, and hairpin opening by the endonucleolytic activity of Artemis (Ma *et al.*, 2002). At least one additional factor required for V(D)J recombination has been implied from studies on patients with immunodeficiencies (Day *et al.*, 2003).

The V(D)J recombination reaction has been conserved through the evolution of jawed vertebrates (Hsu *et al.*, 1989; Litman *et al.*, 1993; Ventura-Holdman *et al.*, 2001), and the NHEJ pathway has been partially conserved in yeast. Recently, an assay has been developed that allowed the study of SJ formation and RAG-mediated transposition in *Saccharomyces cerevisiae* (Clatworthy *et al.*, 2003). In this assay, the RAG1 and RAG2 proteins are expressed from two different plasmids as fusions to the SV40 nuclear localisation signal (NLS), under the control of the inducible GAL1 promoter. A third plasmid contains the recombination signal sequence (RSS) downstream from the actin promoter. Upon induction of RAG proteins expression, signal joints (SJ) are detected

by a PCR based assay. Intramolecular transposition produces inversion circles or deletion circles, depending on which DNA strand is attacked by the 3'OH group present on the RSS. A similar PCR approach is used to detect and distinguish between these products. This assay provided the first *in vivo* evidence of RAG-mediated transposition (Clatworthy *et al.*, 2003).

7.1.4 DNA repair and transposition of the *Sleeping Beauty* transposon

Sleeping Beauty (SB) has received a lot of attention owing to its high transposition activity in vertebrate cells compared to other transposons, and potential to be developed into a vector for gene therapy and gene-targeting in mammals (Ivics *et al.*, 2004; Izsvak *et al.*, 2000). In light of these applications, the DNA repair processes associated with its transposition have been studied recently (Izsvak *et al.*, 2004; Yant and Kay, 2003).

Cells that do not express the DNA-PKcs exhibit reduced levels of *SB* transposition, whereas over-expression of DNA-PKcs increases transposition. In addition to its role in DNA repair, DNA-PKcs plays a role in *SB* transposition which is independent of its kinase activity. A mutant version of DNA-PKcs, inactive for phosphorylation and unable to support DSB repair, was nevertheless able to stimulate *SB* transposition. A direct interaction between Ku and *SB* transposase has been suggested by co-immunoprecipitation experiments, although it is not clear what subunit of the Ku

heterodimer is involved in the interaction. In the absence of NHEJ factors, transposon footprints are still formed, although with a lower efficiency, by a homologous recombination mechanism of DNA repair, most consistent with synthesis-dependent strand annealing (SDSA) followed by micro-homology-directed end joining (Adams *et al.*, 2003; Nassif *et al.*, 1994). This is consistent with a decrease in *SB* transposition in Xrcc3 and Rad51C mutant backgrounds.

Thus, in contrast to the DNA breaks produced by V(D)J recombination which are absolutely dependent on the NHEJ pathway of repair, the DSBs produced by the *Sleeping Beauty* (*SB*) transposon can be repaired both by NHEJ and HR mechanisms. In addition, the NHEJ proteins DNA-PKcs and the product of the Ataxia Telangiectasia-Mutated (ATM) gene seem to be required for efficient transposition of SB in mammalian cells (Izsvak *et al.*, 2004).

7.1.5 *Saccharomyces cerevisiae* as a heterologous system to study DNA transposition

The ability to repair DNA lesions is of capital importance for an organism, and transposons are an important source of DSBs and gaps in the DNA, with the latter eventually leading to DSBs through DNA replication. Until recently, not much was

known about how the DNA is repaired after transposition, and gaps in our knowledge still need to be filled.

The relative ease with which it can be genetically modified, made yeast a preferred model for the study of a large number of eukaryotic biological processes, including DNA repair. A lot is known about how yeast deals with DNA lesions, and importantly the processes proved to be highly conserved in mammalian cells. In addition, the yeast genome does not contain DNA transposons, so their influence on, and interaction with, the host can be assessed in a heterologous system.

An assay to study the behaviour of maize *Ac/Ds* transposons in *Saccharomyces cerevisiae* has been developed (Weil *et al.*, 2000). It consists of a mini-*Ds* transposon inserted into the yeast ADE2 gene, either in the ORF or in the 5' untranslated region. The mini-*Ds* transposon retains the cis sequences required for its mobilisation. The transposase is provided *in trans* from a plasmid construct, expressing the *Ac* transposase linked to the yeast Gal4 nuclear localization signal (NLS) under the inducible GAL10 or GAL1S promoters. Upon transposase induction and *Ds* element excision, ADE^+ revertants are recovered on the appropriate selective media.

The footprints after *Ds* element excision in yeast strains wild-type for DNA repair proteins suggest the presence of a hairpin intermediate in the flanking DNA, which is consistent with the mechanism proposed for another member of the hAT superfamily of

transposons, *Tam3* from *Antirrhinum majus* (Coen *et al.*, 1989), and resembles the V(D)J recombination mechanism (McBlane *et al.*, 1995). Between 60 and 90% of excisions are followed by insertion of the transposon elsewhere in the yeast genome, producing characteristic 8 bp duplications (Weil *et al.*, 2000).

Yeast mutants lacking genes involved in DNA repair showed a reduce ability to repair *Ac/Ds* excision sites. YKU70, YKU80, RAD50, MRE11, XRS2, SAE2 and PSO2 mutants all show a markedly reduced number of ADE^+ revertants (Yu *et al.*, 2004). These proteins show an involvement of NHEJ pathways in DNA repair after Ac/Ds excision. In the presence of a homologous template, more than half of the repair events still occur by end-joining using regions with microhomology (Yu *et al.*, 2004). The yeast PSO2 gene product is not usually required for NHEJ, being involved in interstrand cross-link recognition (Brendel *et al.*, 2001; Lambert *et al.*, 2003). Deletion of the PSO2 gene still allows approximately 10% of the wild type reversion frequency. The transposon footprints in this case retain part of the transposon DNA at one end, and seem to be a result of hairpin opening at aberrant positions by the transposase. It has been suggested that Pso2 protein might bind to the hairpin structures and position the transposase to open the hairpin correctly (Yu *et al.*, 2004).

Thus, yeast can be a useful model to study transposition and DNA repair. By introducing mutations in different yeast DNA repair factors (see for example PSO2 above), their effect on the repair products can be assessed and information can be

obtained with regard to the way cells protect their genomes from the effects of transposition.

7.2 Experimental design

The assay for *Mos1* transposition in *Saccharomyces cerevisiae* is presented in **Figure 7.1c**). A copy of the *Mos1* transposon is inserted between the ADH promoter and the ADE2 gene generating the transposon-donor plasmid (Ytdp). The transposon was not inserted into the ADE2 ORF because potential excision-site repair products that change the reading frame would not be recovered (Yu *et al.*, 2004). A STOP codon replaced the fifth codon of the *Mos1* ORF (GTG) with a STOP codon (TAA). In this way, the element present in Ytdp is inactive, and can be mobilised only by the transposase provided by a second, transposase-expression plasmid (Ytep). The transposase is fused at the N-terminus to the SV nuclear localisation signal (SV-NLS), to direct its transport through the yeast nuclear envelope (Clatworthy *et al.*, 2003) and its expression is driven by the galactose-inducible GAL1-10 promoter.

Potentially, excision could be screened in two ways. By plating the yeast containing both Ytep and Ytdp on plates lacking adenine and containing galactose as carbon source, ADE^+ revertants can be selected for. In this case, the expression of the ADE2 gene needs to be repressed, which was not the case. Alternatively, a PCR approach can

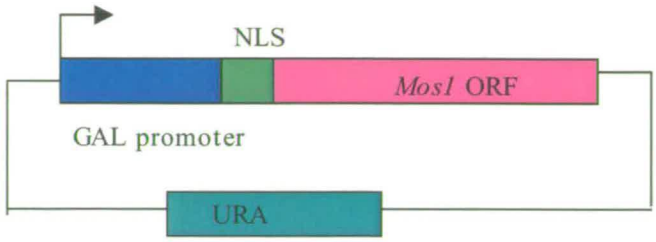

Figure 7.1 a. Transposase expression plasmid (YtepM)

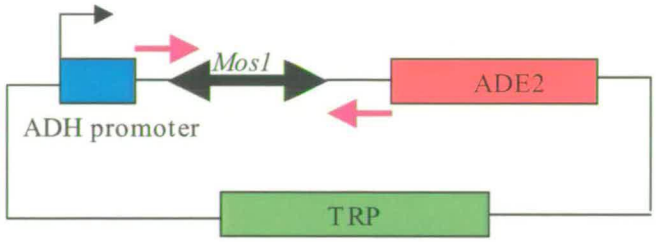

Figure 7.1 b. Transposon donor plasmid (YtdpM)

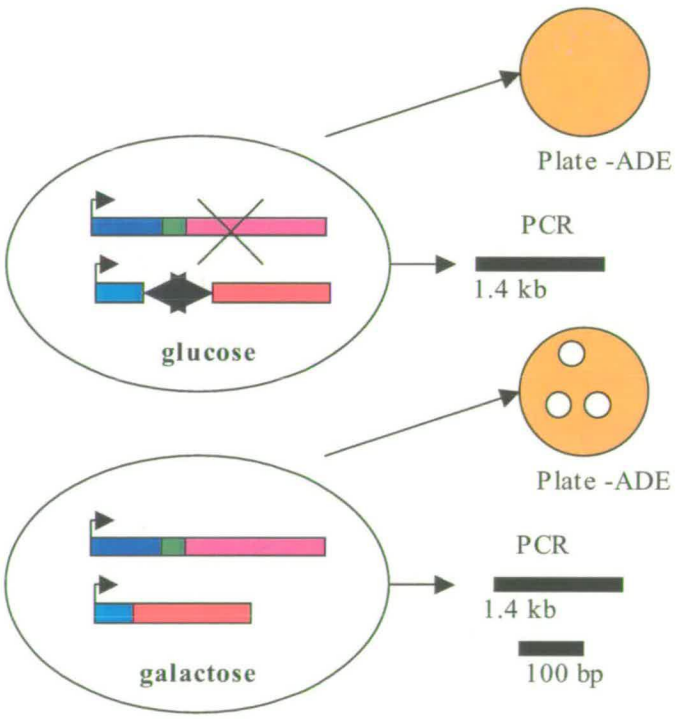

Figure 7.1c. Assay for *Mos1* transposition in *S. cerevisiae*

Figure 7.1 An assay to study *Mos1* transposition in *Saccharomyces cerevisiae*

Figure 7.1 An assay to study *Mos1* transposition in *Saccharomyces cerevisiae*

Figure 7.1a. Transposase expression plasmid (YtepM). The transposase expression plasmid contains the *Mos1* transposase ORF fused in frame with the nuclear localization signal from SV40 (N-terminal fusion). The reading frame is placed under the control of a GAL1 promoter. In this way transposase expression is induced by the presence of galactose. The plasmid contains the yeast URA3 auxotrophic marker and the CEN/ARS sequences which ensure that the plasmid is able to replicate in yeast.

