DNA Binding Activity and Subunit Interactions of the *Mariner* Transposase

Lei ZHANG

A thesis presented for the degree of PhD

University of Edinburgh

2001

I dedicate this thesis to my family.

DECLARATION

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

Lei ZHANG

January , 2001

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Professor David J. Finnegan for all his advice, encouragement and patience during the course of my PhD. David introduced me to molecular biology. I have learned more than I could have imagined, and had a fine time doing it. Thanks for giving me the chance. I would also like to appreciate my second supervisor, Dr. David Leach for his advice and support.

Many thanks to all of the past and present members of the Finnegan lab who provided help, encouragement and friendship: Angela Dawson, Eve Hartswood, Ivan Clark, Laura Bisoni, Mary Canning, David Blesa.

I would like to thank Rob van Nues and Alan Colley in Jean Begg's lab for their advice on my yeast two-hybrid work.

Thanks should also go to my Chinese friends in ICMB: Xinsheng, Debiao, Suling, Guowen and Hongmei, for their friendship and encouragement. I'd like to thank my colleagues in Shanghai Medical University for their warm greetings from my hometown.

I would especially express my gratitude to my husband Xingxing, my son Ben and my whole family for their encouragement and love at all times.

Finally, I would like to acknowledge the Darwin Trust for supporting my PhD studies.

Nearly three years ago I was leaving busy Shanghai and research on Pharmacology to work in beautiful Edinburgh on transposable elements. Looking back, I have had really joyful time. Thanks for all the hard work, inspiration, success, unsuccess, conferences and parties.

Abbreviations

A	adenosine
Amp	Ampicillin
ATP	adenosine triphosphate
Adh	alcohol dehydrogenase
bp	base pair
BSA	bovine serum albumin
C	cytidine
°C	degrees celsius
cDNA	complementary DNA
Ci	curie
DEPC	diethyl pyrocarbonate
DMF	dimethylformamide
dH_2O	de-ionised water
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxyribonucleoside triphosphate
DTT	dithiothreitol
E.coli	*Escherichia coli*
EDTA	ethylene diaminetetra-acetic acid
G	guanosine
HEPES	N-[2-Hydroxyethyl]piperazine-N'-[2-ethane-sulphonic acid]
IPTG	isopropyl-β-D-thiogalactopyanoside
kb	kilobase
KD	kiloDalton
l	litre
LB	Luria Broth
M	molar (moles per litre)
mol	mole
min	minute
MOPS	3-[N-morpholino]propanesulphonic acid
MW	molecular weight
μ	micro ($x10^{-6}$)
n	nano ($x10^{-9}$)

O.D.	optical density
ORF	open reading frame
Ω	ohms
p	pico ($x10^{-12}$)
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethlene glycol
POD	peroxidase
PMSF	phenylmethylsulphonyl fluoride
PVDF	polyvinylidene fluoride
rpm	revolutions per minute
RNA	ribonucleic acid
RNase	ribonuclease
SDS	sodium dodecyl sulphate
T	thymidine
TEMED	N,N,N'-tetramethylethylenediamine
Tris	tris (hydroxymethyl) aminomethane
Tm	melting temperature
UV	ultraviolet
UTR	untranslated transcribed region
V	volt
X-gal	5-bromo-4-chloro-3-indol-β-D-galactopyranoside

AMINO ACIDS

Amino acids	Three letter abbreviation	One letter symbol
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

ABSTRACT

Transposable elements are sequences that can move from one place to another both in prokaryotic and eukaryotic genomes. *Mariner* is a member of the *Tc1/mariner* superfamily of eukaryotic transposons that transpose in a "cut-and-paste" manner via a DNA intermediate. It is one of the most widespread DNA transposons in nature. The functional *mariner* element *Mos1* is 1286 bp long with 28 bp imperfect terminal repeats and contains a single open reading frame encoding a 345 amino acid transposase.

In this thesis, the N-terminal domain of mariner transposase was found to bind specifically to inverted repeats at both ends. The minimal DNA-binding domain was identified to be located between amino acid 1 and 120. This contains a helix-turn-helix (HTH) motif. Site-directed mutagenesis indicates that the HTH is required for sequence-specific recognition of the terminal inverted repeats by Mos1 transposase. The transposase has 5 to 6 times higher affinity for the right end sequence of Mos1 than for the left end.

Yeast two-hybrid assays were performed to identify protein-protein interactions of Mos1 transposase. Deletion mutants suggest that residues required for the interaction of transposase monomers are distributed along the length of the protein. Twelve single point mutations which reduce subunit interactions have been isolated. The L124S mutation is located in neither the DNA-binding nor the catalytic domain. The purified Mos1(L124S) mutant reduces transposase activity in both excision and transposition assay, indicating that subunit interactions are involved in the transposition reaction. A change in target site selection other than the TA dinucleotide observed with this mutant may reflect a change in the conformation of the transpososome.

The promoter activity of *Mos1* was detected in transgenic flies by *P* element-mediated germline transformation. The promoter of *Mos1* lies within nucleotides 1-171 which shows low activity. There is an enhancer-like element in the region of nucleotides 172-516 that stimulates reporter gene expression. No autoregulatory inhibition of Mos1 transposase on its own promoter was observed. This suggests that "overproduction inhibition" may not act at the level of transcription through autorepression of the *Mos1* promoter.

TABLE OF CONTENTS

Chapter 1

Introduction

1.1 Introduction to transposable elements

Transposable elements are sequences that can move from one place to another both in prokaryotic and eukaryotic genomes (Sherratt, 1995; Berg and Howe, 1989). Over 50 elements have been found in *Drosophila melanogaster,* making up about 10% of the total genome (Manning et al., 1975; Young, 1979). It is nearly half a century since the classical genetic experiments of Barbara McClintock suggested existence of transposing DNA elements in maize (McClintock, 1957) which were originally called "controlling elements" because they can cause variegated phenotypes. Transposable elements may be divided into two main classes according to their structure and presumed mechanism of transposition (Fig.1.1-1.2). Class I elements contain genes that are related to reverse transcriptase genes of retroviruses. They transpose by reverse transcription of RNA intermediates. Class II elements have short inverted repeats at their termini and transpose by an excision-insertion mechanism (Finnegan, 1989; 1992).

1.1.1 Class I elements

There are two types of Class I elements. The first type of Class I elements are similar in structure to the integrated provirus of a retrovirus. They have long direct repeats, LTRs, at their termini. Most of these elements have two open reading frames that are similar to the *gag* and *pol* genes of retroviruses (Patience et al., 1997). The *pol* gene codes for a protease, reverse transcriptase and integrase required for retrotransposition. The *Ty* elements of yeast (Boeke, 1989) and the *copia* in Drosophila (Yoshioki et al., 1991) are of this type. These elements share an overall life cycle that resembles that of the retroviruses, except that there is no extracellular or infectious phase. Transcription of the RNA transposition intermediate initiates in one LTR and terminates in the other. This RNA is translated and then packaged into virus-like particles (or VLPs) that are constructed of proteins encoded by the *gag*- and *pol*-like ORFs (Burns et al., 1992). Reverse transcription of RNA is carried out in the VLPs, and the cDNA produced is then transported to the host cell nucleus (Kenna et al., 1998).

A few LTR elements, like *gypsy*, have a third open reading frame that is similar to an *env* gene. In retroviruses, *env* codes for components of the viral envelop that mediate entry of the virus into the host cell. The *gypsy* element which has a similar third ORF has been

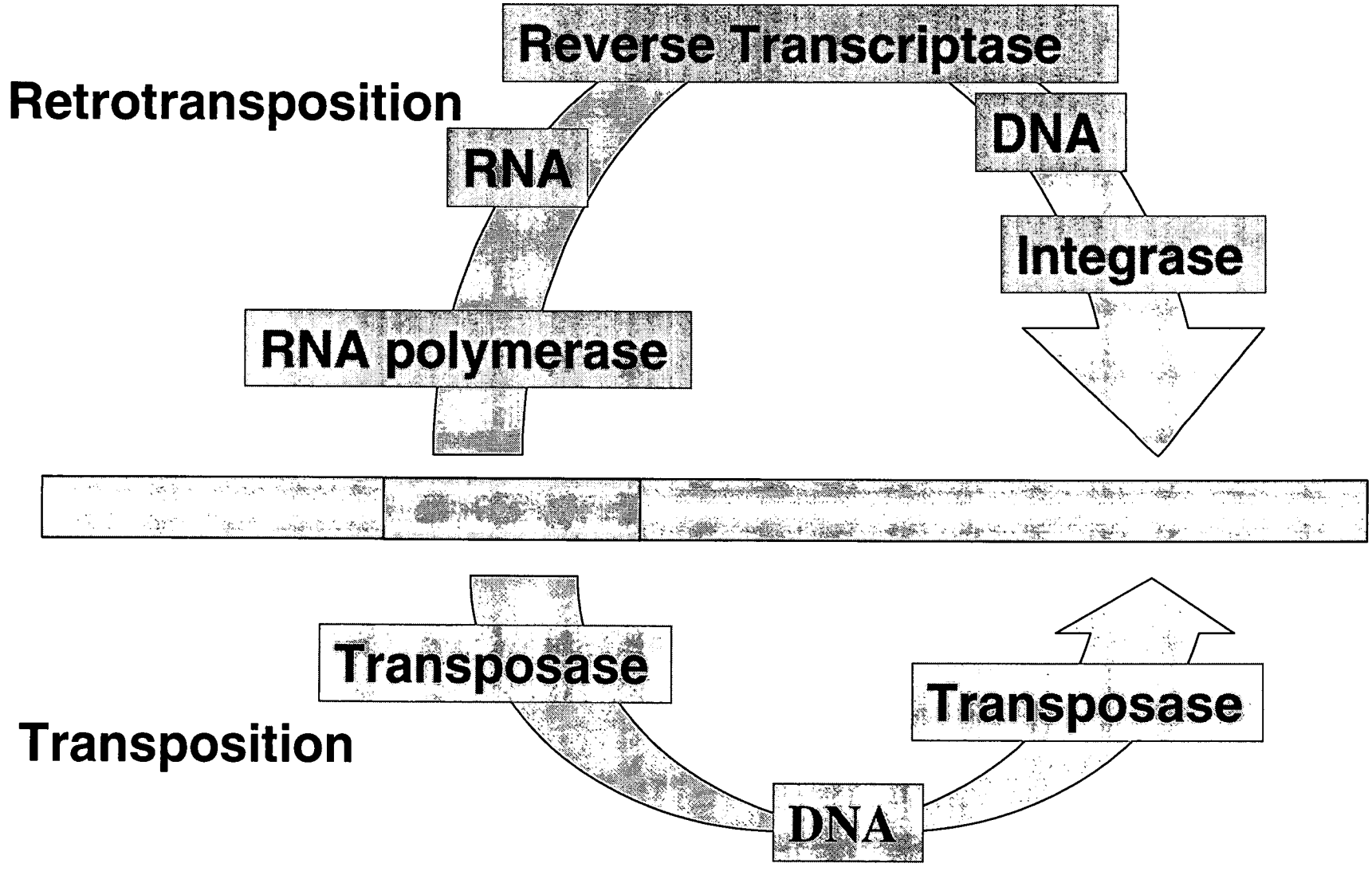


Figure 1.1 Retrotranspoition and transposition

(adapted from Polard and Chander, 1995)

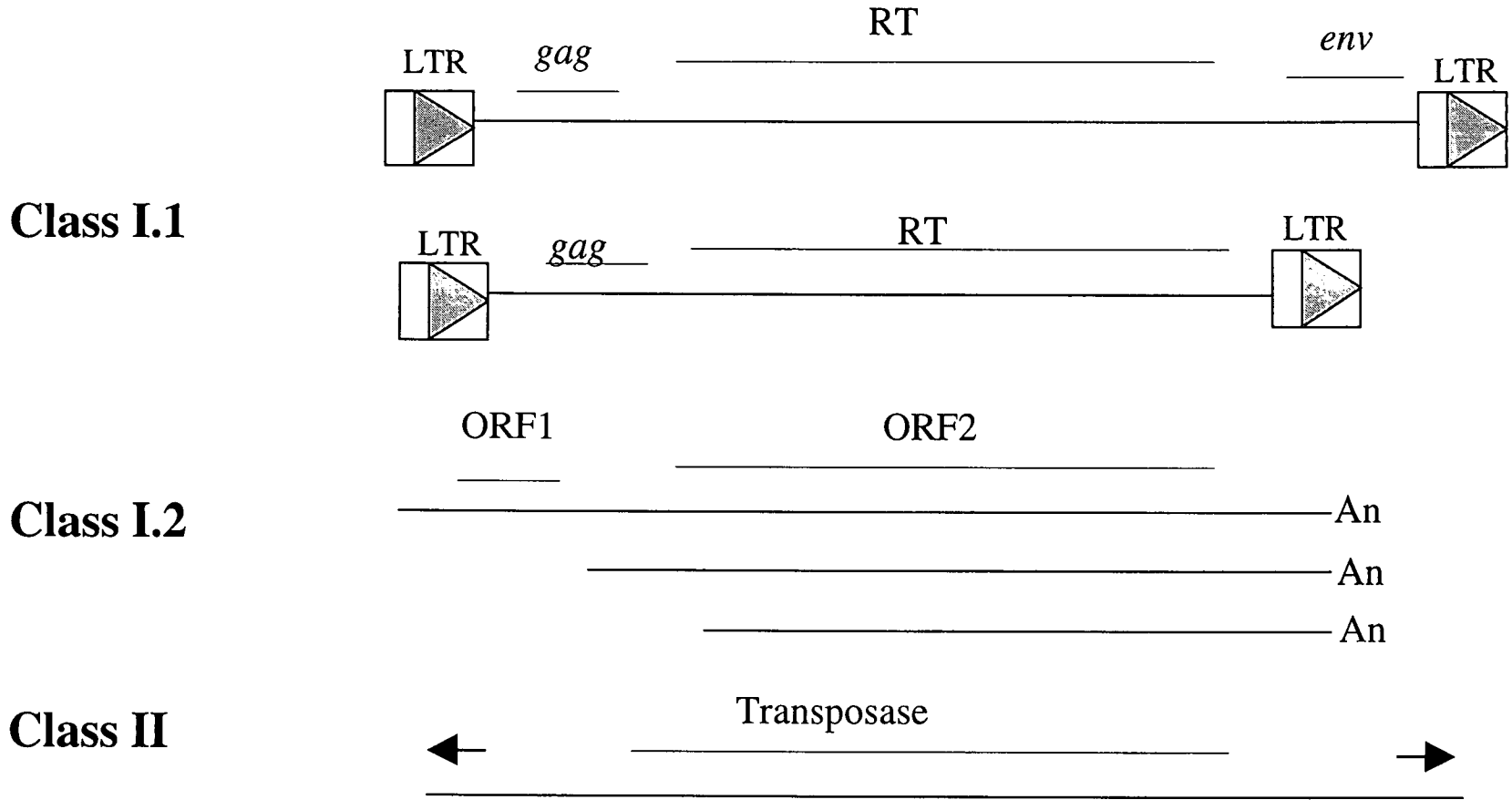


Figure 1.2 Schematic representation of the structure of transposable elements (adapted from Finnegan, 1992)

Class I.1 elements have long terminal repeats (LTRs). They have a *gag* like ORFand an ORF that encodes a reverse transcriptase(RT). Some of them also have a third *env*-like ORF. ClassI.2 elements have two ORFs and A-rich sequences (An) at their 3'ends. Class II elements have short inverted repeats (arrowed) and a gene encoding transposase.

suggested to be an infectious retrovirus since when fly larvae are exposed to *gypsy* viral-like particles, high level of *gypsy* insertion activity is observed (Sony et al., 1994; Kim et al., 1994).