Figure 7.1b. Transposon donor plasmid (YtdpM). YtdpM contains the Mos1 transposon flanked by 100 bp of *Drosophila melanogaster* DNA cloned between the ADH promoter and the ADE2 ORF. The plasmid contains the CEN/ARS replication sequences and the yeast TRP auxotrophic marker. The position of primer sequences used in the PCR detection assay for transposon excision are represented by purple arrows.

Figure 7.1c. Assay for *Mos1* transposition in *S. cerevisiae*. The yeast cells containing both YtepM and YtdpM plasmids are plated on medium containing glucose or galactose. Both media are lacking uracil (URA), triptophan (TRP) and adenine (ADE) to select for the auxotrophic markers present in the plasmids. In the presence of glucose, the transposase expression is repressed, the transposon is present between the promoter and the ADE2 gene, so yeast should not grow on medium lacking ADE. In this case, the PCR assay should give rise to a product of approximately 1.4 kb. In the presence of galactose, transposase expression should be induced, the transposon excised from between the ADH promoter and the ADE2 gene and the ADE2 gene expressed, enabling cells to grow on medium lacking ADE. In this case, the PCR assay for excision should give a product of 100 bp.

be designed, using primers flanking the site of excision. Insertion elsewhere in the yeast genome can be screened by Southern blot, or by an inverse-PCR approach.

7.3 Construction of transposase producing plasmid

The transposase-expressing plasmid Ytep was constructed using the plasmid pBM125 as starting point. The plasmid pBM125 (Rose *et al.*, 1987) (gene accession number: GI 406851) was a gift from Richard Grainger. It contains the CEN4 and ARS1 sequences and the URA3 auxotrophic marker. It also contains the Gal1-10 promoter sequence upstream of a BamHI cloning site.

In order to excise the transposon from the plasmid YtdpM and to insert it into a new location, the transposase expressed from the plasmid YtepM must enter the yeast nucleus. A nuclear localization sequence from the simian virus 40 (SV40-NLS) was used previously to direct the nuclear internalisation of RAG proteins (Clatworthy *et al.*, 2003). A double-stranded DNA fragment containing the SV40 Nuclear Localization Signal was cloned into the pBM125 plasmid downstream of the Gal1-10 promoter by blunt end cloning (see section 2.2.4). The resulting plasmid was named pBM-SV40.

The yeast ADH (Alcohol Dehidrogenase) terminator sequence (ADHter) was used to ensure the presence of polyadenylation signals that are functional in yeast downstream

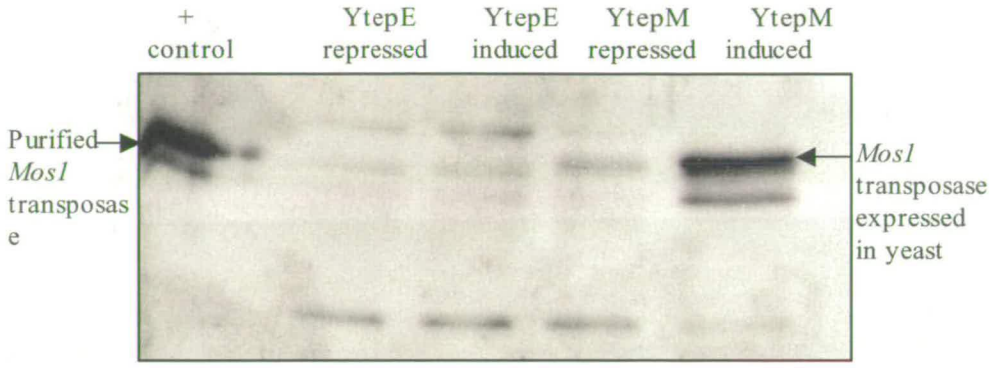

Figure 7.2. Expression of *Mos1* transposase in *Saccharomyces cerevisiae* was detected by Western blott..

Western blott was performed as described in section 2.2.3.3. In the repressed lanes the yeast was grown on glucose. In the induced lanes the yeast was grown on galactose medium. YtepE = empty vector; YtepM = vector containing *Mos1* ORF under the control of GAL1 promoter. As positive control, *Mos1* purified by renaturation was used (Dawson and Finnegan, 2003).

of the *Mos1* ORF. The ADHter sequence was amplified by PCR using the plasmid pEG202 (Serebriiskii *et al.*, 1999) as template and the primers ADHT-sense and ADHT-anti-sense (section 2.1.2.7). The PCR product was cloned into the Ytep by blunt end cloning (see section 2.2.4). The resulting plasmid was designated YtepE.

The *Mos1* ORF was PCR amplified from the plasmid pMos with primers *Mos*-Bam-ATG and *Mos*-Bam-STOP-anti-sense primers (section 2.1.2.7) and cloned into the BamHI site in YtepE in frame with the SV40NLS (section 2.2.4) (**Figure 7.1a**). The in frame junction between the SV-NLS and *Mos1* ORF was checked by sequencing. The resulting plasmid was named YtepM. The expression of *Mos1* transposase was induced as described in Materials and methods (section 2.2.2.4) and checked by western blotting using antibodies against *Mos1* transposase (**Figure 7.2a**).

7.4 Construction of transposon donor plasmid

For the construction of the transposon-donor plasmid, a cassette containing the ADH promoter-ADE2 ORF-ADH terminator was amplified by PCR from the plasmid pAA (Chang *et al.*, 1994) using the primers pAA-sense and pAA-antisense (section 2.1.2.7). The yeast vector pRS314 (Sikorski *et al.*, 1989) was digested with BamHI, and the ends were blunted with the Klenow fragment of *E. coli* polymerase I and dephosphorylated using CIP (see section 2.2.4). The cassette containing the ADH promoter-ADE2 ORF-

ADH terminator was cloned into the pRS314 vector by blunt end cloning generating the plasmid YtepE.

The *Mos1* element was amplified from the plasmid pMos by a mutagenic PCR designed to replace the fifth codon from *Mos1* ORF (GTG) with a STOP codon (TAA) using the primers *Mos*-stop-forward, *Mos*-stop-reverse, p*Mos*-forward, p*Mos*-reverse (section 2.1.2.7). The PCR product containing the mutated element flanked by *Drosophila*-derived DNA was cloned between the ADE2 ORF and the ADH promoter in the plasmid YtdpE generating YtdpM (**Figure 7.1b**). The element was inserted in inverse orientation with respect to the ADH promoter and the ADE2 ORF in the plasmid YtdpM1 or in direct orientation in YtdpM11. The presence and the orientation of the transposon insert were verified by PCR.

7.5 Testing the assay

The yeast strain KY117 was transformed with the plasmids YtepM and YtdpM1 or YtdpM11, and the ability of *Mos1* to repress the yeast growth on medium lacking Ade was assessed (**Figure 7.3**). No difference could be detected when the element was present or absent from between the promoter and the ADE2 reporter gene. This indicates that the ADE2 ORF expression is not blocked, and a PCR approach is required to test for transposon excision. The primer pair p*Mos*-forward and p*Mos*-reverse can be used in this scope (section 2.1.2.7), generating a PCR product of

Figure 7.3. PCR test for *Mos1* excision in *Saccharomyces cerevisiae*.

Primers flanking the *Mos1* element were used (pMos for and pMos rev). Without excision a product of around 1.4 kilo bases is visible Upon excision a product of approximately 100 base pairs should be visible. lane 1 = buffer control; lane 2 = negative control (plasmid DNA from *E. coli*); lanes 3 - 13 = plasmid DNA extracted from *S. cerevisiae* grown on galactose medium was used as template.

YtepM+YtdpE (-Ura-Trp)

YtepM+YtdpE (-Ura-Trp-Ade)

YtepM+YtdpM1 (-Ura-Trp)

YtepM+YtdpM1 (-Ura-Trp-Ade)

YtepM+YtdpM11 (-Ura-Trp)

YtepM+YtdpM11 (-Ura-Trp-Ade)

Figure 7.4. The presence of *Mos1* between the promoter and the ADE2 ORF does not block the expression of the ADE2 gene.

The yeast strain KY117 was transformed with the YtepM and YtdpM1 or YtdpM11 plasmids and plated on glucose media lacking Ura and Trp or Ura, Trp and Ade. Should the transposon block the expression of the ADE2 reporter gene, the yeast would not grow on the –Ade medium. As negative control, Ytdp with no transposon was used (YtdpE). Ytdp = yeast transposon donor plasmid. Ytep = yeast transposase expression plasmid. M denotes the presence of *Mos1* transposon in inverse orientation (M1) or direct orientation (M11).

approximately 1400 bp when the transposon is present. Upon excision and DNA repair, the size of the PCR product should be approximately 100 bp. The PCR assay failed to detect any excision events (**Figure 7.2b**).

7.6 Discussions

An assay to study transposition of *mariner* elements in a system with a wealth of information on DNA repair such as *Saccharomyces cerevisiae* would potentially be a very useful tool. I have attempted to design and assess such an assay, comprising a transposase-producing plasmid and a transposon donor plasmid. The *Mos1* transposase is directed to the nucleus by the SV40 nuclear localisation sequences. Should it be active in yeast, the transposase should excise an inactive copy of *Mos1* present on the transposon donor plasmid. I have not been able to detect any excision events using a PCR assay. One possible explanation is that the transposase is expressed from a single copy plasmid, so the amount of transposase produced might be quite low. If its ability to catalyse transposition is not very efficient in yeast, than no excision would be detected. Further experiments are required to test and improve the assay, allowing excision and transposition to be detected.

Chapter 8. Conclusions and future work

The mechanism of *mariner* elements transposition has only recently started to be revealed in detail and, in the light of their potential use as vectors for gene therapy, more information is required for their development as efficient vectors. Synaptic complexes are crucial intermediates in site-specific recombination reactions, including transposition. Studies in systems such as *Mu*, Tn*10*, Tn*5*, V(D)J recombination, and the *Tc3* and *gamma-delta* resolvases have led to a detailed picture of how these complexes are assembled and their structure.

Mos1 is a member of the *mariner* family of transposable elements and uses a cut-and-paste mechanism for its transposition (**Figure 1.2b**). In this mechanism the transposon is excised from the donor DNA and inserted into a new location in the target DNA. Previous work has shown that the *Mos1* transposase purified through a denaturation-renaturation process is able to catalyse *Mos1* excision and transposition both *in vivo* and *in vitro* (Dawson and Finnegan, 2003). At this stage, a synaptic complex was detected by EMSA, but no information regarding the protein stoichiometry or the pathway leading to the formation of this complex was available. Little work has been done using the *Mos1* transposase purified in a soluble form, with mostly its DNA binding properties being analysed (Auge-Gouillou *et al.*, 2000; Auge-Gouillou *et al.*, 2001). No

synaptic complex has previously been identified using the soluble form of *Mos1* transposase.

In the work presented here, the *Mos1* transposase was purified, as a soluble fusion to the Maltose Binding Protein from *E. coli*, and its ability to undergo all the steps required for synaptic complex assembly was studied. The behaviour of MBP-*Mos1* transposase parallels well that of the previously described *Mos1* transposase purified by renaturation (Dawson and Finnegan, 2003) in its ability to cleave the two DNA strands at the end of *Mos1* and to form protein-DNA complexes with the transposon ITRs. A synaptic complex was detected and its stoichiometry was analysed by a mixing strategy as described in Chapter 4, showing a dimeric form of the transposase in this complex.

In Chapter 5, the transposase forms present in solution were analysed, revealing a mixture of monomers and dimers. The association is not stable under the experimental conditions used with the dimers dissociating and returning to the monomeric state. The experiments presented in the next chapter (Chapter 6) identify the monomer as being the active form of *Mos1* transposase at least at the level of DNA binding.

Thus, through the work presented in this thesis, at present we know that the soluble form of *Mos1* transposase parallels well the behaviour of *Mos1* transposase purified by renaturation regarding the position of DNA cleavages and the protein-DNA complexes formed including its ability to form a synaptic complex. This complex contains a dimer

of transposase, raising interesting questions regarding how the second strand cleavage is achieved without a hairpin intermediate (see also the Discussions for Chapter 4). *Mos1* transposase exists in solution as a mixture of monomers and dimers, with the monomer being the active form at least at the level of DNA binding. Thus, binding of a monomer of transposase to a transposon ITR is an initial stage in synaptic complex assembly (see also Discussions to Chapter 6).

Further biochemical work is required to analyse the *mariner* synaptic complex in more detail. Possible directions include footprinting and protein-DNA crosslinking of the synaptic complex and the monomer and dimer complexes. Together with a wide range of mutagenesis experiments these would provide a much more detailed picture of *mariner* elements transposition reaction.

The assay for *mariner* transposition in *Saccharomyces cerevisiae* failed to detect any excision of the *Mos1* in yeast. This could be due to a number of reasons, such as low transposase activity in yeast, failure of the transposase to enter the nucleus, or low transposase expression. Further experiments are required to assess the ability of *Mos1* to transpose in yeast and to address questions related to the DNA repair after its transposition.

Appendix: Ferguson analysis tables

Protein	Oligomeric form		Molecular weight (Da)
α-lactalbumin (from bovine milk)	monomer		14,200
Carbonic anhydrase (from bovine erytrocytes)	monomer	Charge isomer	29,000
	monomer		
Albumin (from chicken egg)	monomer		45,000
Albumin (from bovine serum)	monomer		66,000
	dimer		132,000
Urease (from Jack bean)	trimer		272,000
	hexamer		545,000

Table 2: Molecular weight of protein markers used in Ferguson analysis

Gel%		5	6	7	8	9	10
Dye (mm)		105	108	114	117	104	108
α-lactalbumin (mm)		80	78	77	70	63	62
Carbonic anhydrase (mm)		30	28	28	27	25	24
Egg albumin (mm)		74	70	68	60	53	51
Serum albumin (mm)	1	66	59	56	45	37	35
	2						
Urease (mm)	3	37	28	22	13	8	6
	6	30	17	8	4	2	1

Table 3: Migration of Bromophenol blue dye and marker proteins on gels with increasing polyacrylamide concentration

Gel%		5	6	7	8	9	10
Rf α-lactalbumine		0.76	0.72	0.67	0.59	0.60	0.57
Rf carbonic anhidrase		0.28	0.25	0.24	0.23	0.24	0.22
Rf egg albumin		0.70	0.64	0.59	0.51	0.50	0.47
Rf serum albumin		0.62	0.54	0.49	0.38	0.35	0.32
Urease	3	0.35	0.25	0.19	0.11	0.07	0.001
	6	0.28	0.15	0.07	0.03	0.02	0.001

Table 4: Mobility of marker proteins relative to the bromophenol blue dye on gels with increasing polyacrylamide concentration

Gel%		5	6	7	8	9	10
α-lactalbumin		188.19	185.91	182.93	177.67	178.17	175.89
carbonic anhydrase		145.48	141.32	138.91	136.17	138.02	134.83
Egg albumin		184.75	181.15	177.52	170.92	170.67	167.39
Serum albumin		179.79	173.71	169.1	158.43	155.02	151.05
urease	3	154.6	141.32	128.33	104.92	88.08	74.03
	6	145.48	119.58	84.5	53.14	30.1	-

Table 5: The 100[log(Rfx100)] values calculated for the marker proteins

Gel%	5	6	7	8	9	10
Dry gel (mm)	110	112	119	120	108	112
Wet gel (mm)	105	108	115	117	104	109
correction	0.95	0.96	0.96	0.97	0.96	0.97

Table 6: Correction coefficients for polyacrylamide gels after drying

Gel%	5	6	7	8	9	10
C1 (mm)	42	30	22	-	17	-
C1 corrected (mm)	39.9	28.8	21.1	-	16.3	-
C2 (mm)	34	22	12	12	7	-
C2 corrected (mm)	32.3	21.1	11.5	11.6	67	-
C3 (mm)	30	18	10	5	3	-
C3 corrected (mm)	28.5	17.2	96	48	-	-

Table 7: Migration of MBP-*Mos1* transposase-DNA complexes C1, C2, C3 and, correction after gel drying

Gel%	5	6	7	8	9	10
RfC1	0.38	0.267	0.183	-	0.149	-
RfC2	0.307	0.195	0.1	0.099	0.064	-
RfC3	0.271	0.159	0.083	0.041	-	-

Table 8: Mobility of MBP-*Mos1* transposase-DNA complexes C1, C2 and C3 relative to the bromophenol blue dye (Rf).

Gel%	5	6	7	8	9	10
C1	157.97	142.65	126.24	-	117.31	-
C2	148.71	129	100	99.56	80.61	-
C3	143.29	120.14	91.9	61.27	-	-

Table 9: The 100x[log(Rfx100)] values for C1, C2 and, C3.

marker protein		slope	transposase-DNA complexes	slope
α-lactalbumin		-2.57	C1	-10.39
carbonic anhidrase		-1.88		
egg albumin		-3.56	C2	-16.56
serum albumin (monomer)		-6.01		
urease	trimer	-16.74	C3	-27.43
	hexamer	-29.72		

Table 10: Calculated slopes for marker proteins and C1, C2 and, C3.

References

Adams, M. D., McVey, M., Sekelsky, J.J. (2003). *Drosophila* BLM in double-strand break repair by synthesis-dependent strand annealing. Science *299*, 265-267.

Adzuma, K., Mizuuchi, K. (1988). Target immunity of *Mu* transposition reflects a differential distribution of MuB protein. Cell *53*, 257-266.

Adzuma, K., Mizuuchi, K. (1991). Steady-state kinetic analysis of ATP hydrolysis by the B protein of bacteriophage *Mu*. Involvement of protein oligomerization in the ATPase cycle. J Biol Chem *266*, 6159-6167.

Aravind, L., Landsman, D. (1998). AT-hook motifs identified in a wide variety of DNA-binding proteins. Nucl Ac Res *26*, 4413-4421.

Auge-Gouillou, C., Bigot, Y., Pollet, N., Hamelin, M.H., Rotival, M.M., Periquet, G. (1995). Human and other mammalian genomes contain transposons of the *mariner* family. FEBS Letters *368*, 541-546.

Auge-Gouillou, C., Hamelin,M.H., Demattei,M.V., Periquet,G., Bigot,Y. (2000). The ITR bnding domain of the *Mariner* Mos-1 transposase. Mol Genet Genomics *265*, 58-65.

Auge-Gouillou, C., Hamelin,M.H., Demattei,M.V., Periquet,M., Bigot,Y. (2001). The wild-type conformation of the *Mos1* inverted terminal repeats is suboptimal for transposition in bacteria. Mol Genet Genomics *265*, 51-57.

Bailin, T., Mo,X., Sadofsky,M.J. (1999). A RAG-1 and RAG-2 tetramer complex is active in cleavage in V(D)J recombination. Mol Cell Biol *19*, 4664-4671.

Baker, T. A., Mizuuchi,M., Mizuuchi,K. (1991). MuB protein allosterically activates strand transfer by transposase of phage *Mu*. Cell *65*, 1003-1013.

Beall, E. L. R., D.C. (1997). *Drosophila P*-element transposase is a novel site-specific endonuclease. Genes Dev *11*, 2137-2151.

Bellen, H. J., O'Kane, C.J., Wilson, C., Grossniklaus, U., Pearson, R.K., Gehring, W.J. (1989). *P*-element mediated enhancer detection: a versatile method to study development in *Drosophila*. Genes Dev *3*, 1288-1300.

Bender, J., and Kleckner,N. (1986). Genetic evidence that Tn*10* transposes by a nonreplicative mechanism. Cell *45*, 801-815.

Benjamin, H. W., and Kleckner, N. (1989). Intramolecular transposition by Tn*10*. Cell *59*, 373-383.

Benjamin, H. W., and Kleckner, N. (1992). Excision of Tn*10* from the donor site during transposition occurs by flush double-strand cleavages at the transposon termini. PNAS USA *89*, 4648-4652.

Benjamin, H. W., Cozzarelli, N.R. (1986). DNA-directed synapsis in recombination: slithering and random collision of sites. Proc Robert A Welch Found Conf Chem Res *29*, 107-129.

Besmer, E., Mansilla-Soto J., Cassard, S., Sawchuk, D.J., Brown, G., Sadofsky, M., Lewis, S.M., Nussenzweig, M.C., Cortes, P. (1998). Hairpin coding end opening is mediated by RAG1 and RAG2 proteins. Mol Cell *2*, 817-828.

Betermier, M., Alazard, R., Ragueh, F., Roulet, E., Toussaint, A., Chandler, M. (1987). Phage *Mu* transposase: deletion of the carboxy-terminal end does not abolish DNA-binding activity. Mol Gen Genet *210*, 77-85.

Bhasin, A., Goryshin, I.Y., Steininger-White, M., York, D., Reznikoff, W.S. (2000). Characterization of a Tn*5* pre-cleavage synaptic complex. J Mol Biol *302*, 49-63.

Bhasin, A., Goryshin,I.Y., Reznikoff,W.S. (1999). Hairpin formation in Tn*5* transposition. J Biol Chem *274*, 37021-37029.

Boeke, J. D., Corces, V.G. (1989). Transcription and reverse transcription of retrotransposons. Annu Rev Microbiol *43*, 403-434.

Bolland, S., Kleckner, N. (1995). The two single strand cleavages at each end of Tn*10* occur in a specific order during transposition. PNAS USA *92*, 7814-7818.

Bolland, S., Kleckner,N. (1996). The three chemical steps of Tn*10*/IS*10* transposition involve repeated utilisation of a single active site. Cell *84*, 223-233.

Boocock, M. R., Brown, J.L., Sherratt, D.J. (1986). Structural and catalytic properties of specific complexes between Tn*3* resolvase and the recombination sites res. Biochem Soc Trans *14*, 214-216.