The second type of Class I elements generally have two open reading frames but they have no terminal repeats and end with A-rich sequences at the 3' end of their coding strands. They are often referred to non-LTR retrotransposons or LINE-like elements as the first examples to be detected were mammalian LINE, or *L1* elements (Evans and Palmiter, 1991). They occur in the genome as families of elements that are truncated at their 5' ends. Other members in this class include the R2Bm element of the silkworm *Bombyx mori* (Luan et al., 1993), the *I* factor of Drosophila (Fawcett et al., 1986) and mobile group II introns (Curio and Belfort, 1996). The transposition mechanism of LINE-like elements is known as target-primed reverse transcription (TPRT). Integration initiates by target DNA breakage at the site of element insertion. Reverse transcription and integration occur simultaneously as the target 3'-OH exposed by this break provides a primer for DNA synthesis using an element RNA as a template (Finnegan, 1997).

1.1.2 Class II elements

Class II elements transpose directly from DNA to DNA without an RNA intermediate. DNA breakage and joining reactions are at the heart of DNA-based transposition reactions. The ends of the element are disconnected from a donor site by DNA breakage reactions, and then these exposed ends are joined to a target site by DNA strand transfer reactions (Mizuuchi, 1992). This pathway is used by bacterial elements such as *Tn5, Tn7, Tn10, Mu,* and the eukaryotic *P* family (Engels, 1989), hAT superfamily (*hobo-Ac-Tam3*) (Atkinson *et al,* 1993), *pogo* superfamily (Tudor et al., 1992) and *mariner/Tc1* superfamily (Plasterk, 1996). Class II elements generally have inverted terminal repeats of 10 - 200 bp and one or more ORFs that encode a protein, transposase, that is required for transposition.

1.2 The discovery of *mariner* transposable elements

Mariner was first identified during analysis of an unstable mutation in the *white* gene of *Drosophila mauritiana*, a sibling species of *Drosophila melanogaster.* This mutation results in a peach eye colour and is due to insertion of a novel transposable element into the

5'-nontranslated leader region of *white* gene. This allele was designated *white-peach* (w^{pch}) (Jacobson and Hartl, 1985; Haymer and Marsh, 1986). The transposable element was isolated and named *mariner* (Jacobson et al., 1986). The *mariner* element is able to excise and transpose in both somatic and germ cells. A somatic excision from the original w^{pch} allele results in eye-colour mosaicism (red spots on a peach background), while germline excision results in progeny with wild-type red eyes in the next generation (Bryan et al., 1987; Bryan and Hartl, 1988). However, the w^{pch} allele contains a nonautonomous copy of the *mariner* element, sometimes called the *peach* element. In most w^{pch} -bearing strains of *Drosophila mauritiana*, a low level of somatic excision of the *mariner* element from w^{pch} occurs, yielding pigmented patches on a peach-coloured background. In the germline, w^{pch} reverts to wild type at a frequency of about 10^{-3} per gene per generation. The infrequent occurrence of such somatic mosaics and germline excision may be due to low activity among the approximately 20-30 copies of *mariner* present in the *D.mauritiana* genome (Jacobson and Hartl, 1985).

Subsequently work from Hartl's laboratory (Hartl, 1989) has led to the isolation of strains which showed enhanced rates of both somatic and germline excision of *mariner* from *white*, and were associated with a trans-acting factor designated *Mos1* (mosaic-1) (Bryan et al., 1987; Medhora et al., 1988). The *Mos1* factor has been shown to be a particular copy of *mariner* capable of high rates of both excision and transposition. For example, in most strains, the *Mos1* results in eye colour mosaicism in virtually every fly and rates of germline reversion of w^{pch} in *Mos1* strains are up to 1%. Sequence comparison showed that the *peach* element differs from *Mos1* at 11 nucleotide sites including four amino acid replacements (Medhora et al., 1991) which may be responsible for the functional differences in activity between these elements. The functional organisation of the *Mos1* element is illustrated in Figure 1.3. The *Mos1* transposable element is 1286 base pairs in length including 28bp inverted repeats with 4 mismatches at its termini. Within the inverted repeats, a single ORF of 1035 nucleotides encodes a transposase of 345 amino acids, with a molecular weight of 40.8 Kda.

1.3 Distribution of *mariner* among organisms

The distribution of the transposable element *mariner* was first examined among the eight closely related species comprising the *melanogaster* species subgroup of Drosophila. The

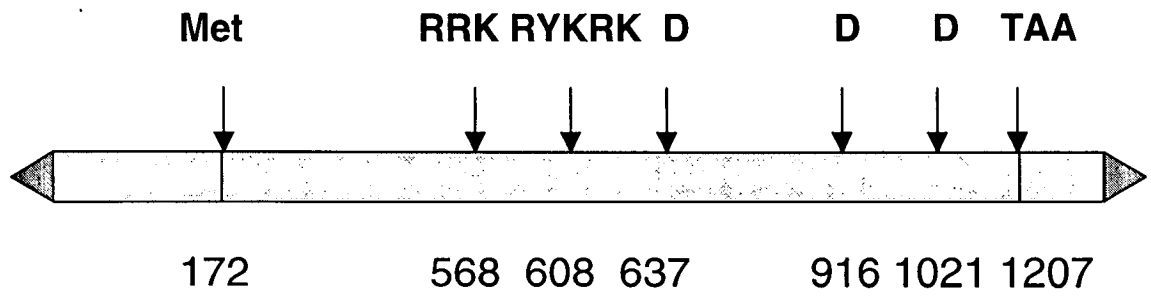


Figure 1.3 Sequence organization of *Mos1* element.

*Mos*1 is an active, autonomous copy of *mariner*. It is 1286 bp in length with 28 bp inverted repeats. The single uninterrupted open reading frame codes for the transposase gene. The D,D(34)D signature residues are required for the catalytic activities of transposase. A bipartite nuclear localization signal (RRK and RYKRK) is indicated. The triangles at the ends represent the inverted repeats. Numbers are nucleotides.

element is present in *D.mauritiana, D.simulans, D.sechellia, D.yakuba and D.teissieri*, but it is absent from *D.melanogaster, D.erecta and D.orena* (Maruyama and Hartl, 1991a; Capy et al., 1991, 1992; Brunet et al., 1996). This discontinuous distribution of *mariner* in the *melanogaster* species subgroup was thought to be due to stochastic loss since the *mariner* gene phylogeny is congruent with the species subgroup phylogeny based on nucleotide sequences of the alcohol dehydrogenase (*Adh*) gene (Maruyama and Hartl, 1991a). According to this hypothesis, *mariner* was present in the common ancestor of the subgroup containing eight species but lost in the lineage to *D.melanogaster, D.erecta and D.orena*. Within the family Drosophilidae, the distribution of *mariner* elements is not uniform and doesn't follow the phylogeny of the host species (Brunet et al., 1994). For instance, *mariner* elements in species belonging to two different genera (*Drosophila mauritiana* and *Zaprionus tuberculants*) can be more closely related than *mariner* elements in species belonging to the same genus (*D.mauritiana* and *D.tsacasi*) (Maruyama and Hartl, 1991b). This may be the results of horizontal transmissions between Drosophila species and /or between Drosophila species and donor species outside the Drosophilidae family. Thus, the history of this transposon within the drosophilids reveals both vertical and horizontal patterns as the basis for its patchy distribution (Capy et al., 1994; Brunet et al., 1994).

The first demonstration of a wide phylogenetic distribution of *mariner*-like elements came from PCR amplification using degenerate oligonucleotide primers complementary to regions corresponding to the WVPHEL and YSPDLAP amino acid domains of the putative transposases of the *mariner* element from *Drosophila mauritiana* and a *mariner*-like element from *Hyalophora cecropia* (Robertson, 1993). This revealed the presence of *mariner* elements in ten other species, representing six additional orders, including insects as diverse as bees, mosquitoes, silverfish, catfleas and earwigs. Sequences of multiple clones from each species revealed a diverse array of *mariner* elements, with multiple subfamilies in the genomes of some insects indicating both vertical inheritance and horizontal transfers.

Mariner elements are also found in invertebrates other than insects, including the centipede *Scutigera coleoptrata* (Robertson and Macleod, 1993), the nematode *Caenorhabditis elegans* (Sedensky et al, 1994); nematode *Trichostrongylus colubriformis* (Wiley et al., 1997); insect parasite nematode *Heterorhabditis bacteriophora* (Grenier et al., 1999); the

planarian *Dugesia tigrina* (Garcia-Fernandez et al., 1993); flatworm *Stylochus zebra* and *Belloura candila* (Robertson, 1997); silkworm moth (Robertson and Asplund, 1996) and hydras (Robertson, 1997). The genome of the planarian *Dugesia tigrina* contains a high copy number of *mariner* elements. Most of them have an intact ORF and probably some of them are still active (Garcia-Fernandez et al., 1995). The high degree of similarity between these elements and those from other arthropods, together with the fact that this element is not found in other planarian species, strongly suggests a case of interphylum horizontal transmission.

Mariner elements are prevalent not only in insects and other invertebrates but also in vertebrates. The presence of *mariner* elements in the genomes of human, mouse, rat, sheep, Chinese hamster and cattle has been reported (Auge-Gouillou et al., 1995). Three subfamilies of *mariner* elements have been identified in the human genome: *cecropia, irritans* and *mellifera* (Oosumi et al, 1995; Smit and Riggs, 1996), consistent with multiple horizontal transfer events. One pair of *mariner* elements has received considerable attention because of its potential role in human diseases. It is located in chromosomal 17p as a hotspot of unique cross-over. The resulting products are either duplicated or deficient for the gene for peripheral myelin protein 22 which cause CMT1A (Charcot-Marie Tooth disease type 1A) and HNPP (Hereditary neuropathy with liability to pressure palsies), respectively (Reiter et al, 1996). The genomic locations of 109 copies of this *mariner* element were determined by primed in situ labelling using primers designed to match the right and left inverted terminal repeats of the transposon. These elements appear to be spread evenly throughout the genome and their locations may coincide with at least 12 locations where homologous recombination is the proposed molecular mechanism responsible for the human disease phenotypes (Reiter et al., 1999).

The copy number of *mariner* elements varies widely from one species to another. It is about 30 copies in *D.mauritiana*, from which *Mos1* was derived. 100-300 copies have been estimated in the human genome. The planarian *Dugesia tigrina* contains about 8000 copies of a *cecropia*-subfamily MLE. The record is held by the horn fly *H.irritants* which has about 17,000 copies of *mariner* elements making up 1% of its genome by mass (Robertson and Lampe, 1995).

Phylogenetic analysis indicates that *mariner* elements can be divided into 15 subfamilies (Robertson 1997), independently of the host phylogeny. In general, the amino acid sequences within a subfamily have greater than 40% amino acid identities, while members of different subfamilies show 25-40% identity Robertson and Lampe, 1995).

1.4 *Mariner* superfamily

The transposases encoded by *mariner* elements in animals, referred to collectively as *mariner*-like elements (MLEs), are members of a large superfamily of transposase and integrase proteins known as the D,D(35)E superfamily. The two aspartic acid (D) residues are separated from one another by a variable distance, followed by a glutamic acid(E) residue that is 34 (in eukaryotic elements) or 35 (in prokaryotic elements) amino acids downstream (Doak et al., 1994; Capy et al., 1996). The MLEs are unique in having a D,D(34)D signature in these positions (Robertson, 1995). This superfamily also includes proteins encoded by prokaryotic elements *Tn7*, *Tn10*, bacteriophage *Mu* and many bacterial insertion sequences as well as retrotransposons and *Tc1* family (TLEs), sister group of *mariner* family (Table 1.1).

The D,D(35)E motif is thought to be a key player in the transposition reaction. The acidic residues may participate in co-ordination of the metal cofactor (Mn^{2+} or Mg^{2+}) required for the catalytic activities of transposase during the DNA cleavage and strand transfer reactions (Kulkosky et al., 1992). The importance of the DDE residues has been demonstrated by site-directed mutagenesis of enzymes of several members of the family, such as *Tc3* (van Luenen et al., 1994), *mariner* (Lohe et al., 1997), *Mu* (Baker and Luo, 1994), *Tn7* (Sarnovsky et al., 1996) and *Tn10* (Junop and Haniford, 1997). The X-ray crystal structures of the catalytic domains of HIV integrase (Dyda et al., 1994), *Mu* transposase (Rice and Mizuuchi, 1995) and ASV integrase (Bujacz et al., 1995) proteins have shown that the D,D(35)E motif forms a plausible binding site for divalent metal ion , which may promote phosphoryl transfers (Grindley and Leschziner, 1995). Several crystal structures of integrase core domains have been solved in the presence of metal ions. Typically, these structures show the binding of Mg^{2+} or Mn^{2+} only at the site located between the two aspartate residues (Bujacz et al., 1996; Goldgur et al., 1998). In the active site of Tn5 transposase/DNA complex, one Mn^{2+} ion is co-ordinated by Asp^{97} and Glu^{326} and by the 3'-OH of the transferred strand of DNA (Davies et al., 2000).

Table 1.1 DDE motifs in transposases and integrases

Transposon	DDE motif	Reference
HIV-1 (IN)	D(51) D(35) E	Goldgur et al., 1998
ASV (IN)	D(56) D(35) E	Bujacz et al., 1995
Mu (MuA)	D(66) D(55) E	Rice and Mizuuchi, 1995
Tn7 (TnsB)	D(87) D(34) E	Sarnovsky et al., 1996
Tn10 (IS10)	D(63) D(130) E	Junop and Haniford, 1997
Tn5 (IS50)	D(90) D(136) E	Davies et al., 1999
IS630	D(80) D(35) E	Capy et al., 1996
IS911	D(59) D(35) E	Capy et al., 1996
Tc1	D(89) D(34) E	Vos and Plasterk, 1994
Tc3	D(86) D(34) E	Van Luenen et al., 1994
Mariner (Mos1)	D(92) D(34) D	Roberston, 1995

The number of residues between the first and second D residue and between the second D and the E residue is shown between brackets.

In addition to the DDE-containing transposases and integrases, structural studies of the catalytic domains of proteins that promote phosphoryltransfer reactions, notably RNaseH (Katayanagi et al., 1990) and RuvC resolvase (Ariyoshi et al., 1994) have revealed remarkably similar topologies. This has led to the notion that such enzymes belong to a superfamily of proteins known as polynucleotidyltransferases (Rice et al., 1996; Dyda et al., 1994).

The *mariner/Tc1* superfamily of transposable elements, which is named after its two best-studied members, is one of the most diverse and widespread groups of Class II transposable elements. *Tc1* was discovered in 1983 as a repeat sequence in the genome of *C.elegans* (Emmons et al., 1983). At present, it seems that members of the *mariner/Tc1* superfamily are found in virtually all animal phyla, from fungi to vertebrates (Plasterk, 1996). The relationship of the various *mariner* subfamilies to the *Tc1* family is illustrated in Figure 1.4. Similar structures and molecular mechanisms of transposition between the *mariner* and *Tc1* families including duplication of a TA target site upon insertion, their activity in both germline and somatic tissues and the length of their putative transposases (about 340-350 a.a.) suggest a common ancestor for them. Amino acid identities between members of the two families range from 10% to 20% (average 16%) (Robertson, 1995). Three of the *Tc1* family elements (*Tc1, Tc3* and *Minos*) have single short introns in their transposase genes. Interestingly, the most striking difference is in the so-called D,D(35)E motif mentioned before. Elements in the *mariner* family have the motif D,D(34)D instead of D,D(34)E in the *Tc1* family, and the "conservative " change from D,D(34)D to D,D(34)E completely obliterates Mos1 transposase activity (Lohe et al., 1997).

1.5 The mechanism of *mariner* transposition

All transposition reactions analysed thus far utilise the same basic chemical strategy for joining transposon ends to target DNA. After the two ends are brought together in a synaptic complex, transposases or integrases make a nick precisely between the transposon end and the flanking DNA at the 3' ends of the transposon, leaving a 3'-OH at the transposon end. The 3'-OH termini then engage in direct nucleophilic attack upon the two strands of target DNA in a symmetrical pair of transesterification reactions (Mizuuchi, 1997). The 5' cleavage of the transposable elements does not seem to play an important role in the transposition and is found at different positions in various elements.

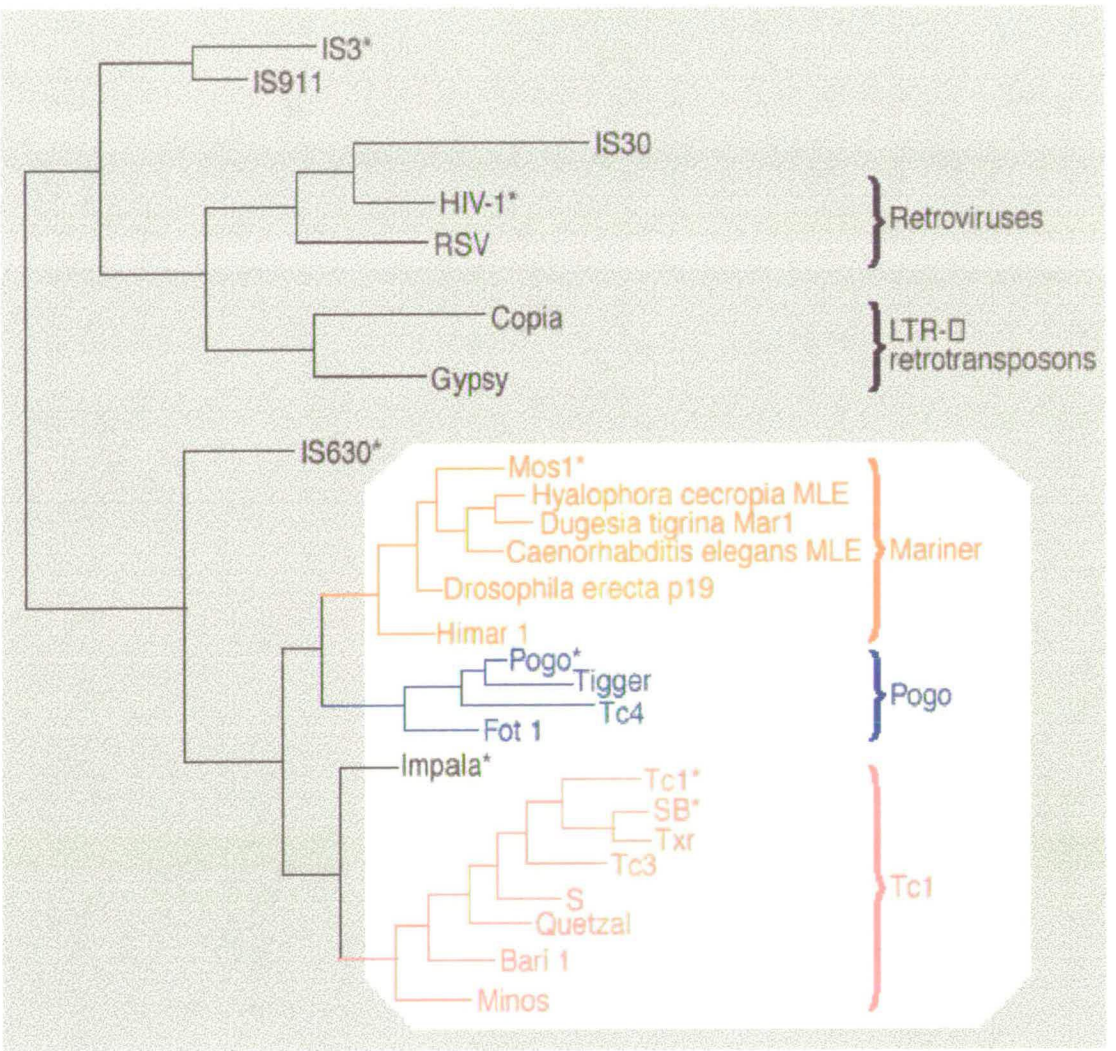


Figure 1.4 Phylogeny of the *Tc1/mariner* superfamily (from Plasterk et al., 1999).

Recombinases that contain the DDE motif are grouped into two major clusters: a DNA-transposon group and a retroelement group. The *Tc1*, *mariner* and *pogo* transposon families are probably monophyletic.

1.5.1 Retroviral Integration

In retroviral integration, the viral RNA genome is converted to linear double-stranded DNA by reverse transcription. This linear viral DNA is the precursor in integration (Fujiwara and Mizuuchi, 1988; Fujiwara and Craigie, 1989). The element-encoded integrase initiates recombination by DNA cleavage a few nucleotides (usually 2 bp) from the 3' ends of the viral DNA, exposing the actual 3'-OH ends of the transposable segment and leaving a two-base 5' overhang (Kutzman et al., 1989). Strand transfer reactions then join these exposed 3' ends to staggered positions on the target DNA. The staggered insertion results in the creation of short, complementary single-strand regions flanking the insertion element. Removal of the overhanging 5' bases and repair of the single-strand region at each end by host factors would then complete integration and generates the characteristic target site duplication (Craig, 1995). The resulting product, in which the retroviral DNA is covalently joined to the target DNA, is called a simple insertion (see Figure 1.5).

1.5.2 *Mu* transposition

Bacteriophage *Mu* uses a replicative transposition strategy which involves cleavage of only one strand at each transposon end, leaving the transposon covalently linked at its uncleaved 5' ends to donor DNA while inserting into a new target site. DNA cleavage reactions are executed by MuA transposase to expose the 3'-OH ends of the transposon. The exposed 3' ends of the transposon are then joined by strand transfer reactions to staggered positions on the target DNA. This transposition reaction is then followed by DNA replication to generate two copies of the element, one end of each copy being attached to the donor site and the other end to the target site, in a structure called a cointegrate (Lavoie and Chaconas, 1996) (Fig 1.5).

1.5.3 Cut and Paste

Many elements, including *Tn10, Tn7, Tn5, P* element, *Tc1/mariner* family, move by an excision-insertion or cut-and-paste pathway in which the element is excised from the donor site by double-strand breaks that disconnect both the 3' and 5' ends of the transposon from the donor backbone and join them to a target site without replication. *Tn5* is excised

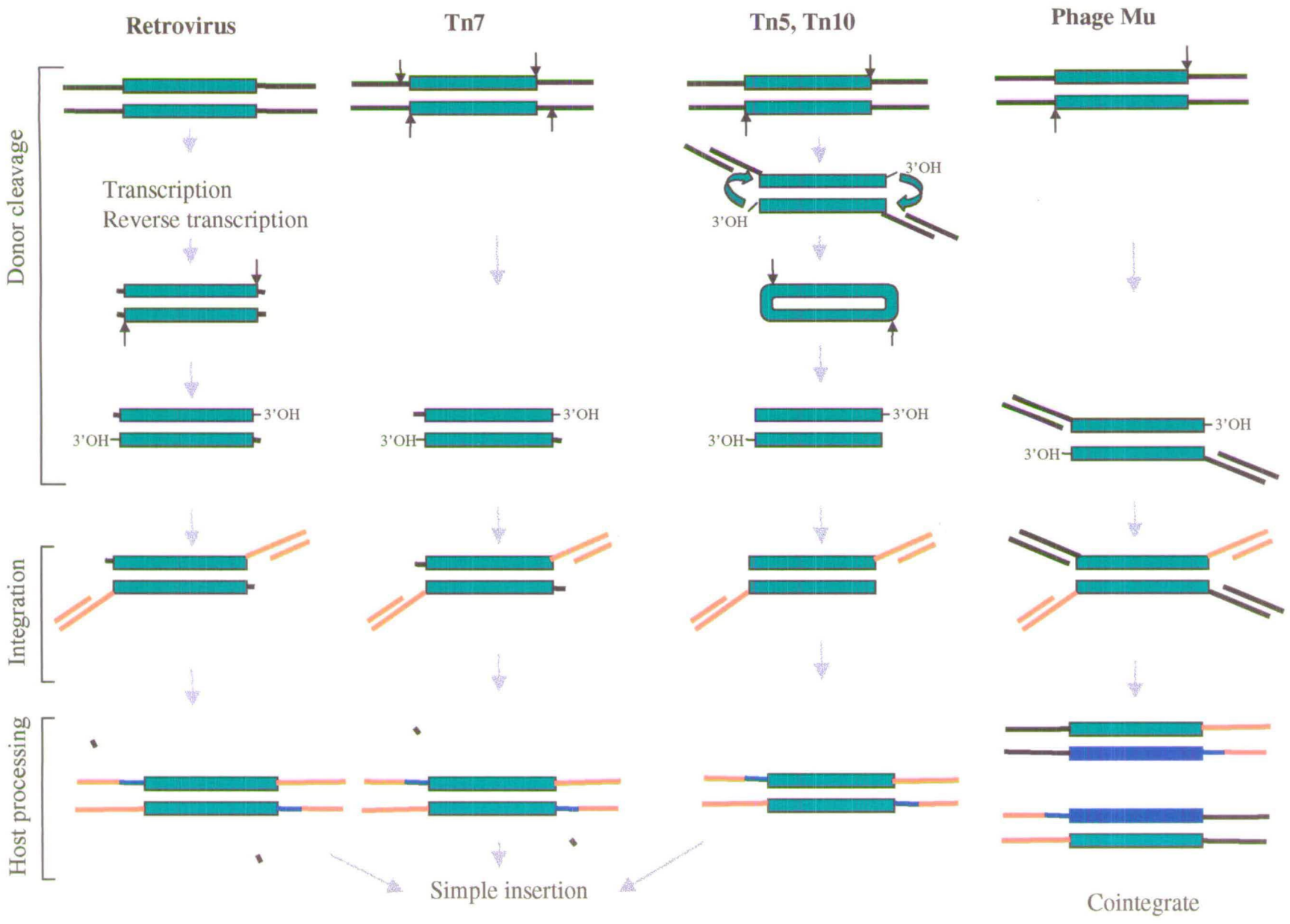

Retrovirus
Tn7
Tn5, Tn10
Phage Mu
Donor cleavage
Transcription
Reverse transcription
3'OH
3'OH
3'OH
3'OH
3'OH
3'OH
3'OH
3'OH
Integration
Host processing
Simple insertion
Cointegrate

Figure 1.5 Different models of DNA transposition (adapted from Craig, 1996; Haren et al., 1999).

Transposon DNA are indicated by green boxes, donor DNA by black lines, and target DNA by red lines. Positions of strand cleavage are indicated by black vertical arrows. The newly synthesized DNA are indicated by blue. Retrovirus is amplified by transcription and reverse transcription before insertion. Phage *Mu* passes through a cointegrate intermediate (note that the newly replicated sequences are indicated in blue). *Tn7* undergoes double-strand cleavage while *Tn5* and *Tn10* undergo double-strand cleavage by way of a hairpin intermediate. See text for detailed description.

from the donor DNA by flush double strand breaks precisely at the end of the element (Goryshin and Reznikoff, 1998) and the same is true of *Tn10* (Morisato and Kleckner, 1984). *Tn7* is excised with three nucleotide 5' overhangs (Bainton et al., 1991) whereas there are 3' overhangs for *Tc1/mariner* family (van Luenen et al., 1994) as a result of staggered cuts at the transposon ends.

It has been recently found that *Tn10* (Kennedy et al., 1998) and *Tn5* (Bhasin et al., 1999) transpose via hairpin intermediates. The first cleavage step is a 3' hydrolytic nick by transposase at the end sequence. The 3'-OH attacks the 5' end, forming a hairpin at the transposon end and releases the flanking donor DNA. The hairpin is resolved by hydrolytic cleavage to regenerate the 3'OH of the transferred strand. This 3'-OH joins to target DNA in a strand transfer reaction. This model allows for a single transposase active site to perform all the necessary chemical steps without having to undergo a large conformational change. *Tn7* has solved this problem in a different way by having its transposase made up of two polypeptides, TnsA and TnsB. TnsB contains the active site responsible for 3' nicking whereas TnsA contains the active sites for 5' nicking (Sarnovsky et al., 1996). Different models of DNA transposition are shown in Figure 1.5.

A model of *Mos1* transposition has been proposed (Bryan et al., 1990; Coates et al., 1995; Smith, 1997) adapted from the mechanism of excision and mismatch repair of *Tc3* (van Luenen et al., 1994). Double-strand breaks with a 3bp stagger are generated, resulting in an excised element. The excised *Mos1* element contains the complete *Mos1* sequence at the 3' ends but lacks the terminal three nucleotides at each 5' end. Integration of *Mos1* results in a duplication of the TA target sequence. After integration of *Mos1*, the 5bp nucleotide gap at each end of the transposon will be repaired by the cellular machinery to produce a complete *Mos1* element flanked by a duplicated TA sequence. The lesions in the chromosome can be repaired by a mismatch mechanism, leaving either the sequence of the first or last three nucleotides of the element (Figure 1.6).

A DNA hairpin molecule has also been detected in *Mos1* transposition. This hairpin species is formed on the flanking DNA rather than at the end of the element. This is consistent with the order of the cleavage: *mariner* transposase cleaves at the 5' end first followed by second strand cleavage at the 3' end of the element (Dawson, personal communications).

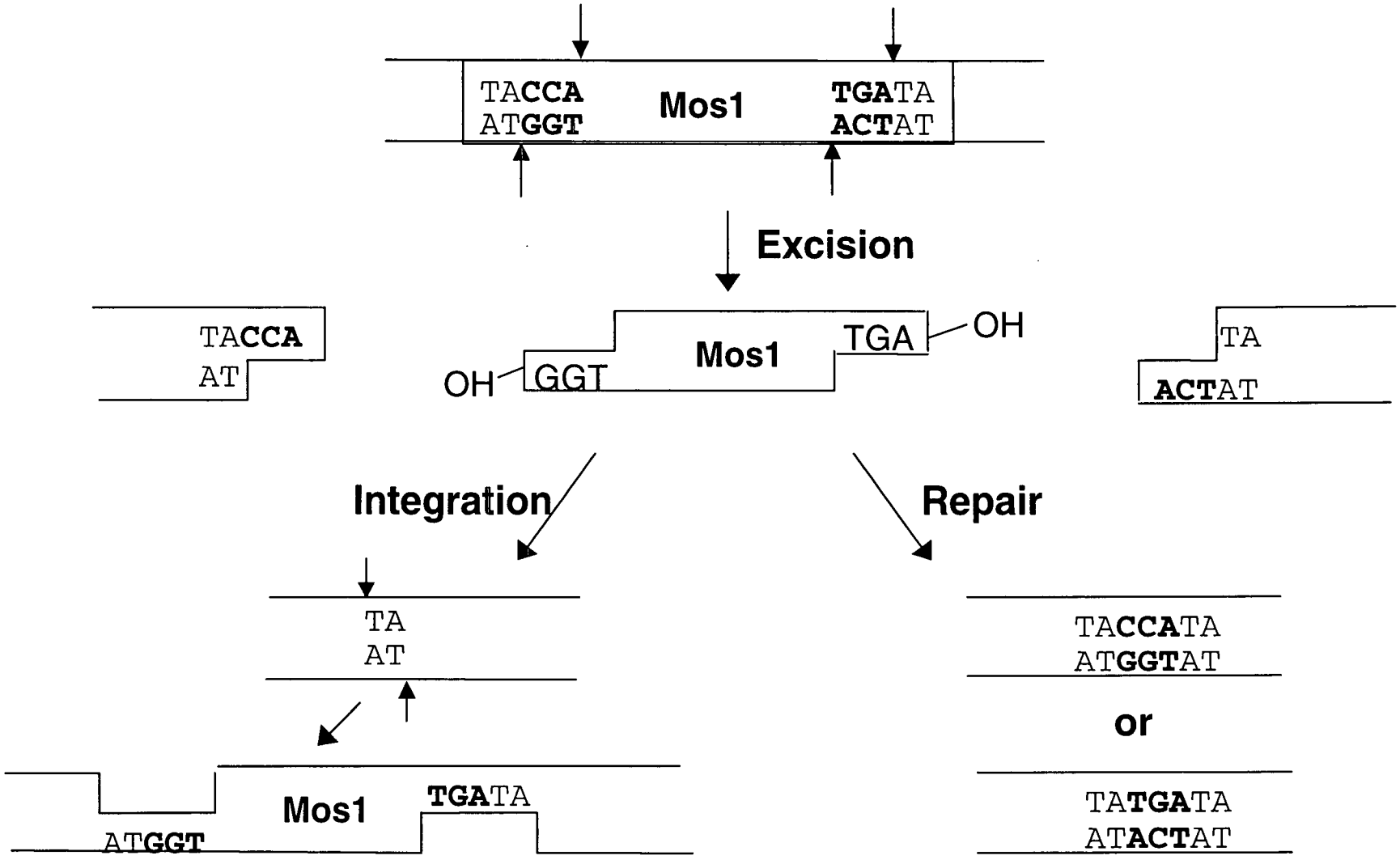


Figure 1.6 Model for *Mos1* transposition.