Boocock, M. R., Brown, J.L., Sherratt, D.J. (1987). Topological specificity in Tn*3* resolvase catalysis. In DNA Replication and Recombination, T. J. Kelly, McMacken, R., ed. (New York, Alan R. Liss).

Bosco, G., Haber, J.E. (1998). Chromosome break-induced DNA replication leads to nonreciprocal translocations and telomere capture. Genetics *150*, 1037-1047.

Boulton, S. J., Jackson, S.P. (1996a). Identification of a *Saccharomyces cerevisiae* homologue: roles in DNA double strand rejoining and in telomeric maintenance. Nucl Ac Res *24*, 4639-4648.

Boulton, S. J., Jackson, S.P. (1996b). *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. EMBO J *15*, 5093-5103.

Bradbury, J. M., Jackson, S.P. (2003). The complex matter of DNA double-strand break detection. Biochem Soc Trans *31*, 40-44.

Brendel, M., Henriques, J.A. (2001). The pso mutants of *Saccharomyces cerevisiae* comprise two groups: one deficien in DNA repair and another with altered mutagen metabolism. Mutat Res *489*, 79-96.

Brookman et al., K. W., Lamerdin, J.E., Thelen, M.P., Hwang, M., Reardon, J.T., Sancar, A., Zhou, Z.Q., Walter, C.A., Parris, C.N., Thompson, L.H. (1996). ERCC4 (XPF) encodes a human nucleotide excision repair protein with eukaryotic recombination homologs. Mol Cell Biol *16*, 6553-6562.

Bryan, G., Garza,D., Hartl,D., (1990). Insertion and excision of the transposable element *mariner* in *Drosophila.* Genetics *125*, 103-114.

Bryan, G. J., Jacobson, J.W., Hartl, D.L. (1987). Heritable somatic excision of a *Drosophila* transposon. Science *235*, 1636-1638.

Bucheton, A., Busseau, I., Teninges, D. (2002). *I* elements in *Drosophila melanogaster.* In Mobile DNA II, N. L. Craig, Craigie, R., Gellert, M., and Lambowitz, A. M.,, ed. (Washington, D.C., ASM Press), pp. 796-812.

Bukhari, A. I. (1975). Reversal of mutator phage *Mu* integration. J Mol Biol *96*, 87-99.

Burke, M. E., Arnold, P.H., He, J., Wenwieser, S.V.C.T., Rowland, S.J., Boocock, M.R., Stark, W.M. (2004). Activating mutations of Tn*3* resolvase marking interfaces important in recombination catalysis and its regulation. Molec Microbiol *51*, 937-948.

Burke, W. D., Eickbush, D.G., Xiong, Y., Jakubczak, J., Eickbush, T.H. (1993). Sequence relationship of retrotransposable elements *R1* and *R2* within and between divergent insect species. Mol Biol Evol *10*, 163-185.

Burke, W. D., Malik, H.S., Lathe, W.C., Eickbush, T.H. (1998). Are retrotransposons long-term hitchhikers? Nature *392*, 141-142.

Butler, A. R., White, J.H., Stark, M.J.R. (1991). Analysis of the response of the *Saccharomyces cerevisiae* cells to *Kluyveromyces lactis* toxin. J Gen Microbiol *137*, 1749-1757.

Caizzi, R., Caggese, C., Pimpinelli, S. (1993). *Bari-1*, a new transposon-like family in *Drosophila melanogaster* with a unique heterochromatic organization. Genetics *133*, 335-345.

Carter, T., Vancurova, I., Sun, I., Lou, W., DeLeon, S. (1990). A DNA-activated protein kinase from HeLa cell nuclei. Mol Cell Biol *10*, 6460-6471.

Casadaban, M. J., Cohen, S.N. (1979). Lactose genes fused to exogenous promoters in one step using a *Mu*-lac bacteriophage: *in vivo* probe for transcriptional control sequences. PNAS USA *76*, 4530-4533.

Chalmers, R., Guhathakurta, H.B., Kleckner, N. (1998). IHF modulation of Tn*10* transposition: sensory transduction of supercoiling status via a proposed protein/DNA molecular spring. Cell *93*, 897-908.

Chang, E. C., Barr,M., Wang,Y., Jung,v., Xu,H.P., Wigler,M.H. (1994). Cooperative interaction of *S. pombe* proteins required for mating and morphogenesis. Cell *79*, 131-141.

Check, E. (2002). Gene therapy: a tragic setback. Nature *420*, 116-118.

Ciubotaru, M., Ptaszek, L.M., Baker, G.A., Baker, S.N., Bright, F.V., Schatz, D.G. (2003). RAG-1 DNA binding in V(D)J recombination. Specificity and DNA induced conformational change revealed by fluorescence and CD spectroscopy. J Biol Chem *278*, 5584-5596.

Clatworthy, A. E., Valencia,M.A., Haber,J.E., Oettinger,M.A., (2003). V(D)J recombination and RAG-mediated transposition in yeast. Molec Cell *12*, 484-499.

Coates, C. G., Jasinskiene, N., Miyashiro, L., James, A.A. (1998a). *Mariner* transposition and transformation of the yellow fever mosquito, *Aedes aegypti*. PNAS USA *95*, 3748-3751.

Coates, C. J., Jasinskiene, N., Miyashiro, L., James, A.A. (1998b). *Mariner* transposition and transformation of the yellow fever mosquito, *Aedes aegypti*. PNAS USA *95*, 3748-3751.

Coates, C. J., Turney, C.L., Frommer, M., O'Brochta, D.A., Atkinson, P.W. (1997). Interplasmid transposition of the *mariner* transposable element in non-drosophilid insects. Mol Gen Genet *253*, 728-733.

Coen, E. S., Robbins, T.P., Almeida, J., Hudson, A., Carpenter, R. (1989). Consequences and mechanism of transposition in *Antirrhinum majus*. In Mobile DNA, D. E. Berg, Howe, M.M., ed. (Washington, D.C., Ammerican Society for Microbiology), pp. 413-436.

Collins, J., Forbes, E., Anderson, P. (1989). The *Tc3* family of transposable genetic elements in *Caenorhabditis elegans*. Genetics *121*, 47-55.

Colloms, S. D., van Luenen, H.G.A.M., Plasterk, R.H.A. (1994). DNA binding activities of the *Caenorhabditis elegans Tc3* transposase. Nucl Ac Res *22*, 5548-5554.

Consortium, I. H. G. S. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Craig, N. L., Craigie, R., Gellert, M., and Lambowitz, A. M., (2002). Mobile DNA II (Washington DC, American Society for Microbiology).

Craigie, R., Arndt-Jovin, D.J., Mizuuchi, K. (1985a). A defined system for the DNA strand-transfer reaction at the initiation of bacteriophage *Mu* transposition: protein and DNA substrate requirements. PNAS USA *82*, 7570-7574.

Craigie, R., Mizuuchi, K. (1985b). Mechanism of transposition of bacteriophage *Mu*: structure of a transposition intermediate. Cell *41*, 867-876.

Craigie, R., Mizuuchi, K. (1987). Transposition of *Mu* DNA: joining of *Mu* to target DNA can be uncoupled from cleavage at the ends of *Mu*. Cell *51*, 493-501.

Czerny, T., Schaffner, G., Busslinger, M. (1993). DNA sequence recognition by Pax proteins: bipartite structure of the paired domain. Genes Dev *7*, 2048-2061.

Davies, D., Goryshin,I.Y., Reznikoff,W.S., Rayment,I., (2000). Three-dimensional structure of the Tn*5* synaptic complex transposition intermediate. Science *289*, 77-85.

Davies, D. R., Braam, L.M., Reznikoff, W.S., Rayment, I. (1999). The three-dimensional structure of a Tn*5* transposase-related protein determined to a 2.9A resolution. J Biol Chem *274*, 11904-11913.

Davis and Symington, A. P., Symington, L.S. (2001). The yeast recombinational repair protein Rad 59 interacts with Rad52 and stimulates single-strand annealing. Genetics *159*, 515-525.

Dawson, A., Finnegan,D.J. (2003). Excision of the Drosophila *mariner* transposon *Mos1*: comparison with bacterial transposition and V(D)J recombination. Molec Cell *11*, 225-235.

Day, Y., Kysela,B., Hanakahi,L.A., Manolis,K., Riballo,E., Stumm,M., Harville,T.O., West,S.C., Oettinger,M.A., Jeggo,P.A., (2003). Nonhomologous end joining and V(D)J recombination require an additional factor. PNAS USA *100*, 2462-2467.

Dewannieux, M., Dupressoir, A., Harper, F., Pierron, G., Heidmann, T. (2004). Identification of autonomous *IAP* LTR retrotransposons mobile in mammalian cells. Nat Genet *36*, 534-539.

diGuana, C., Lib,P., Riggs,P.D., Inouyeb, H. (1988). Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. Gene *67*, 21-30.

Doak, T. G., Doerder, F.P., John, C.L., Herrick, G. (1994). A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common " D35E" motif. PNAS USA *91*, 942-946.

Eastman, Q., Leu, T.M., Schatz, D.G. (1996). Initiation of V(D)J recombination in vitro obeying the 12/23 rule. Nature *380*, 85-88.

Eickbush, T. H. (2002). *R2* and related site-specific non-long terminal repeat retrotransposons. In Mobile DNA II, N. L. Craig, Craigie, R., Gellert, M., and Lambowitz, A. M.,, ed. (Washington, D.C., ASM Press), pp. 813-835.

Emmons, S. W., Yesner, L., Ruan, K., Katzenberg, D. (1983). Evidence for a transposon in *Caenorhabditis elegans*. Cell *32*, 55-65.

Engelman, A., Mizuuchi,K., Craigie,R. (1991). HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. Cell *67*, 1211-1221.

Fadool, J. M., Hartl, D.L., Dowling, J.E. (1998). Transposition of the *mariner* element from *Drosophila mauritiana* in zebrafish. PNAS USA *95*, 5182-5186.

Featherstone, C., Jackson, S.P. (1999). Ku, a DNA repair protein with multiple cellular functions? Mutat Res *434*, 3-15.

Ferguson, K. A. (1964). Starch-gel electrophoresis-application to the classification of pituitary proteins and polypeptides. Metabolism *13*, 985-1002.

Feschotte, C., Swamy, L., Wessler, S.R. (2003). Genome-wide analysis of *mariner*-like transposable elements in rice reveals complex relationships with stowaway miniature inverted repeat transposable elements (MITEs). Genetics *163*, 747-758.

Feschotte, C., Wessler, S.R. (2002). *Mariner*-like transposases are widespread and diverse in flowering plants. PNAS USA *99*, 280-285.

Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. Trends Genet *5*, 103-107.

Fire, A., Xu, S.Q., Montgomery, M.K. Kostas, S.A., Driver, S.E., Mello, C.C. (1998). Potent and specific enetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature *391*, 806-811.