1.6 Regulation of *mariner* elements

Regulation of the activity of a transposable element is critical for the long-term persistence of the transposable element and the host organism, because increased frequencies of transposition would result in reduced viability, reduced fertility and increased frequency of mutations. Restriction of activity of *mariner* elements is even more critical than for most transposable elements, because these elements are active not only in the germline but also in the soma. Three potential mechanisms of regulation have been proposed: titration, dominant-negative complementation , and overproduction inhibition.

1.6.1 Titration

Although *mariner* elements are abundant in many animal genomes, the vast majority of naturally occurring *mariner* elements are non-functional. Many are inactive because they contain multiple chain-termination, deletion or frameshift mutations that disrupt the open reading frame (Robertson, 1993; Brunet et al., 1996). Others have an ORF with only missense replacements but produce a protein that cannot support transposition, such as the *peach* element in *D.mauritiana*. Inactive elements may participate in the regulation of transposition. *Mariner* elements with defective ORFs may still retain their transposase-binding sites, such elements can serve as substrates for transposition, thereby titrating the functional transposase, while their multiplication does not increase the total amount of transposase produced. One particular deletion is especially common in natural populations of *D.teissieri* (Brunet et al., 1996). This deletion spans between nucleotides 544 and 1260, eliminating about half of the 3' end of the ORF. A small repeated sequence of 7bp (AACGAGA) that is present in the complete element (*Mos1*) occurs only once in the deleted element suggesting that it originated through homologous recombination. This element is a strong candidate for a titrator element. Titration effects have also been suggested for defective *P* elements (Simmons and Bucholz, 1985). In the P-M system of hybrid dysgenesis of *Drosophila melanogaster*, some M strains possess defective *P* elements, which could bind transposase and create a shortage of transposase able to catalyse excision of *P* elements at the *singed* locus. A similar titration mechanism might

explain why certain deletions have replaced almost all functional *mariner* elements in a few species.

1.6.2 Dominant-negative complementation

A second mechanism of regulation offered by *mariner* elements with inactive open reading frames is that of direct interference with functional transposase through "poisoning" the transposase with inactive subunits. Among 18 missense mutations in the *Mos1* transposase induced by EMS and site-directed mutagenesis, 7 have no effect on wild-type transposase (Lohe et al., 1997). Another eight of the mutations decreased the rate of w^{pch} excision to about half compared to that of wild-type when they were examined in heterozygous mutant/nonmutant genotypes. The greatest effect was observed in three missense mutations that inhibited the activity of wild-type transposase more than twofold. These dominant-negative mutations are located near the D,D(34)D motif. This effect was attributed to dominant negative complementation from "poison" subunits that combine with wild-type monomers to create oligomers with impaired activity (Lohe et al., 1996). In maize, mutant forms of *Activator* transposase interact with wild-type transposase in a dominant negative manner when assayed by transfection of petunia protoplasts (Kunze et al., 1993). Since two of the mutants are DNA-binding-deficient, it was suggested the inhibition of excision is not caused by competition between wild-type and mutant subunits for binding to transposon DNA, but rather from a formation of non-functional heterooligomers between defective and intact protein.

A particular type of naturally occurring *P* element contains an internal in-frame deletion and encodes a 207-amino-acid KP protein (Black *et al*, 1987). The first 199 amino acids are identical to the N-terminus of transposase enzyme. There are two protein-protein interaction motifs that mediate dimerization : a leucine zipper motif that contains four leucines in a heptad repeat, and a basic C-terminal region of the protein (Lee et al., 1996). This repressor protein acts in a dominant-negative manner to form non-functional oligomers with wild-type transposase. Mutagenesis of the leucine zipper impaired the ability of the mutant proteins to inhibit transposase activity in vivo (Andrews and Gloor, 1995). This suggests the multimer-poisoning model of *P* element repression. If there is an analogue of the KP element in *mariner*, it is the deleted element in *D.teissieri* as described above. This deleted *mariner* element may encode a truncated protein that acts as a

repressor of the transposase encoded by active *mariner* elements. But this truncated protein has not been identified and examined for dominant-negative complementation.

Tn5 is unique among prokaryotic transposable elements in that it encodes a special inhibitor protein identical to the Tn5 transposase except lacking the first 55 amino acids. This protein regulates transposition through non-productive protein-protein interactions with transposase. Tn5 transposase (Tnp) and inhibitor (Inh) protein share two distinct dimerization domain, one of which is located at the C-terminus and the other near amino acids 114-314 (Braam and Reznikoff, 1998). Mutation of the recently discovered dimerization interface in Inh at the C-terminus (Davies et al., 1999) abolished Tnp-Inh heterodimerization and *in vitro* inhibition (Braam, 1999).

1.6.3 Overproduction inhibition

The *mariner* transposase has the remarkable property that overproduction of the wild-type transposase appears to reduce the overall level of transposase activity as assayed by the excision of a nonautonomous *mariner* target element. This phenomenon is called overproduction inhibition (OPI). In the *hsp70:Mos1P:Mos1* construct, the *hsp70* promoter is fused to *Mos1* promoter (*Mos1P*) in the putative *Mos1* TATA box (nucleotide position 58 in *Mos1*) and drives transcription of the *Mos1* transposase reading frame (Lohe and Hartl, 1996a). Increasing the number of copies of the construct decreases the rate of germline excision by 25% at 25^{o}C and 45% with heat shock. The mechanism of overproduction inhibition has not been determined.

Overproduction inhibition results from a dosage effect and may be mediated through any of a number of molecular mechanisms. One possibility is that OPI could act at the level of transcription through autorepression of the *Mos1* promoter. *P* transposase recognises a 10 bp DNA sequence that overlaps the TATA box of the *P* promoter (Kaufman et al., 1989) and full-length transposase can repress transcription *in vitro* by interfering with TFIID-TATA box interactions, thereby blocking the assembly of an RNA polymerase II transcription complex at the *P* element promoter (Kaufman and Rio, 1991). Expression and transposition of the *Suppressor-mutator* (Spm) transposon of maize are controlled by autoregulatory mechanism. TnpA is a multifunctional protein that participates in transposition as well as in both positive and negative regulation of Spm promoter

expression. TnpA activates the inactive, methylated promoter and leads to reduced methylation. By contrast, TnpA represses the active, unmethylated Spm promoter (Schlappi et al., 1994).

Another possibility of OPI is that transposase monomers, when present in excess concentration, form inactive or weakly active oligomers. The inverse dose effect of the maize element *Activator* (Ac) is caused by such an autoregulatory mechanism. As the transposase protein concentration increases with *Ac* copy number, apparently an inverse relation exists between transposase concentration and transposition frequency. It was proposed that transposase levels above a certain threshold inhibit the transposition reaction (Scofield et al., 1993). Based on the observation that the transposase aggregates in vivo, the possibility is discussed that the aggregated transposase is a transpositionally inactive form of the protein (Heinlein et al., 1994).

Overproduction inhibition has been recently described for the reconstructed *Himar1* transposase *in vitro* (Lampe et al., 1998) and the *Sleeping Beauty* transposase *in vivo* (Yant et al., 2000).

1.7 *Mariner* as a transformation vector

Transposable elements can be exploited as valuable research tools for the molecular characterisation of genomes by acting as DNA vectors for genome manipulation and the "creation" of transgenic organisms (Warren and Crampton, 1994). They serve as agents of chromosomal insertion, deletion, or rearrangement, and provide the basis for transformation of somatic or germline cells. The *P* element from *D.melanogaster* has been used in a variety of insect species to create transgenic individuals, and appropriate transformation vector constructs are now used as routine research tools in Drosophila molecular genetics. However, the host range of *P* elements is limited to Drosophila species and attempts to use *P* elements to transform yeast and mammalian cells have been unsuccessful (Rio et al, 1988). Transposition of the *P* elements involves a host-encoded "inverted terminal repeat binding protein" that is thought to be at least one factor limiting its host range to the Drosophila (Beall et al, 1994). In contrast to *P* elements, there are indications that members of the *mariner/Tc1* superfamily may not require a species-specific factor for their transposition. *Mariner* elements are of great interest as potential

transformation vectors because of their extraordinarily wide taxonomic distribution and their apparent indifference to host factors (Kidwell, 1993; Warren and Crampton, 1994)

The standard method of germline transformation in *Drosophila melanogaster* is to coinject two plasmids based on the *P* element into embryos. One is the "helper" plasmid which provides a source of transposase. The other is the transformation vector carrying the terminal sequences of the *P* element. Between which lie a reporter gene for identification of transformed individuals, and a site where foreign DNA can be ligated into the vector.

Two constructs have been made to study germline transformation in *Drosophila melanogaster* mediated by *mariner* elements. One vector, called *pM1wB*, has an 11.9 kb cassette inserted into the *Sac*I site. This cassette includes an *hsp70:mini-white* gene as well as an *E.coli* ß-galactosidase reporter gene and plasmid sequences for transformation rescue while another vector was constructed with only a 4.5 kb *hsp:mini-white* cassette. When coinjected with *Mos1* helper plasmid, both constructs were found to insert into the genome in a *mariner*-dependent manner with the termini of the inverted repeats inserted at a TA dinucleotide. Both constructs exhibit remarkable germline and somatic stability after integration (Lidholm et al, 1993; Lohe and Hartl, 1996b). Although not ideal for purposes of genetic manipulation, the relative stability of the insertions may be of practical significance for some schemes of insect population regulation.

The ability of *mariner* elements to excise and transpose in non-drosophilid insects has also been reported. Plasmid-based excision (Coates et al., 1995) and interplasmid transposition (Coates et al., 1997) of *Mos1* have been observed in developing embryos of the Australian sheep blowfly *Lucilia cuprina* and the Queensland fruitfly *Bactrocera tryoni*. The majority of empty sites following *mariner* excision contained 3 bp from one of the *mariner* inverted repeats, as is observed in *D. mauritiana* chromosomes (Bryan et al., 1990). The transposition products recovered are identical in structure to those recovered from *D.melanogaster* (Lidholm et al., 1993). These results increase the likelihood that *mariner* is capable of plasmid to chromosome transposition, potentially forming the basis of a germline transformation system in these and other non-drosophilid species, especially those insects of medical and agricultural importance. Such efforts have been successful in the mosquito (Coates et al., 1998; Moreira et al., 2000).

The *mariner* transposable element is capable of interplasmid transposition in the embryonic soma of the yellow fever mosquito, *Aedes aegypti*. Subsequently, a genetic transformation experiment showed that *Mos1* could integrate into the germline of the mosquito (Coates et al., 1998). This experiment exploited the recently demonstrated ability of a wild-type copy of the *Drosophila melanogaster cinnabar* (cn^+) gene to complement the white-eye phenotype of the *kynurenine hydroxylase-white* (kh^w) strain of the *Ae. Aegypti*. The analysis of *mariner* has shown that it functions as a heritable, stable and efficient mediator of gene insertion into *Ae. aegypti*. The availability of a *mariner* transformation system greatly enhances our ability to study and manipulate this important vector species, and eventually generate strains that are incapable of transmitting a specific pathogen.

Gueiros-Filho and Beverley have reported the trans-kingdom transposition of the *Drosophila mariner* within the protozoan *Leishmania* (Gueiros-Filho and Beverley, 1997). They transfected *L.major* cells with two types of plasmid, one containing an intact *Mos1* element and the other containing the *Mos1* transposase-coding region fused with DNA sequences allowing trans-RNA splicing and gene expression in *Leishmania*. Transposition in *Leishmania* was efficient, occurring in more than 20 percent of random transfectants. Their findings reinforce the impression that *mariner* can cross distant species boundaries even different kingdoms separated by an evolutionary distance of more than 1 billion years. This is the widest evolutionary distance yet shown to be traversed by transposable elements *in vivo* and suggests that *mariner*'s potential ability may be comparably broad.

Not only can *mariner* elements be shuttled from flies into mosquitoes and protozoa, but they can also be used as transgene vectors for vertebrates. Transgenic animals have been generated by microinjection of *mariner*-containing plasmid vectors into eggs of zebrafish (Fadool et al., 1998) and chicken (Sherman et al., 1998). One to two-cell-stage zebrafish embryos were coinjected with a supercoiled plasmid carrying the nonautonomous *mariner* element *peach* and mRNA encoding the transposase. The strategy was to rely upon the host's cellular machinery to translate the mRNA into functional Mos1 transposase. Four of the 12 founders transmitted the element to their progeny, and inheritance of the transgene from the F1 to the F2 generation showed a mendelian pattern. A plasmid carrying an active *mariner* element was injected into the cytoplasm of the germline disc of chick zygotes. Germline transmission of *mariner* from one of three surviving birds confirmed

transposition. Once integrated into the chicken genome, *mariner* elements are stable as is observed after transposition into the *D.melanogaster* genome. These results, along with the data showing transposition of *mariner* in zebra fish, support the development of *mariner* as a vector for transgenesis in vertebrates. A general transformation system for insertional mutagenesis and gene tagging therefore can be developed in a vertebrate model in the presence of the transposase.

The *Mos1* element is only one member of the *mariner/Tc1* superfamily which has potential for development as vectors for transgenesis. Successful transposition of other members within this family in cell lines has been obtained for diverse vertebrate species (Izsvak et al., 2000) including mouse ES cells (Luo et al., 1998; Izsvak et al., 2000) and human cells (Zhang et al., 1998; Schouten et al., 1998; Izsvak et al., 2000). *Sleeping Beauty*, a synthetic transposable element made from defective copies of an ancestral *Tc1*-like fish element has recently been shown to function in adult mice (Yant et al., 2000). Chromosomal transposition resulted in long-term expression of human blood coagulation factor IX at levels that were therapeutic in a mouse model of haemophilia B. This is the first demonstration of transposition -mediated gene transfer in an adult animal and represents an advance in the development of stable non-viral gene transfer system. Furthermore, these data suggest that members of the *mariner/Tc1* family may have widespread use as transformation vectors in mammals.

1.8 Organisation and dynamics of transpososome

Transposition is initiated by the formation of a highly organised nucleoprotein synaptic complex or transpososome (Mizuuchi et al., 1992; Kleckner et al., 1996). The assembly of this elaborate transposition machinery proceeds through a series of ordered steps, requiring multiple proteins, multiple DNA sites and co-operative protein-protein and protein-DNA interactions. The assembly process ensures that all the pieces are firmly bolted in place before the machine will run. This tight regulatory safeguard ensures that transposition occurs in the right place at the right time, and avoids damage to the host genome by preventing partial reactions that cannot be completed. This is understood in some detail from studies on phage *Mu*, *Tn10*, *Tn7* and *Tn5*.

1.8.1 Protein-DNA complexes in *Mu* transposition

Mu transposition occurs in the context of several higher order protein-DNA complexes. The earliest characterised synaptic complex is the LER in which the two *Mu* ends (L and R) and the enhancer (E) are brought together by a complex of the MuA transposase and the host HU and IHF proteins (Waston and Chaconas, 1996). The LER is converted to the type 0 complex (or SSC, stable synaptic complex), in which the enhancer has been released and the two *Mu* ends are held together by a tetramer of MuA, no strand cleavage has yet occurred (Mizzuchi et al., 1992). In the presence of Mg^{2+}, transposase catalyses 3'-end cleavage to yield the cleaved donor complex (CDC, or Type 1 transpososome) (Craigie and Mizuuchi, 1987). In the CDC, MuA is stably bound to its three endmost binding sites (L1, L2 and R1) (Mizuuchi et al., 1991). In the presence of MuB, strand transfer of 3' ends into a random site on a target DNA molecule generates a Type 2 complex (or STC, strand transfer complex) (Surette et al., 1987). After strand transfer, MuA is liberated from the complex and exchanged for the replication machinery necessary to complete formation of the cointegrate transposition product. The action of the ClpX chaperone induces a conformational change in the Type 2 transpososome and weakens the interaction which holds it together (Kruklitis et al., 1996). The product of this transformation is the Type 3 or STC2 transpososome. Previous studies led to the proposal of a model in which complete transposition require DDE residues from each subunit of the tetramer (Baker et al., 1994). More recent studies show that two active sites, each containing a set of DDE residues from a separate transposase subunit, catalyse complete transposition (both cleavage and strand transfer of the same *Mu* DNA end). Of the two paired *Mu* DNA ends bound within the transpososome, the subunit supplying the DDE residues for cleavage of one DNA end is bound to the partner *Mu* DNA end (catalysis in trans). The catalytic activity of these two active sites is coupled such that the co-ordinated joining of both *Mu* DNA ends is favoured during transposition (Williams et al., 1999).

1.8.2 Protein-DNA complexes in *Tn10* transposition

Tn10, a composite bacterial transposon, comprises two *IS10* elements (R and L) flanking sequences including a gene for tetracycline resistance (Kleckner et al., 1996). *Tn10* transposition involves assembly of a stable protein-DNA transpososome that includes a pair of transposon ends plus transposase protein. A precleavage synaptic complex, also called a paired ends complex (PEC) has been defined *in vitro* (Sakai et al., 1995). More

recent studies show that stable *Tn10* pre-cleavage transpososomes occur in two conformations: a folded form which contains the DNA-bending factor IHF and unfolded form which lacks IHF. Transpososome assembly must proceed via a folded intermediate which, however, must be unfolded in order for strand transfer to target DNA (Sakai et al., 2000). In the presence of Mg^{2+} or Mn^{2+}, the PEC undergoes double-strand cleavage to generate single-end break complex (SEBc) or double-end break complex (DEBc) (Sakai et al., 1995). Kinetic data suggest that the DEBc is generated from the SEBc (Haniford and Kleckner, 1994). Cleavage of two strands at each end occurs in a sequential order: nicking of the transferred strand always precedes nicking of the nontransferred strand (Bolland and Kleckner, 1995). This allows the liberated 3'-OH of the first strand to attack the nontransferred strand forming a hairpin intermediate (Kennedy et al., 1998). *Tn10* synaptic complex can capture target DNA only after transposon excision (Sakai and Kleckner, 1997). A stable non-covalent DEB-target DNA cocomplex has been identified (Junop and Haniford, 1997). Postcleavage target capture is remarkable for *Tn10*, while *Mu* and *Tn7* select a target site prior to cleavage (Baker et al., 1991; Bainton et al., 1993). *Tn10* transposase carries out four distinct chemical reactions: first-strand nicking, hairpin formation, hairpin resolution (or second-strand nicking), and strand transfer. A single active site on a single monomer (per transposon end) is used repeatedly to catalyse all four steps (Bolland and Klekner 1996; Kennedy et al., 1998). A transposase dimer within the synaptic complex may carry out the entire transposition reaction (Bolland and Klekner, 1996).

1.8.3 Protein-DNA complexes in *Tn7* tranposition

Bacterial transposon *Tn7* (Barth et al., 1976) is distinguished among transposons in that its transposition requires multiple transposon-encoded proteins: TnsA, TnsB, TnsC, TnsD and TnsE (Craig 1996). The different combinations of Tns proteins provided regulated and differential signals for transposition. TnsABC+D mediates *Tn7* insertion into a unique site *attTn7* on the *E.coli* chromosome at high frequency, while TnsABC+E direct the transposon into many non-*attTn7* sites at a relatively low frequency (Rogers et al., 1986; Waddell and Craig 1988; Kubo and Craig 1990). *Tn7* transposase is composed of TnsA and TnsB, which introduces double-strand breaks at the transposon ends and joins the element to the target DNA (Bainton et al., 1991; Bainton et al., 1993). The restriction enzyme-like protein TnsA (Hickman et al., 2000) carries out DNA breakage at the 5' ends

of the transposon, while the retroviral integrase-like protein TnsB cuts the 3' ends and catalyses subsequent strand transfer (May and Craig, 1996; Sarnovsky et al., 1996). TnsA and TnsB also play architectural roles in the synaptic complex in addition to possessing enzymatic activities. The C-terminal domain of TnsA is likely to mediate protein-protein interactions with another member of the transpososome (Hickman et al., 2000). TnsC, a transposition regulator, activates the TnsAB transposase by protein-protein interaction between TnsA and TnsC (Lu and Craig. 2000). *Tn7* is directed to *attTn7* by the sequence-specific binding of the TnsD targeting protein to *attTn7* (Bainton et al., 1993). The binding of the TnsD to *attTn7* can recruit TnsC to *attTn7*. TnsC interacts with TnsA+B to recruit the transposon ends bound by TnsA+B to this target site and to provoke the DNA breakage and joining reactions (Craig, 2000). A notable feature of *Tn7* transposition is that the double-strand breakage does not begin until the specific target DNA is recognised. For transposition into *attTn7*, all the components of the reaction: TnsA, TnsB, TnsC, TnsD, the donor transposon DNA and the target *attTn7* DNA must be present for initiation of recombination (Bainton et al., 1991; Bainton et al., 1993). This assembly requirement is an important "check point" control that can decrease the possibility of inappropriate recombination by preventing initiation in the absence of an appropriate target and also by co-ordinating events at both ends of the element (Craig, 1997).

1.8.4 Protein-DNA complexes in *Tn5* transposition

Tn5 is a composite, prokaryotic transposon consisting of two *IS50* elements (termed *IS50R* and *IS50L*) flanking a region of DNA encoding three antibiotic resistance genes (Reznikoff, 1993). Tn5 transposes by a cut and paste mechanism (Goryshin and Reznikoff, 1998). First, a monomer of *Tn5* binds each OE (outside end) sequence (Zhou and Reznikoff, 1997). Then, interactions of these bound transposase monomers through transposase-transposase dimerization form a synaptic complex which must precede DNA cleavage and has a precise rigid architecture (Goryshin et al., 1994). The *Tn5* pre-cleavage synaptic complex (PEC) is the simplest transposition complex described to date. It contains a dimer of transposase, two transposon ends, and is competent for DNA cleavage in the presence of Mg^{2+} (Bhasin et al., 2000). The C-terminal helix of transposase (residues 458-468 of 476 total amino acids) is required for synaptic complex formation during *Tn5* transposition (Steiniger-White and Reznikoff, 2000). The three-dimensional structure of *Tn5* transposase complexed with *Tn5* transposon end DNA demonstrate that the molecular

assembly is dimeric, where each double-strand DNA molecule is bound by both protein subunits, orienting the transposon ends into the active sites (Davies et al., 2000). Comparison with the transposase-DNA complex of phage *Mu* indicates a similar architecture: a subunit bound to one DNA end cleaves and joins the other end (cleavage of DNA in *trans*). These two examples show a common arrangement of protein-DNA interactions within the synaptic complex and thus provide a framework for coupling cleavage and joining between the two ends during transposition (Williams and Baker, 2000).

Tn5 is unique among prokaryotic transposable elements in that it encodes a special inhibitor protein identical to the *Tn5* transposase except lacking a 55 amino acids N-terminal DNA binding sequence (Johnson et al., 1982). This protein regulates transposition through non-productive protein-protein interactions with transposase. The inhibited complex containing one molecule of inhibitor protein and one molecule of DNA-bound transposase is defective for later stages of the transposition reaction (de la Cruz et al., 1993; York and Reznikoff, 1996). Transposase/inhibitor share two distinct dimerization domains (Weinreich et al., 1994; Braam and Reznikoff, 1998) and the C-terminal domain is necessary for inhibition (Braam et al., 1999), correlating with the ability of the inhibitor protein to homodimerize via this domain (Davies et al., 1999).

1.9 Scope of the thesis

As an initial step of the transposition process, a transposase must first recognise and interact with the ends of its cognate transposable element to form a highly organised transposition complex. The transposase can then catalyse the endonuclease cleavage and strand transfer reactions that follow. Specific DNA binding ability is therefore an essential characteristic of every transposase. *Mariner* transposase, like other enzymes of this type, must have a sequence specific DNA binding ability that allows it to recognise its substrate, the termini of *mariner* element. The DNA binding activity of *Mos1* transposase will be described in Chapters 3 and 4.

Dimerization is another fundamental property of many transposase of both prokaryotic and eukaryotic origins. It is important in the assembly of the active synaptic complex, in ensuring collaboration between two ends for catalysis and in regulating transposase

activity. The role of protein-protein interactions in *Mos1* transposition will be described in Chapters 5 and 6.

Little is known about the promoter activity of *mariner* transposable element. The primary aim of Chapter 7 is to identify those sequences within *Mos1* gene which direct transcription of a reporter gene.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Bacterial Media

Luria Broth (LB)

Difco Bacto tryptone, 10 g; Difco Bacto yeast extract, 5 g; NaCl, 5 g; per litre adjusted to pH7.2.

Luria agar (L-agar)

Luria broth with 15 g/L Difco Bacto agar.

Ampicillin was added to LB or L-agar to a final concentration of 100 μg/ml when required.

SOC

100 ml LB with 1.8 ml 20% glucose, 1 ml 1 M $MgSO_4$ and 1 ml 1 M $MgCl_2$

2.1.2 *Drosophila* Media

French fly food

Oxide No.3 agar, 7.5 g; polenta, 55 g; dried flake yeast, 55 g; nipagen (150mg/ml made up in 95% ethanol) , 10 ml; dH_2O, 100 ml.

Apple juice agar

Difco Bacto agar , 9 g; sucrose, 10 g boiling in 300 ml dH_2O, then 100 ml pure apple juice added and boiled.

Tomato juice agar

Difco Bacto agar , 4 g boiling in 160 ml dH_2O, then 40 ml pure tomato juice added and boiled.

2.1.3 Yeast Media

YPDA

Difco Bacto peptone, 10 g; Difco yeast extract 10 g; glucose 20 g; per litre. For plates add 2g Difco Bacto agar.

Drop-out yeast medium

Yeast nitrogen base without amino acids, 6.7 g; D-glucose 20 g; drop-out powder mix , 2 g; per litre. For plates add 20 g Difco Bacto agar. Amount of 1M 3-Amino 1,2,4 Triazole (3-AT) is added after cooling the agar to 60°C.

M9 Drop-out plates

0.5 g drop-out mix; 10 g Bacto agar in 368 ml dH_2O and autoclaved. Let cool to 60°C, add 125 ml M9x4; 5 ml 20% glucose; 1 ml 1M $MgSO_4$; 1 ml 100 mM $CaCl_2$; 0.5 ml 100 mg/ml ampicillin.

2.1.4 Solutions

10 x TBE

Tris, 108 g; Boric acid , 55 g; EDTA, 7.44 g; made up to 1 L with dH_2O.

10 x TAE

Tris 48.4 g; Acetic acid, 11.4 ml; EDTA, 9.5 g; made up to 1 L with dH_2O.

10 x TGS

Tris, 30 g; Glycine, 144 g; SDS, 10 g; made up to 1 L with dH_2O.

10 x PBS

NaCl, 80 g; KCl, 2 g; Na_2HPO4, 14.4 g; KH_2PO4, 2.4 g; per litre adjusted to pH7.4 with NaOH.

Stacking gel buffer

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

Running gel buffer

Tris, 18.15 g; SDS, 0.4 g, per 100ml adjusted to pH8.9 with HCl.

Stain

1 g Commassie brilliant blue R-250 dissolved in 250 ml methanol , then add 250 ml dH_2O, 50 ml acetic acid.

Destain

50 ml methanol , 50 ml acetic acid, 400 ml dH_2O.

6 x Boiling Mix

Stacking gel buffer, 1 ml; 25% SDS, 0.8 ml; ß-mercaptoethanol, 0.5 ml; 100% glycerol, 1 ml; Bromophenol blue, 0.05%.

Denaturation buffer

NaCl, 87.66 g; NaOH, 20 g, made up to 1 L with dH_2O.

Neutralisation buffer

NaCl, 87.66 g; Tris, 60.5 g; per litre, adjusted to pH 7.5 with HCl.

20 x SSC

Tris-sodium citrate, 88.23 g; NaCl, 175.32 g, per litre.

Prehybridisation buffer

Na_2HPO4, 25.56 g; NaH_2PO4, 8.4 g; SDS. 35 g; 0.5 M EDTA, 1 ml; made up to 500 ml with dH_2O.

6 x Agarose gel loading buffer

30% glycerol; 0.1% bromophenol blue.

TE

10 mM Tris; 50 mM EDTA; adjust to pH 8.0

Western blot transfer buffer

Tris, 9 g; glycine, 43.2 g; made up to 3 L with dH_2O.

10 x TBS

Tris, 60.5 g; NaCl, 87.6 g; per 1 L adjusted to pH 7.5 with HCl. For TBST, 0.1% Tween 20 was added into 1 x TBS.

10 x MOPS

MOPS, 20.9 g; NaAc, 5.44 g; 0.5 M EDTA, 10 ml, per 500 ml adjusted to pH 7.0 with NaOH.

Drosophila DNA extraction buffer

100 mM Tris, pH 9; 100 mM EDTA, pH 8; 1% SDS.

Injection buffer

5 mM KCl; 0.1 mM $NaPO_4$, pH 6.8.

Yeast plasmid extraction buffer

10% Triton X-100, 4 ml; 10% SDS, 2 ml; 5M NaCl, 0.4 ml; 1M Tris-HCl (pH 8.0), 0.2 ml; 0.5 M EDTA (pH 8.0), 40µl; and 13.36 ml dH_2O; per 20 ml.

2.1.5 Isotopes

α-^{32}P-dCTP (3000 Ci/mM, 10 µCi/µl).

2.1.6 Bacterial strains (*E.coli*)

Strain	Genotype	Comments
XL1-Blue (Stratagene)	*Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 Lac[F'proAB lac*Iq*lacZ*ΔM*15*Tn*10(tet*^{r}*)*]	for recombinant DNA manipulation with blue/white screening, Tet-resistant
DH5α (Hanahan, 1983)	*supE44 Δlac*U*169 (ø80 lacZ*ΔM*15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	for recombinant DNA manipulation with blue/white screening

MC1066 (P. Legrain, Institute Pasteur)	*galU galK strA^r leuB6 trpC-9830 pyrF74::Tn5(Kn^r)hsdr$^-$ Δ(LacIPOZYA)74*	for selecton on Leu and Trp in yeast two-hybrid system.
DH10B (GibcoBRL)	*F mcrA Δ(mrr-hsdRMS-mcbBC) ϕ80dlacZΔM15 ΔlacX74 deoR recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ$^-$ rpsL nupG*	high efficiency competent cells for *in vitro* transposition assay
BL21(DE3) (Novagen)	*F ompT hsdS$_B$(r$_B^-$m$_B^-$) gal dcm (DE3)*	For high-level expression of recombinant proteins using bacteriophage T7 promoter.

2.1.7 Yeast Strain (*S.cerevisiae*)

Strain	**Genotype**	**Comments**
L40 (Hollenberg et al., 1995)	*MATa, ade2, his3Δ200, trp1-901, leu2-3, -112, ade2, lys2-801am, LYS2::(lexAop)4-HIS3, URA3::(lexAop)8-lacZ*	containing *HIS3* and *lacZ* reporter genes downstream of LexA binding sites.

2.1.8 *Drosophila melanogaster* strain

Strain	**Genotype**	**Comments**
w1118 (Hazelrigg et al., 1984)	white$^-$	for P element-mediated germline transformation.

2.1.9 Plasmids

Name	Reference	Comments
pGEM-T	Promega	for cloning of PCR products; white/blue screen; Amp-resistance
pMOS	Medhora et al., 1991.	pBluescript containing the *Mos1* element present in a 5kb genomic clone.
pET-15b	Novagen	for high-level expression of recombinant protein; fusion protein could contain C-terminal 20 aa HisTag; Amp-resistance.
pACTIIST	Gift from Jean Begg's lab	Activation domain (AD) cloning vector used in two-hybrid system. The fusion protein contains the GAL4 AD.
pBTM116	Gift from Jean Begg's lab	Binding domain (BD) cloning vector used in two-hybrid system. The fusion protein contains the complete LexA coding sequence (1-202).
pCaSpeR-AUG-ßgal	Thummel et al. 1988	P-element-mediated transformation vector. This vector provides a Drosophila start codon for translation of the *lacZ* gene.
pCaSpeR-ßgal	Thummel et al. 1988	P-element-mediated transformation vector. It requires the insertion of foreign promoter and a start codon in-frame with the *lacZ* gene.
pπ25.7wc	Karess and Rubin, 1984	Helper plasmid encodes P transposase used in P-element-mediated transformation.
pBCP378	Velterop et al., 1995	For direct, overexpression of recombinant proteins from *trc* promoter.