Fishmann-Lobell, J., Haber, J.E. (1992). Removal of nonhomologous DNA ends in double-strand break recombination: the role of the yeast ultraviolet repair gene RAD1. Science *258*, 480-484.

Frank-Vaillant, M., marcand, S. (2001). NHEJ regulation by mating type is exercised through a novel protein Lif2p, essential to the ligase IV pathway. Genes Dev *15*, 3005-3012.

Garcia-Fernandez, J., Bayascas-Ramirez, J.R., Marfany, G., Munoz-Marmol, A.M., Casali, A., et al. (1995). High coly number of highly similar *mariner*-like transposons in planarian (Platyhelminthe): evidence for a trans-phyla horizontal transfer. Mol Biol Evol *12*, 421-431.

Gehring, W. J. (1993). Exploring the homeobox. Gene *135*, 215-221.

Gehring, W. J., Qian, Y.Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A.F., Resendez-Perez, D., Affolter, M., Otting, G., Wuthrich, K. (1994). Homeodomain-DNA recognition. Cell *78*, 211-223.

Gellert, M. (2002). V(D)J recombination: RAG proteins, repair factors, and regulation. Annu Rev Biochem *71*, 101-132.

Gilfillan, S., Dierich, A., Lemeur, M., Benoist, C., Mathis, D. (1993). Mice lacking TdT: mature animals with an immature lymphocyte repertoire. Science *261*, 1175-1178.

Gill, R., Heffron, F., Dougan, G., Falkow, S. (1978). Analysis of sequences transposed by complementation of two classes of transposition defective mutants of transposition element Tn*3*. J Bacteriol *136*, 742-756.

Godderz, L. J., Rahman, N.S., Risinger, G.M., Arbuckle, J.L., Rodgers, K.K. (2003). Self association and conformational properties of RAG1: implication for formation of the V(D)J recombinase. Nucl Ac Res *31*, 2014-2023.

Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T.E., Mann, M., Lieber, M.R. (1997). Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. Nature *388*, 428-429.

Greenwald, I. (1985). lin-12, a nematode homeotic gene, is homologous to a set of mammalian proteins that includes epidermal growth factor. Cell *43*, 583-590.

Grindley, N. D. F. (1983). Transposition of *Tc3* and related transposons. Cell *32*, 3-5.

Grindley, N. D. F. (2002). The movement of Tn*3*-like elements: transposition and cointegrate resolution. In Mobile DNA II, N. L. Craig, Craigie, R., Gellert, M., and Lambowitz, A. M.,, ed. (Washington, D.C., ASM Press), pp. 272-302.

Gueiros-Filho, F. J., Beverley, S.M. (1997). Trans-kingdom transposition of the *Drosophila* element *mariner* within the protozoan *Leishmania*. Science *276*, 17161719.

Haber, J. E. (1998). The many interfaces of Mre11. Cell *95*, 583-586.

Haber, J. E. (2000). Partners and pathways: repairing a double-strand break. Trends Genet *16*, 259-264.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol *166*, 557-580.

Haniford, D. B. (2002). Transposon Tn10. In Mobile DNA II, N. L. Craig, Craigie, R., Gellert, M., and Lambowitz, A. M.,, ed. (Washington DC, American Society for Microbiology Press).

Haren, L., Polard, B., Ton-Hoang, B., Chandler, M. (1998). Multiple oligomerisation domains in the IS*911* transposase: a leucine zipper motif is essential for activity. J Mol Biol *283*, 29-41.

Harshey, R. M., Getzoff, E.D., Baldwin, D.L., Miller, J.L., Chaconas, G. (1985). Primary structure of phage *Mu* transposase: homology to *Mu* repressor. PNAS USA *82*, 7676-7680.

Hartl, L. D., Lohe, A.R., Lozovskaya, E.R. (1997). MODERN THOUGHTS ON AN ANCYENT *MARINER*E: Function, Evolution, Regulation. Annu Rev Genet *31*, 337-358.

Herrmann, G., Lindahl, T., Schar, P. (1998). *Saccharomyces cerevisiae* LIF1: a function involved in DNA double-strand break repair related to mammalian XRCC4. EMBO J *17*, 4188-4198.

Hiom, K., Gellert, M. (1998a). Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination. Mol Cell *1*, 1011-1019.

Hiom, K., Gellert, M. (1998b). A stable RAG1-RAG2-DNA complex that is active in V(D)J cleavage. Cell *88*, 65-72.

Hiom, K., Melek,M., Gellert, M. (1998c). DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. Cell *94*, 463-470.

Horie, K., Yusa, K., Yae, K., Odajima, J., Fischer, S.E.J., Keng, V.W., Hayakawa, T., Mizuno, S., Kondoh, G., Ijiri, T., Matsuda, Y., Plasterk, R.H., Takeda, J. (2003). Characterisation of *Sleeping Beauty* transposition and its application to genetic screening in mice. Mol Cell Biol *23*, 9189-9207.

Houck, M. A., Clark, J.B., Peterson, K.R., Kidwell, M.G. (1991). Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. Science *253*, 1125-1129.

Howe, M. M., bade, E.G. (1975). Molecular biology of bacteriophage *Mu*. Science *190*, 624-632.

Hsu, E., Schwager, J., Alt, F.W. (1989). Evolution of immunoglobulin genes: V_H familiesin the amphibian *Xenopus*. PNAS USA *86*, 8010-8014.

Isberg, R. R., Lazaar, A.A., Syvanen, M. (1982). Regulation of Tn*5* by the right-repeat proteins: control at the level of the transposition reaction? Cell *30*, 883-892.

Ivics, Z., Hackett, P.B., Plasterk, R.H.A., Izsvak, Z. (1997). Molecular reconstruction of *Sleeping Beauty*, a *Tc1*-like transposon from fish, and its transposition in human cells. Cell *91*, 501-510.

Ivics, Z., Izsvak, Z. (2004). Transposable elements for transgenesis and insertional mutagenesis in vertebrates: a contemporary review of experimental strategies. Meth Mol Biol *260*, 255-276.

Ivics, Z., Izsvak, Z., Mintr, A., Hackett, P.B. (1996). Identification of functional domains and evolution of *Tc1*-like transposable elements. PNAS USA *93*, 5008-5013.

Izsvak, Z., ivics, Z., Plasterk, R.H. (2000). *Sleeping Beauty*, a wide host-range transposon vector for genetic transformation in vertebrates. J Mol Biol *302*, 93-102.

Izsvak, Z., Khare, D., Behlke, J., Heinemann, U., Plasterk, R., Ivics, Z. (2002). Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in *Sleeping Beauty* transposition. J Biol Chem *277*, 34581-34588.

Izsvak, Z., Stuwe,E.E., Fiedler,D., Katzer,A., Jeggo,P.A., Ivics,Z., (2004). Healing the wounds inflicted by *Sleeping Beauty* transposition by double-strand break repair in mammalian somatic cells. Molec Cell *13*, 279-290.

Jackson, S. P., MacDonald, J.J., Lees-Miller, S., Tjian, R. (1990). GC box binding induces phosphorylation of Sp1 by a DNA-dependent protein kinase. Cell *63*, 155-165.

Jacobson, J. W., Hartl, D.L. (1985). Coupled instability of two X-linked genes in *Drosophila mauritiana*: germinal and somatic mutability. Genetics *111*, 57-65.

Jacobson, J. W., Medhora,M.M., Hartl,D.L., (1986). Molecular structure of a somatically unstable transposable element in *Drosophila.* PNAS USA *83*, 8684-8688.

Jakubczak, J., Burke, W.D., Eickbush, T.H. (1991). Retrotransposable elements *R1* and *R2* interrupt the rDNA genes of most insects. PNAS USA *88*, 3295-3299.

Jeggo, P. A. (1998). DNA breakage and repair. Adv Genet *38*, 185-218.

Johnson, R. C., Reznikoff, W.S. (1984). Role of the IS*50*R proteins in the promotion and control of Tn*5* transposition. J Mol Biol *177*, 645-661.

Johnson, R. C., Yin, J.C., Reznikoff, W.S. (1982). Control of Tn*5* transposition in *Escherichia coli* is mediated by protein from the right repeat. Cell *30*, 873-882.

Jones, J. M., Gellert,M., (2002). Ordered assembly of the V(D)J synaptic complex ensures accurate recombination. EMBO J *21*, 4162-4171.

Junop, M. S., and Haniford,D.B. (1997). Factors responsible for target selection in Tn*10* transposition: a role for the DDE motif in target DNA capture. EMBO Journal *16*, 2646-2655.

Kaiser, J. (2003). Seeking the cause of induced leukemias in X-SCID trial. Science *299*, 457-608.

Kegel, A., Sjostrand, J.O., Astrom, S.U. (2001). Nej1p, a cell type-specific regulator of nonhomologous end joining in yeast. Curr Biol *11*, 1611-1617.

Kellerman, O. K., Ferenci,T. (1982). Maltose binding protein from *E. coli.* Methods Enzymol *90*, 459-463.

Kennedy, A. K., Guhathakurta, A., Kleckner, N. and Haniford, D.B. (1998). Tn*10* transposition via a DNA hairpin intermediate. Cell *95*, 125-134.

Ketting, R. F., Haverkamp, T.H.A., van Luenen, H.G.A.M., Plasterk, R.H.A. (1999). *mut-7* of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell *99*, 133-141.

Kilbride, E., Boocock, M.R., Stark,W.M. (1999). Topological selectivity of a hybrid site-specific recombination system with elements from Tn*3* res/resolvase and bacteriophage P1 loxP/Cre. J Mol Biol *289*, 1219-1230.

Kim, D. R., Oettinger, M.A. (1998). Functional analysis of coordinated cleavage in V(D)J recombination. Mol Cell Biol *18*, 4679-4688.

Kiyoshi, N., Perutz, M.F., Poyart, C. (1985). Oxygen binding properties of human mutant hemoglobins synthesized in *Escherichia coli*. PNAS USA *82*, 7252-7255.

Kleckner, N. (1979). DNA sequence analysis of Tn*10* insertions: origin and role of the 9-bp flanking repetitions during Tn*10* translocation. Cell *16*, 711-720.

Komori, T., Okada, A., Stewart, V., Alt, F.W. (1993). Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. Science *261*, 1171-1175.

Kruklitis, R., Nakai, H. (1994). Participation of bacteriophage MuA protein and host factors in the initiation of *Mu* DNA synthesis in vitro. J Biol Chem *269*, 16469-16477.

Kuo, C. F., Zou, A.H., Jayaram, M., Getzoff, E., Harshey, R. (1991). DNA-protein complexes during attachment-site synapsis in *Mu* DNA transposition. EMBO J *10*, 1585-1591.

Lafaille, J. J., DeCloux, A., Bonneville, M., Takagaki, Y., Tonegawa, S. (1989). Junctional sequences of T cell receptor gamma delta genes: implications for gamma delta T cell lineages and for a novel intermediate of V-(D)-J joining. Cell *59*, 859-870.