2.1.10 Oligonucleotide primers

Name	Sequences (5' to 3')	Comments
N6799	GC*CATATG*TCGAGTTTC GTGCC	+strand primer from 172-188 of *Mos1*, *Nde*I site.
A6499	CG*CATATG*TTAATGCAA AAACGACTTCC	-strand primer from 621-605 of *Mos1*, *Nde*I site, stop codon.
N30(+)	G*CATATG*CTTGTTGAAG CCTTTGG	+strand primer from 262-281 of *Mos1*, *Nde*I site.
N130(-)	G*CATATG*TTACTCCATC TGCCTCTCGTTC	-strand primer from 561-543 of *Mos1*, *Nde*I site. stop codon.
N120(-)	G*CATATG*TTACACCCAT CTACCGACCTT	-strand primer from 531-514 of *Mos1*, *Nde*I site, stop codon.
N100(-)	G*CATATG*TTATTGACTT ACTTCCAACTG	-strand primer from 471-454 of *Mos1*, *Nde*I site, stop codon.
Y1(+)	CATG*CCATGG*ACATGTC GAGTTTCGTGCCG	+strand primer from 170-189 of *Mos1*, *Nco*I site.
Y345(-)	CCG*CTCGAG*TTATTCAA AGTATTTGCCGTCG	-strand primer from 1209-1188 of *Mos1*, *Xho*I site, stop codon.
Y170(+)	CATG*CCATGG*CATACGT TGATCCTGGACAAC	+strand primer from 680-700 of *Mos1*, *Nco*I site.
Y177(-)	CCG*CTCGAG*TTACGGTT GTCCAGGATCAACG	-strand primer from 702-684 of *Mos1*, *Xho*I site, stop codon.
Y30(+)	CATG*CCATGG*TGCTTGT TGAAGCCTTTGGC	+strand primer from 263-283 of *Mos1*, *Nco*I site.
Y90(+)	CATG*CCATGG*AACAACT CGCAGAGCAGTTG	+strand primer from 440-459 of *Mos1*, *Nco*I site.

Y300(-)	CCG*CTCGAG*TTA**GAAGC GCTGCTCAGCGAG**	-strand primer from 1071-1054 of Mos1, XhoI site, stop codon.
pACT(+)	TACCACTACAATGGATG ATG	+strand primer, 110 bp upstream of *Nco*I cloning site of pACTIIst.
pACT(-)	GAAATTGAGATGGTGCA CGATGCAC	-strand primer, 50 bp downstream of *Xho*I cloning site of pACTIIst.
pBTM(+)	GACCTTCGTCAGCAGAG CTTC	+strand primer, 50 bp upstream of EcoRI cloning site of pBTM116.
pBTM(-)	GAGTCACTTTAAAATTT GTATACAC	-strand primer, 70 bp downstream of *Pst*I cloning site of pBTM116.
R106(+)	**GTTTCCAATGCCTTGC GAGAGATG**	+strand primer from 478-501 of *Mos1* for R106A mutation.
R106(-)	**CATCTCTCGCAAGGCA TTGGAAAC**	-strand primer from 501-478 of *Mos1* for R106A mutation.
K112(+)	**GAGATGGGAGCGATTC AGAAGGTC**	+strand primer from 496-519 of *Mos1* for K112A mutation.
K112(-)	**GACCTTCTGAATCGCT CCCATCTC**	-strand primer from 519-496 of *Mos1* for K112A mutation.
P(+)	CG*GAATTC***CCAGGTGTA CAAGTAGGGAATG**	+strand primer from 1-22 of *Mos1*, *Eco*RI site.
P(-)	CG*GGATCC***GTTGACTGC ACTGAGAGTAAAC**	-strand primer from 171-150 of *Mos1*, *Bam*HI site.
P344(-)	CG*GAATTC***GTTCAAAGT ATTTGCCGTCGC**	-strand primer from 1206-1187 of *Mos1*, *Eco*RI site.

P1	CG*TCTAGA***CCAGGTGTA CAAGTAGGGAATG**	+strand primer from 1-22 of *Mos1*, *Xba*I site.
P861(-)	CG*GAATTC***GAAGCGCAC GGTTCAAATTG**	-strand primer from 862-843 of *Mos1* , *Eco*RI site.
P516(-)	CG*GAATTC***GCTTCTGAA TCTTTCCCATCTC**	-strand primer from 516-496 of *Mos1*, *Eco*RI site.
LacZ(+)	GCCGGTCTGGGAGGCAT TGGTCTG	+strand primer containing the Adh sequences of pCaSpeR-AUG-ßgal. RT-PCR for *lacZ*.
LacZ(-)	GGCCTCAGGAAGATCGC ACTCCAG	- strand primer containing *lacZ* sequences of pCaSpeR-AUG-ßgal. RT-PCR for *lacZ*.
Cys(-)	*CTCGAG***TTATTCAAAGT ATTTGCCGTCGCTAGC TACAGCTTTTTC**	- strand primer from 1209-1171 of *Mos1*, *Xho*I site for C336A mutation.
T345(-)	C*GGATCC***TTATTCAAAG TATTTGCCGTCG**	- strand primer from 1209-1188 of *Mos1*, *Bam*HI site for subcloning mutant *Mos1* into pBCP378.
T7	TAATACGACTCACTATA GG	+ strand sequencing primer of pGEM-T and pET-15b.
SP6	ATTTAGGTGACACTATA GAATAC	-strand sequencing primer of pGEM-T.

Mos1 sequences are shown in bold letters.

Restriction sites are shown in italics.

Mutated codons are underlined.

2.2 Methods

2.2.1 Manipulation of bacteria

2.2.1.1 Bacterial growth and storage

A single colony of *E.coli* was inoculated into LB supplemented with the appropriate antibiotic (100 μg/ml ampicillin) in conical flasks or glass bottles with a capacity of 5 times of the culture. The culture was incubated at 37°C with vigorous shaking overnight. For long term storage, 0.85 ml of culture was mixed with 0.15 ml of sterile glycerol and stored at -70°C. For recovery, the bacteria was streaked out from frozen culture onto a LB-agar plate or LB with antibiotic when required.

2.2.1.2 Transformation of bacteria

Calcium chloride method

A single colony of *E.coli* was added into 100 ml of LB broth in a 1 litre flask and incubated with shaking for approximately 3-6 h until the cell density was 4-7 x 10^7 viable cells/ml (OD_{600} of 0.4-0.6). The culture was transferred into ice-cold 50 ml tubes, chilled on ice for 10 min and then centrifuged at 4000 rpm for 10 min at 4°C. The pellet was then resuspended in the original culture volume of an ice-cold $CaCl_2$/glycerol solution (60 mM $CaCl_2$ and 15% glycerol, filter sterilised) incubated on ice for 30 min. The cells were then collected by centrifugation as above, resuspended in 1/2 of the original culture volume of an ice-cold $CaCl_2$/glycerol solution, and incubated on ice for 10 min. The cells were then collected by centrifugation again, resuspended in 1/30 of the original culture volume of an ice-cold $CaCl_2$/glycerol solution, and stored on ice overnight. The following morning the cell suspension was dispensed as 80 μl aliquots into prechilled eppendorfs and stored at -70°C. 2 μl of plasmid DNA in ligation buffer or 100 ng of plasmid DNA were added to 80 μl of competent cells thawed slowly on ice. After 30 minutes on ice, cells were heat shocked at 42°C for 2 min. 80 μl of SOC buffer was then added and incubated at 37°C for 1 hour. The culture was spread onto a suitable selective plate and incubated at 37°C overnight. For blue/white selection, 20 μl of 100mM IPTG and 35 μl of 50 mg/ml X-gal were spread onto the plates which were then incubated for 30 minutes for absorption prior to use.

Electroporation method

A single colony of *E.coli* was inoculated onto 10 ml of 2xYT medium and grown at 37°C overnight with vigorous shaking. The cells were chilled on ice and harvested by centrifugation at 4,000 rpm for 10 min at 4°C. Cells were washed twice by resuspension in

50 ml ice-cold dH_2O followed by centrifugation. Finally the cells were resuspended in 140 µl of ice-cold dH_2O. 40 µl cells were mixed with 1 µl DNA and transferred to an electroporation cuvette (0.2cm, Invitrogen). A single pulse at 2.5kV, 15µF, 200Ω was applied. 1 ml SOC buffer was added immediately and the mixture transferred to a small glass bottle. Cells were then incubated at 37°C with shaking for 20-60 min before plating out on a suitable selectable plate.

2.2.2 Nucleic acid preparation and manipulation techniques

2.2.2.1 Plasmid DNA preparations

Mini-preparation of plasmid DNA

Plasmid DNA was prepared by QIAprep Spin Miniprep Kit (QIAGEN), according to manufacturer's instructions. Single colony or frozen culture bacteria were inoculated into 5 ml of LB broth containing antibiotic and grown overnight at 37°C with shaking. The cells were collected by centrifugation at 3,000-4,000 rpm for 10 min and resuspended in 250 µl of Buffer P1 (100 µg/ml RNase A, 10 mM EDTA, 50 mM Tris-Cl , pH 8.0). The cells were then lysed by the addition of 250 µl of Buffer P2 (200 mM NaOH, 1% SDS) and mixed gently by inversion of the tube several times. This was neutralised with 350 µl of chilled Buffer P3 (3.0 M potassium acetate, pH 5.5), mixed immediately, and then centrifuged at 13,000 rpm for 10 min. The supernatant was transferred to QIAprep spin columns and then centrifuged at 13,000 rpm for 1 min. The columns were washed by adding 0.5 ml of buffer PB and centrifuging at 13,000 rpm for 1 min, then adding 0.75 ml of buffer PE and centrifuging twice at 13,000 rpm for 1 min. Finally the column received 30 µl of dH_2O and was centrifuged at 13,000 rpm for 1 min.

Midi-prep of plasmid DNA

Large scale preparations of Plasmid DNA were prepared using QIAfilter Plasmid Midi Kit (QIAGEN), following the manufacturer's instructions. A single colony from a freshly streaked selective plate was inoculated into 5 ml of LB broth containing antibiotic and grown for 8 hours at 37°C with shaking. 75 µl of the starter culture was transferred to 50 ml of selective LB medium and grown 12-16 h at 37°C with shaking. The bacteria were harvested by centrifugation at 6,000 x g for 15 min at 4°C and the pellet resuspended in 4 ml of buffer P1 as above. The cells were lysed by adding 4 ml of buffer P2 (200 mM NaOH, 1% SDS). The lysate was mixed gently and incubated at room temperature for 5

min. After the addition of 4 ml of buffer P3 (3 M potassium acetate pH 5.5) and mixing, the solution was then poured into the QIAfilter Cartridge and incubated at room temperature for 10 min. The cell lysate was filtered into the previously equilibrated QIAGEN-tip 100. The tip was washed with buffer QC (1.0 M NaCl; 50 mM MOPS, pH7.0; 15% isopropanol) and DNA was eluted in buffer QF (1.25 M NaCl; 50 mM Tris-Cl, pH 8.5; 15% isopropanol). The DNA was then precipitated by 3.5 ml of isopropanol, and centrifuged immediately at 13,000 x g for 40 min at 4 ^{o}C. The DNA pellet was washed with 2 ml of 70% ethanol and centrifuged at 13,000 rpm for 10 min. The DNA pellet was air-dried and then redissolved in 120 µl of TE.

Plasmid "Bulk prep" for germline transformation

The first several steps are same as Midi-prep. After the cell lysate was poured into the Qiafilter Cartridge and incubated at room temperature for 10 minutes and filtered, the filtrate was mixed with 6 ml of isopropanol and centrifuged at 6,000 rpm for 15 minutes. The DNA pellet was washed with 2 ml of 70% ethanol and dissolved in 1 ml dH_2O. An equal volume of cold 5M LiCl was added and the tube was kept on ice for 5 minutes. After centrifugation, the supernatant was precipitated by equal volume of isopropanol. The DNA pellet was dissolved in 200 µl of dH_2O. An equal volume of PEG/NaCl (15% PEG, 1.6M NaCl) was added and the tube was kept on ice for 5 minutes. After centrifugation, the pellet was air-dried and redissolved in 300 µl of dH_2O. The DNA solution was then extracted with phenol/chloroform and ethanol precipitated before resuspension in 300 µl of dH_2O for injection.

2.2.2.2 RNA preparations

Total RNA from Drosophila ovaries was isolated using RNeasy Mini Kit (QIAGEN). 30 pairs of ovaries were homogenised with 350 µl of Buffer RLT (50 mM Tris-Cl, pH8.0; 140 mM NaCl; 1.5 mM $MgCl_2$; 0.5% NP-40; 1% ß-mercaptoethanol) and a rotor stator homogenizer. The lysate was then centrifuged at 13,000 rpm for 3 min. The supernatant was transferred to a fresh tube and mixed with an equal volume of 70% ethanol. 700 µl of the sample was applied to an RNeasy mini spin column and centrifuged at 13,000 rpm for 15 sec. The RNA was washed with Buffer RW1 and Buffer PRE. The RNA was eluted twice with 30 µl of RNase-free dH_2O after the column was incubated at 65^{o}C for 5 min. The A_{260} and A_{280} of the preparation was measured to verify that the A_{260}/A_{280} ratio was higher then 1.9.

2.2.2.3 Quantification of DNA and RNA

The concentration of DNA and RNA in solution was estimated by measuring their absorbences at 260 nm in spectrophotometer. It was assumed that an OD_{260} of 1.0 is equivalent to a concentration of 50 µg/ml for double stranded DNA, 35 µg/ml for oligonucleotides, and 40 µg/ml of single-stranded RNA. The purity of DNA and RNA was estimated by measuring its absorbance at 260 nm and 280 nm in a spectrophotometer. A good DNA preparation should have a A_{260}/A_{280} ratio higher than 1.8 while a pure RNA preparation has an A_{260}/A_{280} ratio of 1.8-2.1.

2.2.2.4 Phenol extraction of proteins from DNA

DNA was thoroughly mixed with an equal volume of phenol: chloroform: isoamylalcohol (25:24:1). The phases were separated by centrifugation for 5 min. The upper aqueous phase was then extracted with an equal volume of chloroform: isoamylalcohol (24:1) to remove any residual phenol. After spinning the upper aqueous phase was transferred to a fresh tube.

2.2.2.5 Ethanol precipitation of DNA

DNA in solution was precipitated by addition of 1/10 volume 3M sodium acetate (pH5.2) and 2.5 volumes of 100% ethanol. The solution was mixed thoroughly and incubated on ice for 20 min. DNA was pelleted by centrifugation at 13,000 rpm for 15 min. Following removal of the supernatant, the pellet was washed with ice-cold 70% ethanol and spun again for 3 min. The supernatant was removed and the DNA pellet was dried for 15 min at room temperature before dissolved in dH_2O or TE.

2.2.2.6 Restriction enzyme digestion of DNA

DNA was digested by restriction enzyme in 20 µl of appropriate 1x restriction buffer, in a 37°C water bath for 1-20 h.

2.2.2.7 Agarose gel electrophoresis

Agarose gel eletrophoresis was carried out in 0.6-2% agarose made up in 1x TAE or 0.5X TBE containing 0.5 mg/ml ethidium bromide. Prior to loading, DNA samples were mixed with 1/6 volume of 6x loading buffer. A potential difference of 1-10V per cm gel was used

to separate DNA fragments. DNA was visualised and photographed on a UV transilluminator.

2.2.2.8 Purification of PCR products/DNA fragments

DNA fragments of 70bp-10kb were purified from agarose gels using QIAquick Gel Extraction Kits (QIAGEN Inc.). PCR reaction or enzyme-digested DNA products were separated on agarose gel and then the DNA fragments were excised from the gel. 3 volumes of Buffer QG was added to 1 volume of gel and incubated at 50°C for 10 min. 1 gel volume of isopropanol was added to the sample, pipetted into a QIAquick spin column and then centrifuged at 13,000 rpm for 1 min. 0.5 ml of buffer QG was added to the column and it was centrifuged again. 0.75 ml of buffer PE was added to the column and it was centrifuged twice at 13,000 rpm for 1 min. 30 µl of water was added to the column and it was allowed to stand for 1 min and then centrifuged at 13,000 rpm for 1 min.

2.2.2.9 Subcloning of PCR products into pGEM-T vectors

Ligations into pGEM-T vectors were carried out using the Promega pGEMT cloning kit following the manufacturer's instructions. *Taq* DNA polymerase adds a single deoxyadenosine to 3'-ends of both strands for "TA" cloning. Briefly, for each ligation reaction, 2 µl of gel-purified PCR product were mixed with 1 µl of pGEM-T vector, 1 µl of 10x ligation buffer (0.5 M Tris pH 7.4, 0.1 M $MgCl_2$, 0.1 M dithiothreitol, 10 mM spermidine, 10 mM ATP, 1 mg/ml BSA), 1 µl of T4 DNA ligase (3 Weiss Units/µl, Promega) and 5 µl of dH_2O. Ligation mixes were incubated at 16°C overnight, and 2 µl of mixes taken for transformation of *E.coli*.

2.2.2.10 Subcloning of DNA fragments into plasmids

10 µl of plasmid (3 µg) and 10 µl of insert DNA (3 µg) were cut with suitable restriction enzyme in total volume of 20-100 µl at 37°C for 4 h or overnight. When plasmid was cut with single restriction enzyme, 0.2 units of alkaline phosphatase was added into the digestion mix and then incubated at 37°C for 30 min to dephosphorylate the ends and prevent the self-ligation of compatible ends of digested plasmid. Purified cut plasmid and insert DNA were then ligated. Briefly, for each ligation reaction, 2 µl insert DNA, 1 µl of vector, 1 µl of 10x ligation buffer, 1 µl of T4 DNA ligase (3 Weiss Units/ml for cohesive ends and 20 Weiss Units/ml for blunt ends) and 5 µl of dH_2O. The ligation mix was

incubated at 16°C overnight or room temperature for 3 h, and 2 µl of mix used for transformation of *E.coli*.

2.2.2.11 Polymerase chain reaction (PCR)

The PCR is used for enzymatic synthesis of specific DNA sequences The single or double stranded DNA is first denatured, then annealed with specific primers [Tm = 2°C x (A+T) + 4°C x (G+C)] corresponding to each end of the target sequence, and to extend the sequence. This cycle is repeated 20-40 times and has the potential of amplifying the target sequence as much as 10^7-10^9 times. Due to the high temperature required for the denaturation step, the thermostable DNA polymerase, so-called *Taq* DNA polymerase was normally used in the thermal reactions. However, the *pfu* and Vent DNA polymerase could be used instead of *Taq* to improve the PCR accuracy.

PCR was performed in 0.5 ml Eppendorf tubes. Each reaction (100 µl) contained 100 ng of template DNA; 10µl of 10xPCR buffer (500 mM KCl, 200 mM Tris-HCl pH 8.4); 1.5mM of $MgCl_2$, 200µM of each dNTP; 6 µl of each primer at 30 µg/ml or 5 pmol/ml, 2.5 units of *Taq* DNA polymerase (Promega or Qiagen) or Vent® DNA polymerase (New England Biolabs). The reaction mixture was overlaid with mineral oil and PCR was performed using HYBAID Omnigene thermal cycler.

The PCR reaction normally started by incubation the tube at 94°C for 5 min to denature the template, followed by 30 cycles of denaturation at 94°C for 1 min, annealing at 50°C for 1 min and extension at 72°C for 2min. Finally the tube was incubated at 72°C for additional 10 min. The reaction temperature may vary, depending on the primers and templates.

2.2.2.12 Reverse transcriptase-PCR (RT-PCR)

Total RNA was isolated using RNeasy Mini Kit as described in 2.2.2.2. Single-stranded cDNA was reverse transcribed from total RNA according to the GeneAmp RNA-PCR kit protocol (Perkin Elmer, CA). Each reaction (20 µl) contained 1 µl (1 µg) of total RNA, 1 µl of oligo $d(T)_{16}$ (2.5 µM) and 18 µl of master mix (4 µl of 25 mM $MgCl_2$, 2 µl of 10x PCR buffer II, 2 µl of DEPC treated DI water, 2 µl of each dNTP, 1 µl of RNase inhibitor, 1 µl of MuLV reverse transcriptase). All tubes were incubated at room temperature for 10 min to allow for the extension of the oligo $d(T)_{16}$ by reverse transcriptase. The extended primers will then remain annealed to the RNA template upon raising the reaction

temperature to 42°C. The reaction mixture was overlaid with mineral oil and performed: 42°C for 15 min (up to 60 min for synthesis of longer RNA transcripts), 99°C for 5 min (to inactivate the reverse transcriptase), and 55°C for 5 min. The tubes were placed on ice for immediate PCR amplification or stored at -70°C for later use. The PCR conditions were set up as described above using gene-specific primers. The amplified PCR products were analysed by agarose gel electrophoresis and ethidium bromide staining.

2.2.2.13 Rapid PCR screen of recombinant colony

A colony from plate was picked using a 200 µl pipet tip which was touched to a plate to keep a copy of the colony. The tip was touched to 50 µl of sterile water in a 1.5 ml tube, stirred gently to mix. The tubes were placed in boiling water or a heat block at 90-100°C for 5-10 min to lyse the cells and denature DNases, then centrifuged at 12,000 g for 1-2 min. 5 µl of the supernatant was transferred to a 0.5 ml tube for PCR. A master reaction mixture was made by combining the following: pre reaction containing 16.25 µl of sterile water, 2.5 µl of 10x buffer containing 15 mM $MgCl_2$, 0.5 µl of 10 mM dNTP mix, 0.5 ml each primer 1 and 2 (5 pmol/ml or 30 µg/ml), 0.25 µl *Taq* DNA polymerase (5U/µl). 20 µl of the master mix was added to each sample, mixed gently, and overlaid with 2 drops of mineral oil. The reactions were placed in a thermal cycler with the following program: 35 cycles of: 94°C for 1min, 55°C for 1min and 72°C for 2 min; followed by 1 cycle of 72°C for 5min. The PCR reaction products were visualised on a 1% agarose containing ethidium bromide.

2.2.2.14 Site –directed mutagenesis by overlap extension using PCR

PCR amplification with complementary oligonucleotide primers was used to generate two DNA fragments having overlapping ends. These fragments were combined in a subsequent 'fusion' reaction in which the overlapping ends anneal, allowing the 3' overlap of each strand to serve as a primer for the 3' extension of the complementary strand. The resulting fusion product was amplified further by PCR. Specific alterations in the nucleotide sequence were introduced by incorporating nucleotide changes into the overlapping primers. The method is illustrated in Figure 2.1.

2.2.2.15 DNA automatic sequencing

Automated sequencing of plasmid DNA was carried out using dye terminators with the PRISM™ cycle sequencing kit (Applied Biosystems Inc.) and the 377 automated

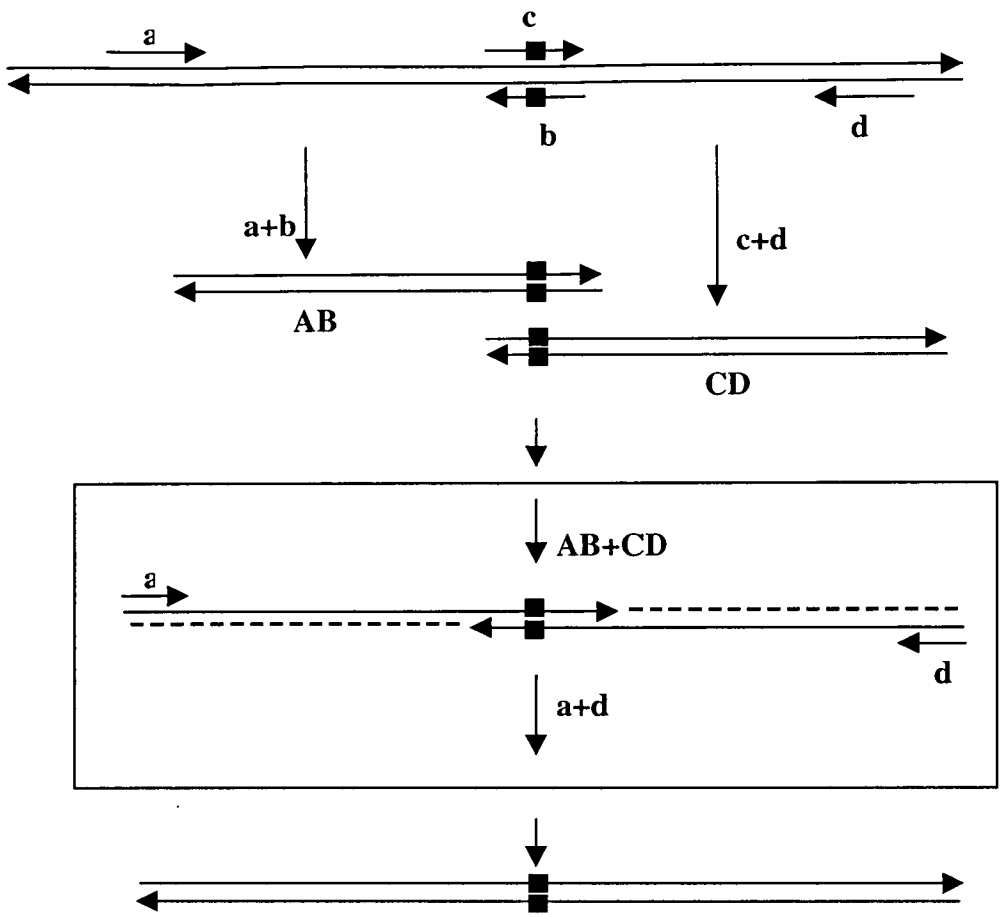


Figure 2.1 Schematic diagram of site-directed mutagenesis by overlap extension.

In separate PCRs, two fragments of the target gene sequence are amplified. Each reaction uses one flanking primer that hybridises at one end of the target sequence (primer "a" or "d") and one internal primer that hybridises at the site of the mutation and contains the mismatched bases (primer "b" or "c").

The two fragments AB and CD, generated in the first PCR, are gel purified. A second PCR is conducted by using equimolar amounts of AB and CD as template and the external primers (a and d) to amplify the mutant fusion product.

The site of mutagenesis is indicated by the small rectangle. Primers are denoted by lower-case letters and PCR products are denoted by pairs of upper-case letters corresponding to the oligo primers used to generate that product.

sequencer (ABI). Reactions were carried out according to the manufacturer's instructions (ABI PRISM dye Terminator protocol, P/N 402078). For each reaction, the reagents were aliquoted into a 0.5 ml PCR tube as follows: 4 µl Terminator Ready Reaction Mix, 2 µl dsDNA Template (0.2 µg/ml), 0.5 µl sequencing primer (20-30 µg /ml), 3.5 µl H_2O with total volume 10 µl. Cycle sequencing was performed on a HYBAID DNA thermal cycler (OmniGene) with the following conditions: 25 cycles of (96°C for 30 sec, 50°C for 15 sec, 60°C for 4 min). After completion, reaction products were mixed with 1µl of 3 M sodium acetate (pH4.6 or 5.2) and 25 µl of cold 95% ethanol, and placed on ice for 10 min. After centrifugation at 13,000 rpm for 20 min, the precipitated pellets were washed again using 125 µl of cold 70% ethanol and finally resuspended in 3 µl loading buffer (5 parts of deionized formamide and 1 part of 50 mg/ml Blue dextran in 25 mM EDTA, pH8.0). Samples were heated at 90°C for 3 min and 1.8 ml of each sample was loaded. 6% Denaturing gels were run at 50 W for 7 h. Chromatograms were viewed by Sequence Navigator program and data was edited using the EditSeq and Align programs.

2.2.2.16 Southern blotting

Blotting DNA was digested with restriction enzymes at 37°C for 24 h and then separated on a 0.8% agarose gel. Following electrophoresis, agarose gels were soaked in denaturing solution then in neutralisation solution, each for 30 min with gentle agitation. DNA was transferred to Hybond N^+ membrane (Amersham) with 20x SSC by capillary blotting. After overnight transfer the membrane was removed and cross-linked with UV prior to hybridisation.

Radioactive random primed labelling of probe The DNA fragments were labelled with α-^{32}P-dCTP using T7 Quickprime^{TM} kit (Pharmacia) according to the instructions of the manufacturer. The DNA template (25-50 ng in TE buffer) and 5 µl of primer were first denatured by boiling for 5 minutes then 10 µl of Labelling buffer was added followed by 50 µCi α-^{32}P-dCTP and 2 µl (2 units) of Klenow enzyme. The mixture was incubated at 37°C for 10 minutes and the reaction was stopped by addition of 0.5mM EDTA. Removal of unincorporated nucleotides was performed by passing through Nuctrap® probe purification columns (Stratagene).

Hybridisation All hybridisations were performed in a Hybridisation oven (Techne). Blots were prehybridised at 65°C in 30 ml prehybridisation solution for 2 h. The boiled, denatured probe was then added and hybridisation continued overnight at 65°C. Non-

specifically bound nucleotides were removed by washing the membrane in 1x SSC , 0.1%SDS and then in 0.1x SSC, 0.1% SDS at 65°C for 15 min each. Filters were then wrapped in Saran film and autoradiographed.

Autoradiography Radioactive filters were exposed to X-ray film (HA west or AGFA Curix) in autoradiographic cassette for an appropriate time. The cassette containing ^{32}P probed filters were stored at -70°C. Film was developed using a X-OGRAPH COMPACT X-2 automatic processor.

2.2.2.17 Radioactive end labelling of DNA

DNA probes used in gel retardation were labelled by the following method. DNA fragment was digested with restriction enzymes to create recessed 3' ends which could be filled by Klenow polymerase. 200 ng of purified DNA was mixed with 3 µl of 20mM dGTP, dATP and dTTP, 20 µCi α-^{32}P-dCTP, and 5 units of Klenow polymerase in 1x DNA polymerase buffer. Following a 30 min incubation at room temperature , the mixture was passed through Nuctrap probe purification column and stored at -20°C.

2.2.3 Manipulation of *Drosophila* flies, cells and tissues

2.2.3.1 Cell culture

Schneider 2 cells (Invitrogen) were maintained at 24°C in Schneider's Drosophila medium supplemented with 10% foetal bovine serum (GibcoBRL), 50u/ml penicillin and 50µg/ml streptomycin.

2.2.3.2 Calcium phosphate transfection of Schneider 2 cells

3 µg of test plasmid and 37.5 µl of 2M $CaCl_2$ was brought to a final volume of 300 µl with tissue culture sterile water. This DNA/$CaCl_2$ mix was added slowly dropwise to 300 µl of the 2xHBS (50mM HEPES, 1.5 mM NaH_2PO_4, 280 mM NaCl, pH7.1) with gentle vortexing. The resulting transfection mix was incubated at room temperature for 30 min to allow formation of the calcium phosphate-DNA precipitate. The 600 µl precipitate was added dropwise to the surface of the Schneider 2 cells which were incubated for 16-24 hours at 24°C. After calcium phosphate solution was removed, the cells were incubated for a further two days with fresh medium.

2.2.3.3 ß-Galactosidase staining

ß-Galactosidase activity was measured by using ß-Galactosidase Staining Kit (Stratagene). The cells were incubated with a glutaraldehyde-formaldehyde fixing solution and then with a staining solution that contained X-Gal. When present, ß-Galactosidase cleaved 5-bromo-4-chloro-3-indoyl-ß-D-galactopyranoside (X-Gal) to produce a blue stain.