Lamberg, A., Nieminene, S., Qiao, M., Savilahti, H. (2002). Efficient insertion mutagenesis strategy for bacterial genomes involving electroporation of in vitro-assembled DNA transposition complexes of bacteriophage *Mu*. Appl Environm Microbiol *68*, 705-712.

Lambert, S., Mason, S.J., Barber, L.J., Hartley, J.A., Pearce, J.A., Carr, A.M., McHugh, P.J. (2003). *Schizosaccharomyces pombe* checkpoint response to DNA interstrand cross-links. Mol Cell Biol *23*, 4728-4737.

Lampe, D. J., Churchill,M.E.A., Robertson,H.M., (1996). A purified *mariner* transposase is sufficient to mediate transposition *in vitro*. EMBO J *15*, 5470-5479.

Lander, E. S., linton, L.M., Birren, B., Nusbaum, C., Zody, M., et. al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Landree, M. A., Kale, S.B., Roth, D.B. (2001). Functional organization of single and paired V(D)J cleavage complexes. Mol Cell Biol *21*, 4256-4264.

Landree, M. A., Wibbenmeyer, J.A., Roth, D.B. (1999). Mutational analysis of RAG1 and RAG2 identifies three catalytic amino0acids in RAG1 crytical for both cleavage steps of V(D)J recombination. Genes Dev *13*, 3059-3069.

Langin, T., Capy, P., Daboussi, M.J. (1995). The transposable element *impala*, a fungal member of the *Tc1-mariner* superfamily. Mol Gen Genet *246*, 19-28.

Lavoie, B. D., Chan, B.S., Allison, R.G., Chaconas, G. (1991). Structural aspects of a higher order nucleoprotein complex: induction of an altered DNA structure at the *Mu*-host junction of the *Mu* Type 1 transpososome. EMBO J *10*, 3051-3059.

Leung, P. C., Harshey, R.M. (1991). Two mutations of phage *Mu* transposase that affect strand transfer or interactions with B protein lie in distinct polypeptide domains. J Mol Biol *219*, 189-199.

Levchenko, I., Luo, L., Baker, T.A. (1995). Disassembly of the *Mu* transposase tetramer by the ClpX chaperone. Genes Dev *9*, 2399-2408.

Lewis, S., Gifford, A., Baltimore, D. (1985). DNA elements are assymetrically joined during the site-specific recombination of kappa immunoglobulin genes. Science *228*, 677-685.

Li, Z., Otevrel, T., Gao, Y., Cheng, H.L., Seed, B., Stamato, T.D., Taccioli, G.E., Alt, F.W. (1995). The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. Cell *83*, 1079-1089.

Liao, L. W., Rosenzweig, B., Hirsh, D. (1983). Analysis of a transposable element in *Caenorhabditis elegans*. PNAS USA *80*, 3585-3589.

Lieber, M. R., Hesse, J.E., Mizuuchi, K., Gellert, M. (1988). Lymphoid V(D)J recombination: nucleotide insertion at signal joints as well as coding joints. PNAS USA *85*, 8588-8592.

Lipkow, K., Buisine, N., Chalmers, R. (2004a). Promiscuous target interactions in the *mariner* transposon *Himar1*. J Biol Chem 279, 48569-48575.

Lipkow, K., Buisine, N., Lampe, D.J., Chalmers, R. (2004b). Early intermediates of *mariner* transposition: catalysis without synapsis of the transposon ends suggests a novel architecture of the synaptic complex. Molec Cell Biol *24*, 8301-8311.

Litman, G. W., Rast, J.P., Shamblott, M.J., Haire, R.N., Hulst, M., Roess, W., Litman, R.T., Hinds-Frey, K.R., Zilch, A., Amemiya, C.T. (1993). Phylogenetic diversification of immunoglobulin genes and the antibody repertoire. Mol Biol Evol *10*, 60-72.

Lohe, A. R., DeAguiar,D., Hartl,D.L. (1997). Mutations in the *mariner* transposase: the D,D(35)E consensus sequence is non-functional. Proc Natl Acad Sci USA *94*, 1293-1297.

Lohe, A. R., Hartl, D.L. (1996a). Autoregulation of *mariner* transposase activity by overproduction and dominant-negative complementation. Mol Biol Evol *13*, 549-555.

Lohe, A. R., moriyama, E.N., Lidholm, D.A., Hartl, D.L. (1995). Horizontal transmission, vertical inactivation, and stochastic loss of *mariner*-like transposable elements. Mol Biol Evol *12*, 62-72.

Lohe, A. R., Sullivan, D.T., Hartl, D.L. (1996b). Subunits interactions in the *mariner* transposase. Genetics *144*, 1087-1095.

Lovell, S., Goryshin, I.Y., Reznikoff, W.R., Rayment, I. (2002). Two-metal active site binding of a Tn*5* transposase synaptic complex. Nat Struct Biol *9*, 278-281.

Luo, G., ivics, Z., Izsvak, Z., Bradley, A. (1998). Chromosomal transposition of a *Tc1/mariner*-like element in mouse embryonic stem cells. PNAS USA *95*, 10769-10773.

Ma, Y., Pannicke,U., Schwarz,K., Lieber,M.R., (2002). Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. Cell *108*, 781-794.

Marshall, E. (1999). Gene therapy death prompts review of adenovirus vector. Science *286*, 2244-2245.

Maruyama, K., Hartl, D.L. (1991). Evidence for interspecific transfer of the transposable element *mariner* between *Drosophila* and *Zaprionus*. J Mol Evol *33*, 514-524.

Maxwell, A., Craigie, R., Mizuuchi, K. (1987). B protein of bacteriophage *Mu* is an ATPase that preferentially stimulates intermolecular DNA strand transfer. PNAS USA *84*, 699-703.

Maynard-Smith, S., Leach, D., Coelho, A., Carey, J., Symonds, N. (1980). The isolation and characteristics of plasmids derived from the insertion of MupAp1 into pML2: their behaviour during transposition. Plasmid *4*, 34-50.

McBlane, J. F., van Gent,D.C., Ramsden,D.A., Romeo,C., Cuomo,C.A., Gellert,M., Oettinger,M.A. (1995). Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. Cell *83*, 387-395.

McClintock, B. (1950). The origin and behaviour of mutable loci in maize. PNAS USA *36*, 344-349.

McCormack, W. T., Tjoelker, L.W., Carlson, L.M., Petryniak, B., Barth, C.F., Humphries, E.H., Thompson, C.B. (1989). Chicken IgL gene rearrangement involves deletion of a circular episome and addition of single non-random nucleotides to both coding segments. Cell *56*, 785-791.

Medhora, M. M., MacPeek,A.H., Hartl,D.L., (1988). Excision of the *Drosophila* transposable element *mariner*: identification and characterization of the *Mos* factor. EMBO J *7*, 2185-2189.

Medhora, M. M., Maruyama,K., Hartl,D., (1991). Molecular and functional analysis of the *mariner* mutator element *Mos1* in Drosophila. Genetics *128*, 311-318.

Mezard, C., Pompon, D., Nicolas, A. (1992). Recombination between similar but not identical DNA sequences during yeast transformation occurs within short stretches of identity. Cell *70*, 659-670.

Mhammedi-Alaoui, A., Pato, M., gama, M.J., Touissant, A. (1994). A new component of bacteriophage *Mu* replication transposition machinery: the *Escherichia coli* ClpX protein. Mol Microbiol *11*, 1109-1116.

Miller, J. L., Anderson, S.K., Fujita, D.J., Chaconas, G., Baldwin, D., Harshey, R.M. (1984). The nucleotide sequence of the B gene of bacteriophage *Mu*. Nucl Ac Res *12*, 8627-8638.

Milne, G. T., Jin, S., Shannon, K.B., Weaver, D.T. (1996). Mutations in two Ku homologs define a DNA end-joining repair pathway in *Saccharomyces cerevisiae*. Mol Cell Biol *16*, 4189-4198.

Miskey, C., Izsvak, Z., Plasterk, R.H., Ivics, Z. (2003). The *Frog Prince*: a reconstructed transposon from *Rana pipiens* with high transpositional activity in vertebrate cells. Nucl Ac Res *31*, 6873-6881.

Mizuuchi, K. (1992a). Transpositional recombination: mechanistic insights from studies of *Mu* and other elements. Annu Rev Biochem *61*, 1011-1051.

Mizuuchi, M., Baker, T.A., Mizuuchi, K. (1992b). Assembly of the active form of the transposase-*Mu* DNA complex: a critical control point in *Mu* transposition. Cell *70*, 303-311.

Mizuuchi, M., Mizuuchi, K. (1989). Efficient *Mu* transposition requires interaction of transposase with a DNA sequence at the *Mu* operator: implications for regulation. Cell *58*, 399-408.

Moerman, D. G., Benian, G.M., Waterston, R.H. (1986). Molecular cloning of the muscle gene unc-22 in *Caenorhabditis elegans* by *Tc1* transposson tagging. PNAS USA *83*, 2579-2583.

Mombaerts, P., Iacomini, J., Johnson, R.S., Herrup, K., Tonegawa, S., Papaioannou, V.E. (1992). RAG-1 deficient mice have no mature B and T lymphocites. Cell *68*, 869-877.

Montini, E., et al. (2002). *In vivo* correction of murine tyrosinemia type I by DNA-mediated transposition. Mol Ther *6*, 759-769.

Moore, J. K., Haber, J.E. (1996). Cell cycle and genetic requirements of two pathways of non-homologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. Mol Cell Biol *16*, 2164-2173.

Moran, J. V., Gilbert, N. (2002). Mammalian LINE-1 retrotransposons and related elements. In Mobile DNA II, N. L. Craig, Craigie, R., Gellert, M., and Lambowitz, A. M.,, ed. (Washington, D.C., ASM Press), pp. 836-869.

Moreira, L. A., Edwards, M.J., Adhami, F., Jasinskiene, N., James, A.A., Jacobs-Lorena, M. (2000). Robust gut-specific gene expression in transgenic *Aedes aegypti* mosquitoes. PNAS USA *97*, 10895-10898.

Morgan, G. T. (1995). Identification in the human genome of mobile elements spread by DNA-mediated transposition. J Mol Biol *254*, 1-5.

Morisato, D., Kleckner, N. (1984). Transposase promotes double strand breaks and single strand joints at Tn*10* termini in vivo. Cell *39*, 181-190.

Moshous, D., Callebaut, I., de Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., Tezcan, I., Sanal, O., Bertrand, Y., Phillipe, N., Fischer, A., de Villartay, J.P. (2001). Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immunne deficiency. Cell *105*, 177-186.

Mueller, J. E., Smith, D., Bryk, M., Belford, M. (1995). Intron-encoded endonuclease I-TevI binds as a monomer to effect sequential cleavage via conformational changes in the td homing site. EMBO J *14*, 5724-5735.

Mundy, C. L., Patenge,N., Matthews, A.G.W., Oettinger,M.A., (2002). Assembly of the RAG1/RAG2 synaptic complex. Molecular and Cellular Biology *22*, 69-77.

Nagai, K., Thogersen,H.C. (1984). Generation of beta-globin by sequence specific proteolysis of a hybrid protein produced in *Escherichia coli.* Nature reviews *309*, 810-812.

Nakayama, C., Teplow, D.B., Harshey, R.M. (1987). Structural domains in phage *Mu* transposase: identification of site-specific DNA-binding domain. PNAS USA *84*, 1809-1813.

Namgoong, S. Y., Harshey, R.M. (1998). The same two monomers within a Mu A tetramer provide the DDE domains for the strand cleavage and strand transfer steps of transposition. EMBO J *17*, 3775-3785.

Nassif, N., Penney, J., Pal, S., Engels, W.R., Gloor, G.B. (1994). Efficient copying of nonhomologous sequences from ectopic sites via *P*-element-induced gap repair. Mol Cell Biol *14*, 1613-1625.

Naumann, T. A., Reznikoff, W.S. (2000). Trans catalysis in Tn*5* transposition. PNAS USA *97*, 8944-8949.

Normand, C., Duval-Valentin,G., Haren,L., Chandler,M. (2001). The terminal inverted repeats of IS*911*: requirements for synaptic complex assembly and activity. J Mol Biol *308*, 853-871.

Oettinger, M. A., Schatz,D.G., Gorka,C., Baltimore,D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science *248*, 1517-1523.

Ohlfest, J. R., Frandsen, J.L., Fritz, S., Lobitz, P.D., Perkinson, S.G., Clark, K.J., Nelsestuen, G., Key, N.S., McIvor, R.S., Hackett, P.B., Largaespada, D.A. (2004a). Phenotypic correction and long-term expression of Factor VIII in hemophilic mice by immunotolerization and nonviral gene transfer using the *Sleeping Beauty* transposon system. Blood Epub December *2004*, 1-31.

Ohlfest, J. R., Lobitz, P.D., Perkinson, S.G., Laraespada, D.A. (2004b). Integration and long-term expression in xenografted human glioblastoma cells using a plasmid-based transposon system. Mol Ther *10*, 260-268.

Ooi, S. L., Shoemaker, D.D., Boeke, J.D. (2001). A DNA microarray-based genetic screen for nonhomologous end-joining mutants in *Saccharomyces cerevisiae*. Science *294*, 2552-2556.

Oosumi, T. (1995). *Mariner* transposons in humans. Nature *378*, 672.

Orchard, K., May, G.E. (1993). An EMSA-based method for determining the molecular weight of a protein-DNA complex. Nucl Ac Res *21*, 3335-3336.

Ortiz-Urda, S., et al. (2003). Sustainable correction of junctional epidermolysis bullosa via transposon-mediated nonviral gene transfer. Gene Ther *10*, 1099-1104.

Paques, F., Haber, J.E. (1999). Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. Microbiol Molec Biol Rev *63*, 349-404.

Pierce, A. J., Hu, P., Han, M., Ellis, N., Jasin, M. (2001). Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mammalian cells. Genes Dev *15*, 3237-3242.

Plasterk, R. H. A., Izsvak, Z., ivics, Z. (1999). Resident aliens: the *Tc1/mariner* superfamily of transposable elements. Trends Genet *15*, 326-332.

Priimagi, A. F., Mizrokhi, L.J., Ilyin, Y.V. (1988). The *Drosophila* mobile element jockey belongs to LINEs and contains coding sequences homologous to some retroviral proteins. Gene *70*, 253-262.

Reed, R. R. (1981a). Transposon-mediated site-specific recombination: a defined in vitro system. Cell *25*, 713-719.

Reed, R. R., Grindley, N.D.F. (1981b). Transposon-mediated site-specific recombination in vitro: DNA cleavage and protein-DNA linkage at the recombination site. Cell *25*, 721-728.

Resnick, M. A., Martin, P. (1976). The repair of double-strand breaks in DNA: A model involving recombination. J Theoret Biol *59*, 97-106.

Rice, P. A., Steitz, T.A. (1994a). Model for a DNA-mediated synaptic complex suggested by crystal packing of gamma delta resolvase subunits. EMBO J *13*, 1514-1524.

Rice, P. A., Steitz, T.A. (1994b). Refinement of gamma delta resolvase reveals a strikingly flexible molecule. Structure *2*, 371-384.

Robertson H.M., m., E.G. (1993). Five major subfamilies of *mariner* transposable elements in insects, including the Mediterranean fruit fly, and related artropods. Insect Mol Biol *2*, 125-139.

Robertson, H. M. (1993). The *mariner* element is widespread in insects. Nature *362*, 241-245.

Robertson, H. M. (1995a). The *Tc1-mariner* superfamily of transposons in animals. J Insect Physiol *41*, 99-105.

Robertson, H. M. (1997). Multiple *mariner* transposons in flatworms and hydras are related to those of insects. J Hered *88*, 195-201.

Robertson, H. M., Lampe, D.J. (1995b). Recent horizontal transfer of a *mariner* transposable element among and between *Diptera* and *Neuroptera*. Mol Biol Evol *12*, 850-862.

Rodgers, K. K., Villey, I.J., Ptaszek, L., Corbett, E., Schatz, D.G., Coleman, J.E. (1999). A dimer of the lymphoid protein RAG1 recognizes the recombination signal sequence and the complex stably incorporates the high mobility group protein HMG2. Nucl Ac Res *27*, 2938-2946.

Rose, M. D., Novick, P., Thomas, J.H., Botstein, D., Fink, G.R. (1987). A *Saccharomyces cerevisiae* genomic plasmid bank based on a centromere-containing shuttle vector. Gene *60*, 237-243.

Roth, D. B., Porter, T.N., Wilson, J.H. (1985). Mechanisms of nonhomologous recombination in mammalian cells. Mol Cell Biol *5*, 2599-2607.

Rothkamm, K., Kruger, I., Thompson, L.H., Lobrich, M. (2003). Pathways of DNA double-strand break repair during the mammalian cell cycle. Mol Cell Biol *23*, 5706-5715.

Rubin, G. M., Spradling, A.C. (1982). Genetic transformation of *Drosophila* with transposable element vectors. Science *218*, 348.

Sakai, J., and Kleckner,N. (1997). The Tn*10* synaptic complex can capture a target DNA only after transposon excision. Cell *89*, 205-214.

Sakai, J., Chalmers,R.M., and Kleckner,N. (1995). Identification and characterization of a pre-cleavage synaptic complex that is an early intermediate in Tn*10* transposition. EMBO J *14*, 4374-4383.

Sakai, J. S., Kleckner,N., Yang,X., and Guhathakurta,A. (2000). Tn*10* transpososome assembly involves a folded intermediate that must be unfolded for target capture and strand transfer. EMBO J *19*, 776-785.

Sanderson, M. R., Freemont, P.S., Rice, P.A., Goldman, A., Hatfull, G.F., Grindley, N.D., Steitz, T.A. (1990). The crystal structure of the catalytic domain of the site-specific recombination enzyme gamma delta resolvase at 2.7 A resolution. Cell *63*, 1323-1329.

Sandmeyer, S. S., Aye, M., Menees, T. (2002). *Ty3*, a position-specific, *gypsy*-like element in *Saccharomyces cerevisiae*. In Mobile DNa II, N. L. Craig, Craigie, R., Gellert, M., and Lambowitz, A. M.,, ed. (Washington, D.C., ASM Press), pp. 663-683.

Sarkis, G. J., Murley, L.L., Leschziner, A.E., Boocock, M.R., Stark, W.M., Grindley, N.D.F. (2001). A model for the gamma-delta resolvase synaptic complex. Mol Cell *8*, 623-631.

Savilahti, H., Mizuuchi,K., (1996). *Mu* transpositional recombination: donor DNA cleavage and strand transfer in trans by the *Mu* transposase. Cell *85*, 271-280.

Savilahti, H., Rice, P.A., Mizuuchi, K. (1995). The phage *Mu* transpososome core: DNA requirements for assembly and function. EMBO J *14*, 4893-4903.

Schar, P., Herrmann, G., Daly, G., Lindahl, T. (1997). A newly identified DNA ligase of *Saccharomyces cerevisiae* involved in RAD52-independent repair of DNA double-strand breaks. Genes Dev *11*, 1912-1924.

Schumann, W., Bade, E.G. (1979). *In vitro* constructed plasmids containing both ends of bacteriophage *Mu* DNA express phage functions. Mol Gen Genet *169*, 97-105.

Sedensky, M. M., Hudson, S.J., Everson, B., Morgan, P.H. (1994). Identification of a *mariner*-like repetitive sequence in C. elegans. Nucl Ac Res *22*, 1719-1723.

Segall, A. M. (1998). Analysis of higher order intermediates and synapsis in the bent-L pathway of bacteriophage lambda site-specific recombination. J Biol Chem *273*, 24258-24265.

Serebriiskii, I., Khazak, V., Golemis, E. A. (1999). A two-hybrid dual bait system to discriminate specificity of protein interactions. J Biol Chem *274*, 17080-17087.

Shapiro, J. A. (1979). Molecular model for the transposition and replication of bacteriophage *Mu* and other transposable elements. PNAS USA *76*, 1933-1937.

Sherman, A., Dawson, A., Mather, C., Gilhooley, H., Li, Y., Mitchell, R., Finnegan, D.J., Helen, S. (1998). Transposition of the Drosophila element *mariner* into the chicken germ line. Nature Biotech *16*, 1050-1053.

Shinkai, Y., Rathburn, G., Lam, K.P., Oltz, E.M., Stewart, V., Mendelsohn, M., Charron, J., Datta, M., Young, F., Stall, A.M., et. al. (1992). RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. Cell *68*, 855-867.

Sijbers et al.,, A. M., van Der Spek, P.J., Odijk, H., van der Berg, J., van Duin et al., M. (1996). Mutational analysis of the human nucleotide excision repair gene ERCC1. Nucl Ac Res *24*, 3370-3380.

Sikorski, R. S., Hieter,P., (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics *122*, 19-27.

Smit, A. F., Riggs, A.D. (1996). *Tiggers* and other DNA transposon fossils in the human genome. PNAS USA *93*, 1443-1448.

Smith, G. C. M., Jackson, S.P. (1999). The DNA-dependent protein kinase. Genes Dev *13*, 916-934.

Sokolsky, T. D., Baker, T.A. (2003). DNA gyrase requirements distinguish the alternate pathways of *Mu* transposition. Mol Microbiol *47*, 397-409.

Spradling, A. C., Rubin, G.M. (1982). Transposition of cloned *P* elements into *Drosophila* germline chromosomes. Science *218*, 341.

Spradling, A. C., Stern, D.M., Kiss, I., Roote, J., Laverty, T., Rubin, G.M. (1995). Gene disruptions using *P* transposable elements: an integral component of the *Drosophila* genome project. PNAS USA *92*, 10824-10830.

Stark, W. M., Boocock, M.R. (1995). Topological selectivity in site-specific recombination. In Mobile Genetic Elements, D. J. Sherat, ed. (New York, Oxford University Press), pp. 101-129.

Stark, W. M., Boocock, M.R., Sherratt, D.J. (1992). Catalysis by site-specific recombinases. Trends Genet *8*, 432-439.

Stark, W. M., Parker, C.N., Halford, S.E., Boocock, M.R. (1994). Stereoselectivity of DNA catenane fusion by resolvase. Nature *368*, 76-78.

Stark, W. M., Sherratt, D.J., Boocock, M.R. (1989). Site-specific recobination by Tn*3* resolvase: topological changes in the forward and reverse reactions. Cell *58*, 779.

Steiniger-White, M., Bhasin, A., Lovell, S., Rayment, I., Reznikoff, W.S. (2002). Evidence for "unseen" transposase-DNA contacts. J Mol Biol *322*, 971-982.

Steitz, T. A., Smerdon, S.J., Jager, J., Joyce, C.M. (1994). A unified polymerase mechanism for non-homologous DNA and RNA. Science *266*, 2022-2025.

Sugawara et al., N., Ira, G., Haber, J.E. (2000). DNA length dependence of the single-strand annealing pathway and the role of Saccharomyces cerevisiae RAD59 in double-strand break repair. Mol Cell Biol *20*, 5300-5309.

Surette, M. G., Buch, S.J., Chaconas, G. (1987). Transpososomes: stable protein-DNA complexes involved in the *in vitro* transposition of bacteriophage *Mu* DNA. Cell *49*, 253-262.

Surette, M. G., Chaconas, G. (1992). The *Mu* transpositional enhancer can function in trans: requirement of the enhancer for synapsis but not strand cleavage. Cell *68*, 1101-1108.

Swanson, P. C. (2002a). Fine structure and activity of discrete RAG-HMG complexes on V(D)J recombination signals. Mol Cell Biol *22*, 1340-1351.

Swanson, P. C. (2002b). A RAG-1/RAG-2 tetramer supports 12/23-regulated synapsis, cleavage, and transposition of V(D)J recombination signals. Mol Cell Biol *22*, 7790-7801.

Swanson, P. C., Desiderio,S. (1999). RAG-2 promotes heptamer occupancy by RAG-1 in the assembly of V(D)J initiation complex. Mol Cell Biol *19*, 3674-3683.

Szostak, J. W., Orr-Weaver, T.L., Rothstein, R.J., Stahl, F.W. (1983). The double-strand-break repair model for recombination. Cell *33*, 25-35.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., Mello, C.C. (1999). The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. Cell *99*, 123-132.

Takata, M., Sasak, M.S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A., Takeda, S. (1998). Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. EMBO J *17*, 5497-5508.

Tavakoli, N. P., DeVost, J., Derbyshire, K.M. (1997). Defining functional regions of the IS*903* transposase. J Mol Biol *274*, 491-504.

Taylor, A. L. (1963). Bacteriophage-induced mutation in *E. coli*. PNAS USA *50*, 1043-1051.

Teo, S. H., Jackson, S.P. (1997). Identification of *Saccharomyces cerevisiae* DNA ligase IV: involvement in DNA double-strand break repair. EMBO J *16*, 4788-4795.

Thomas, C. E., Ehrhardt, A., Kay, M.A. (2003). Progress and problems with the use of viral vectors for gene therapy. Nature Rev Genet *4*, 346-358.

Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature *302*, 575-581.

Tsukamoto, Y., Kato, J., Ikeda, H. (1996). Effects of mutations of RAD50, RAD51, RAD52, and related genes on illegitimate recombination in *Saccharomyces cerevisiae*. Genetics *142*, 383-391.

Tsukamoto, Y., Kato, J., Ikeda, H. (1997). Silencing factors participate in DNA repair and recombination in *Saccharomyces cerevisiae*. Nature *388*, 900-903.

Turlan, C., Chandler, M. (2000). Playing second fiddle: second strand processing and liberation of transposable elements from donor DNA. Trends Microbiol *8*, 268-274.

Valencia, M., Bentele, M., Vaze, M.B., Herrmann, G., Kraus, E., Lee, S.E., Schar, P., Haber, J.E. (2001). NEJ1 controls non-homologous end joining in *Saccharomyces cerevisiae*. Nature *414*, 666-669.

van Duin et al., M., de Wit, J., Odijk, H., Westerveld, A., Yasui, A., Koken, H.M., Hoeijmakers, J.H., Bootsma, D. (1986). Molecular characterization of the human excision repair gene ERCC1: cDNA cloning and amino acid homology with the yeast DNA repair gene RAD 10. Cell *44*, 913-923.

van Gent, D. C., Hiom, K., Paull, T.T., Gellert, M. (1997). Stimulation of V(D)J cleavage by high mobility group proteins. EMBO J *16*, 2665-2670.

van Gent, D. C., Ramsden, D.A., Gellert, M. (1996). The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. Cell *85*, 107-113.

van Luenen, H. G. A. M., Colloms, S.D., Plasterk, R.H.A. (1994). The mechanism of transposition of *Tc3* in *C. elegans*. Cell *79*, 293-301.

van Pouderoyen, G., Ketting, R.F., Perrakis, A., Plasterk, R.H.A., Sixma, T.K. (1997). Crystal structure of the specific DNA-binding domain of *Tc3* transposase of *C. elegans* in complex with transposon DNA. EMBO J *16*, 6044-6054.

vanGent, D. C., Hiom,K., Paull,T.T., Gellert,M. (1997). Stimulation of V(D)J cleavage by high mobility group proteins. EMBO J *16*, 2665-2670.

Vastenhouw, N. L., Plasterk, R.H.A. (2004). RNAi protects the *Caenorhabditis elegans* germline against transposition. Trends Genet *20*, 314-319.

Venter, J. C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., et. al. (2001). The sequence of the human genome. Science *291*, 1304-1351.

Ventura-Holdman, T., Lobb, C.J. (2001). Structural organization of the immunoglobulin heavy chain locus in the channel catfish: the IgH locus represents a composite of two gene clusters. Mol Immunol *38*, 557-564.

Vos, J. C., De Baere, I., Plasterk, R.H.A. (1996). Transposase is the only nematode protein required for *in vitro* transposition of *Tc1*. Genes Dev *10*, 755-761.

Vos, J. C., van Luenen, H.G.A.M., Plasterk, R.H.A. (1993). Characterization of the *Caenorhabditis elegans Tc1* transposase *in vivo* and *in vitro*. Genes Dev *7*, 1244-1253.

Voytas, D. F., Boeke, J.D. (2002). *Ty1* and *Ty5* of *Saccharomyces cerevisiae*. In Mobile DNA II, N. L. Craig, Craigie, R., Gellert, M., and Lambowitz, A. M.,, ed. (Washington, D.C., ASM Press), pp. 631-662.

Watanabe, T., Furuse, C. and Sakaizumi, S. (1968). Transduction of various R factors by phage P1 in *Escherichia coli* and by phage P22 in *Salmonella typhimurium*. Journal Bacteriol *96*, 1791-1795.

Watkins, S., van Pouderoyen, G., Sixma, T.K. (2004). Structural analysis of the bipartite DNA-binding domain of *Tc3* transposase bound to transposon DNA. Nucl Ac Res *32*, 4306-4312.

Wearne, S. J. (1990). Factor Xa cleavage of fusion proteins: elimination of non-specific cleavage by reversible acylation. FEBS Lett *263*, 23-26.

Wei, S. Q., Mizuuchi, K., Craigie, R. (1997). A large nucleoprotein assembly at the ends of the viral DNA mediates retroviral DNA integration. EMBO J *16*, 7511-7520.

Weil, C. F., Kunze, R. (2000). Transposition of maize *Ac/Ds* transposable elements in the yeast *Saccharomyces cerevisiae*. Nature Genet *26*, 187-190.

Weinreich, M. D., Gasch, A., Reznikoff, W.S. (1994a). Evidence that the cis preference of the Tn*5* transposase is caused by non-productive multimerisation. Genes Dev *8*, 2363-2374.

Weinreich, M. D., Mahnke-Braam, L., Reznikoff, W.S. (1994b). A functional analysis of the Tn*5* transposase. Identification of domains required for DNA binding and multimerization. J Mol Biol *241*, 166-177.

West, R. B., Lieber, M.R. (1998). The RAG-HMG1 complex enforces the 12/23 rule of V(D)J recombination specifically at the double-hairpin formation step. Mol Cell Biol *18*, 6408-6415.

Wijffelman, C., Lotterman, B. (1977). Kinetics of *Mu* DNA synthesis. Mol Gen Genet *151*, 169-174.

Wilkins, M. R., Lindskog, I., Gasteiger, E., Bairoch, A., Sanchez, J-C., Hochstrasser, D.F., Appel, R.D. (1997). Detailed peptide characterisation using PEPTIDEMASS-a World-Wide webb accessible tool. Electrophoresis *18*, 403-408.

Williams, T. L., Jackson,E.L., Carritte,A., Baker,T.A., (1999). Organization and dynamics of the *Mu* transpososome: recombination by communication between two active sites. Genes Dev *13*, 2725-2737.

Wilson, T. E., Grawunder, U., Lieber, M.R. (1997). Yeast DNA ligase IV mediates non-homologous DNA end joining. Nature *388*, 495-498.

Yang, J. Y., Jayaram, M., Harshey, R.M. (1996). Positional information within the *Mu* transposase tetramer: catalytic contributions of individual monomers. Cell *85*, 447-455.

Yang, J. Y., Kim, K., Jayaram, M., Harshey, R.M. (1995a). A domain sharing model for active site assembly within the MuA tetramer during transposition: the enhancer may specify domain contributions. EMBO J *14*, 2374-2384.

Yang, W., Steitz, T.A. (1995b). Crystal structure of the site-specific recombinase gamma delta resolvase complexed with a 34 bp cleavage site. Cell *82*, 193-207.

Yant, S. R., et al (2000). Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. Nat Genet *25*, 35-41.

Yant, S. R., Kay,M.A., (2003). Nonhomologous-end-joining factors regulate DNA repair fidelity during *Sleeping Beauty* element transposition in mammalian cells. Molec Cell Biol *23*, 8505-8518.

Yu, J., Marshall,K., Yamaguchi,M., Haber,J.E., Weil,C.F. (2004). Microhomology-dependent end joining and repair of transposon-induced DNA hairpins by host factors in *Saccharomyces cerevisiae*. Molec Cell Biol *24*, 1351-1364.

Yu, K., Lieber, M.R. (2000). The nicking step in V(D)J recombination is independent of synapsis: implications for the immune repertoire. Mol Cell Biol *20*, 7914-7921.

Zhang, L., Dawson,A., Finnegan,D.J. (2001). DNA-binding activity and subunit interaction of the *mariner* transposase. Nucl Ac Res *29*, 3566-3575.

Zhang, L., Sankar, U., Lampe, D.J., Robertson, H.M., Graham, F.L. (1998). The *Himar1 mariner* transposase cloned in a recombinant adenovirus vector is functional in mammalian cells. Nucl Ac Res *26*, 3687-3693.