2.2.3.4 Preparation of *Drosophila* genomic DNA

20 flies were frozen for 5 minutes at -70°C, then homogenised in solution A (0.1 M Tris-HCl, pH9.0; 0.1M EDTA, pH8.0; 1% SDS). After incubation at 70°C for 30 minutes, 56 µl of 8M potassium acetate was added and samples were left on ice for 30 minutes. Samples were centrifuged at 4°C for 15 minutes at 12,000rpm. The supernatant was removed and recentrifuged. 200 µl of isopropanol was added and DNA left to precipitate for 10 minutes at -70°C. DNA was recovered by centrifugation for 5 minutes, washed with 70% ethanol, dried and resuspended in 40 µl TE.

2.2.3.5 *P* element-mediated germline transformation

Newly eclosed flies from w^{1118} were transferred into fresh bottles, supplemented with yeast paste for 2-3 days, then placed into laying pots. The flies were allowed to lay eggs on the apple juice plates with a little fresh yeast paste dabbed in the centre. Embryos were collected every 40 minutes, dechorionated in 50% bleach and washed onto a gridded filter. After being lined up and stuck onto a coverslip using double-sided sticky tape, embryos were desiccated for 5 minutes in a petri dish containing silica gel. Embryos were covered with halocarbon oil for injection.

A DNA mixture containing 500µg/ml of the plasmid (with the gene of interest) and 250µg/ml of pπ25.7wc (helper plasmid) was back-loaded into the injection needle and injected into the posterior pole of preblastoderm embryos using a transjector 5246(Eppendorf). Older embryos were killed by tearing the vitellin membrane. After incubation on tomato juice plates with yeast paste for 2 days, larvae were transferred to french food. Adults were mated with three w^{1118} virgins of the opposite sex and the progeny examined for coloured-eyed flies. Transformants were mated to three w^{1118} virgins of the opposite sex to obtain males and females with the same insertion. Homozygous lines were then established by sibling matings.

2.2.3.6 Histochemical staining of ovaries and testes

Histochemical staining was carried out on ovaries in females and testes in males. The tissue were dissected from 3-5 day-old flies in Ringer's solution at room temperature. The dissected tissues were placed in staining solution (as described in 2.2.3.3) in micrometer wells for 16-20 hours. They were mounted with 80% glycerol in PBS and photographed.

2.2.3.7 β-galactosidase activity assay

The ovaries were dissected from ten flies and homogenised thoroughly in 500 µl ice-cold homogenisation buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol). Samples were centrifuged at 12,000 rpm in cold room for 15 minutes. 400 µl of the supernatant was preincubated at 37°C for 5 minutes before 600 µl of O-Nitrophenyl-β-galactopyranoside(ONPG) was added and the incubation continued at 37°C. The enzyme reactions were followed by measuring OD_{420nm} at 10 minutes time intervals. The protein concentration of each sample was determined by Bradford method.

2.2.4 Manipulation of yeast

2.2.4.1 Yeast transformation

A single colony of the yeast strain to be transformed was used to inoculate 5 ml of YPDA or 10 ml selectable drop-out medium, and the culture was grown overnight at 30°C until saturation. Cells were then diluted in 50 ml fresh YPDA to an OD_{600} of 0.1 and incubated at 30°C. Upon reaching an OD_{600} of 0.4-0.8, the cells were harvested by centrifugation at 4500 rpm for 5 min. Cells were resuspended in 25 ml dH_2O, recentrifuged and washed in 1 ml of 100mM LiAC, and finally resuspended in 400 µl of 100mM LiAc. Aliquots of 50 µl of the yeast cell suspension were centrifuged for 15 sec at top speed and the LiAC solution was removed. The "transformation mix" consisting of 240 µl of 50% PEG; 36 µl of 1M LiAc; 25 µl of 2 mg/ml denatured salmon sperm DNA and 1 µg plasmid DNA in 50 µl dH2O was added onto each cell pellet in the order listed. The tube was vortexed vigorously until the cell pellet was completely mixed. Cells were incubated at 30°C for 30min, followed by heat shocking at 42°C for 20 min. Cells were pelleted at 6000rpm for 15 sec and resuspended in 1 ml dH_2O. 2 to 200µl of the transformation mix was plated onto selective medium and incubated for 2-3 days to recover transformants.

2.2.4.2 X-Gal-overlay assay

X-Gal-overlay assay was performed directly on the selective medium plates. 10 ml overlay mixture (containing 0.25 M Na_2HPO_4; 0.5% agar; 0.1% SDS; 7% DMF and 0.04% X-GAL) between 50°C and 45°C was pipetted over the plates. The mixture was let to flow out gradually where there was no colonies growing. When the top layer had set, the plates were incubated at 30°C and checked for blue colonies.

2.2.4.3 Yeast plasmid rescue

The experiment allows the recovery of a plasmid from yeast cells by transformation of *E.coli* with a yeast cellular extract. In the two-hybrid system, the diploid cells contain two different plasmids carrying the TRP1(W) and the LEU2(L) markers, respectively. A bacterial strain MC1066 carrying the trp and leu auxotrophies can be complemented by TRP1 and LEU2 yeast genes.

A single colony was inoculated into 2 ml of selective medium overnight at 30°C with shaking. After centrifugation, the pellets were resuspended in 200 µl of extraction buffer and 200 µl of glass beads was added. Samples were then extracted with 200 µl phenol/chloroform and vortexed vigorously for 5 mins. 160 µl supernatant was precipitated with ethanol/NH_4Ac before final resuspension in 10 µl of dH_2O.

One µl of each sample was used to transform MC1066 cells by electroporation. The transformed cells were spread on M9 drop-out plates and the plates were incubated at 37°C for 20 hours. A miniprep was then performed to obtain plasmid DNA.

2.2.4.4 Western blotting

Yeast transformants were grown in 5 ml of selectable medium overnight at 30°C. Cells were then diluted into 50 ml YPDA and incubated to an OD_{600} of approximately 0.5. 3 ml of cells were pelleted and resuspended in 80 µl of "crack buffer" (2% SDS; 80 mM Tris-HCl, pH6.8 ; 10% glycerol; 10 mM EDTA; 0.4mg/ml bromophenol blue; 0.1 M DTT, 2mM PMSF). About 40 µl of glass beads were added and the cells were vortexed for 1 min. After centrifuged for 3 min at 13,000 rpm, the supernatants were boiled for 3 min and fractionated on a 10% SDS polyacrylamide gel.

Proteins separated by SDS-PAGE were transferred onto PVDF (Boehringer Mannheim) membrane, which were pre-soaked in methanol. Transfer was conducted at 40 V for 1.5 h in transfer buffer(9 g Tris, 43.2 g Glycine to 3L) using BIO-RAD Trans-Blot^{TM} cell. The

membranes were incubated in 1% blocking solution (Boehringer Mannheim) overnight, and then incubated with primary antibody (usually 1:1000-2000 dilution in 0.5% block) for 2 h. After washing by TBST for 1 hour with four wash changes, the membranes were incubated with peroxidase-conjugated rabbit anti-mouse IgG (Santa Cruz) for 1 hour. Following washing in several changes of TBST for 1 hour, the bound antibodies were detected by chemiluminescence using Boehringer Mannheim Chemiluminescence Western blotting kit. The membrane was incubated with detection solution for 1 min at room temperature and then exposed with film for 2-30 sec.

2.2.5 Protein detection

2.2.5.1 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

Protein were mixed with 6x SDS-PAGE loading buffer, boiled for 5 min. Samples were loaded on a 10% acrylamide resolving gel, 5% acrylamide stacking gel and run in a BIORAD^{TM} Miniprotein II gel tank in 1x TGS for 40 min at 200V.

To visualise protein, gels were immersed in Commassie brilliant blue stain for 45 min with shaking, then destained overnight.

2.2.5.2 Determination of protein concentration

Protein concentration was estimated by the Bradford method. BSA standards in identical buffer conditions were used to plot a standard curve. Protein samples were mixed with 1 ml Commassie Protein Assay Reagent (Pierce) and absorbance was measured at 595 nm after 2 min incubation at room temperature.

2.2.6 Protein-DNA interactions

2.2.6.1 Gel retardation assay

3 nM of *Mos1*-N150 protein was incubated with 50 ng of non-specific competitor poly dI-dC on ice in binding buffer containing 25 mM HEPES pH7.9, 100 mM NaCl , 5 mM DTT, 5 mM spermidine, 1mg/ml BSA, 20% glycerol, 5 mM $CaCl_2$ for 10 min. 2 ng of DNA probe was added to the mixture and the incubation continued for 1 hours in the cold room. In the competition assay, 10-100 fold molar excess of unlabelled competitor DNA was added. The DNA-protein complexes were loaded on a 5% polyacrylamide gel in 1 x TBE at 150 V in the cold. The gel was then dried under vacuum and autoradiographed.

2.2.6.2 DNase I footprinting

End-labelled DNA fragments were incubated with various amounts of protein in 100 µl of gel retardation buffer for 4 hours in cold room. 250 ng of DNase I was added for a further 2 min of incubation. Digestions were terminated by addition of 325 µl of ethanol, 25 µl of saturated ammonium acetate and 2.5 µl of 1 mg/ml tRNA. Precipitated DNA was washed with 70% ethanol and dried DNA was resuspended in 10ul of TE and 10µl of loading buffer. Samples were heated to 90°C for 5 min before being applied to a 8% polyacrylamide -urea sequencing gel. The G+A sequencing ladder was run in parallel with the samples as a marker to localise the position of the footprints.

2.2.6.3 Preparation of the G+A sequencing ladder

A Maxam-Gilbert guanine and adenine specific cleavage reaction was used to prepare G+A sequencing ladder from the end–labelled DNA fragment used in DNase I footprinting.

2 µl of end-labelled DNA fragment was mixed with 1.5 µl of 1 mg/ml calf thymus DNA in a total volume of 10 µl. The mixture was chilled on ice and 1.35 µl of formic acid added. The reaction was incubated at 37°C for 14 minutes and chilled on ice prior to the addition of 150 µl of freshly prepared 1M piperidine. The cap of the tube was wrapped tightly with parafilm and the reaction was heated at 90°C for 30 minutes. After incubation, the tube was cooled on ice and spun for a few seconds in microcentrifuge. The tube was then spun for 90 minutes under vacuum to remove piperidine. The pellet was resuspended in 100 µl of dH_2O and transferred to fresh tubes before drying again for 30 minutes in a vacuum centrifuge. After the above washing steps were repeated twice, the pellet was resuspended in 10 µl of sequencing loading buffer.

2.2.7 *In vitro* activity of Mos1 transposase

2.2.7.1 Excision assay

200 ng of supercoiled *pMos* plasmid was incubated with Mos1 transposase in the following buffer: 25 mM HEPES (pH 7.5); 100mM NaCl; 10% glycerol; 2 mM DTT; 0.5 mM EDTA; 1µg of BSA; 1 µg of poly d(I-C).d(I-C); 5 mM $Mn(CH_3COO)2$ to a total volume of 20 µl. Samples were incubated in 30°C water bath for 1 hour and then terminated by the addition of 40 µl of STOP buffer (50 mM Tris-Cl, pH7.5; 10 mM EDTA; 0.5 mg/ml

proteinase K; 250 µg/ml yeast tRNA) and incubated at 37°C for 30 min. The reactions were extracted with 25:24:1 phenol-$CHCl_3$-isoamyl alcohol and ethanol precipitated before resuspension in 10 µl of TE buffer. The excised Mos1 element was analysed by Southern blotting following separation on a 1% agarose gel.

2.2.7.2 Transposition assay

300 ng of donor plasmid, pRJM345MosTet was incubated with 200 ng of pBSKA+tetramer in the presence of varying concentrations of Mos1 transposase. The reactions were carried out at 30°C in 25 mM HEPES, pH 7.9; 100 mM NaCl; 10% glycerol; 200 µg/ml BSA and 2 mM DTT for 2 hours in a final volume of 20 µl. The reactions were stopped and extracted as for the excision assay. The nucleic acid pellets were resuspended in 10 µl of sterile water after ethanol precipitation.

One microlitre of each sample was electroporated into DH10B cells. The transformed cells were incubated at 37°C for 45 min with shaking before plating out 0.5 ml on L broth plates containing ampicillin (100 µg/ml) and tetracycline (12 µg/ml) and incubating overnight at 37°C. Viable counts were performed at the same time and the number of ampicillin resistant colonies calculated.

Chapter 3

Overexpression and purification of Mos1-N150 protein

3.1 Introduction

The interactions of a transposase with the ends of its cognate transposable element is the initial step of transposition. The transposase can then catalyse the endonuclease cleavage and strand transfer reactions that follow. Computer analysis of the amino acid sequences of the transposases of several members of *Tc1/mariner* family suggest that they contain a helix-turn-helix motif near their N-termini (Pietrokovski and Henikoff, 1997). The HTH motif has been crystallized in a complex with double-stranded DNA from the amino terminus of the Tc3 transposon in C.*elegans* (van Pouderoyen et al., 1997). For the Mos1 protein, the predicted HTH motif runs from residues 87-108.

In order to investigate the DNA binding properties of the Mos1 transposase and to determine the sequence that it recognises, the work described in this chapter aimed to express and purify soluble C-terminal truncated polypeptide which contains the first 150 amino acids including the HTH motif.

The pET system was used to express the Mos1-N150 fusion protein. The Mos1-N150 fusion protein was constructed to contain a N-terminal histidine tag, facilitating binding to Ni^{2+}-NTA. Although the overexpressed Mos1-N150 was only partially soluble, sufficient soluble protein was produced for the studies described in this thesis.

3.2 Results

3.2.1 Construction of *Mos1*-N150/pET-15b expression vector

Amino acid residues 1-150 of Mos1 transposase were amplified by PCR using *pMos* as template. Primers N6799 and A6499 were designed with *Nde*I restriction site at the 5' ends for insertion and in-frame expression with the pET-15b vectors (Novagen, Madison, WI). The resulting 466 bp product was ligated into pGEM-T-Vector (Promega). By direct cloning of the *Nde*I/*Nde*I PCR product into a "TA" vector, subsequent cloning steps which require restriction at the *Nde*I site can be performed efficiently. The positive transformants were examined by PCR colony screening method. The pET-15b vector was digested with *Nde*I followed by dephosphorylation and ligated to *Nde*l cut *Mos*1-N150/pGEM-T. The positive constructs were sequenced by automatic cycle sequencing (Perkin Elmer) to confirm that the cloned DNA was in the correct reading frame for fusion protein expression. Fig 3.1 shows the fusion protein sequence.

```
        10         20         30         40         50         60
         *          *          *          *          *          *
ATGGGCAGCAGCCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCAT
 M  G  S  S  H  H  H  H  H  H  S  S  G  L  V  P  R  G  S  H>

        70         80         90        100        110        120
         *          *          *          *          *          *
ATGTCGAGTTTCGTGCCGAATAAAGAGCAAACGCGGACAGTATTAATTTTCTGTTTTCAT
 M  S  S  F  V  P  N  K  E  Q  T  R  T  V  L  I  F  C  F  H>

       130        140        150        160        170        180
         *          *          *          *          *          *
TTGAAGAAAACAGCTGCGGAATCGCACCGAATGCTTGTTGAAGCCTTTGGCGAACAAGTA
 L  K  K  T  A  A  E  S  H  R  M  L  V  E  A  F  G  E  Q  V>

       190        200        210        220        230        240
         *          *          *          *          *          *
CCAACTGTGAAAACGTGTGAACGGTGGTTTCAACGCTTCAAAAGTGGTGATTTTGACGTC
 P  T  V  K  T  C  E  R  W  F  Q  R  F  K  S  G  D  F  D  V>

       250        260        270        280        290        300
         *          *          *          *          *          *
GACGACAAAGAGCACGGAAAACCGCCAAAAAGGTACGAAGACGCCGAACTGCAAGCATTA
 D  D  K  E  H  G  K  P  P  K  R  Y  E  D  A  E  L  Q  A  L>

       310        320        330        340        350        360
         *          *          *          *          *          *
TTGGATGAAGACGATGCTCAAACGCAAAAACAACTCGCAGAGCAGTTGGAAGTAAGTCAA
 L  D  E  D  D  A  Q  T  Q  K  Q  L  A  E  Q  L  E  V  S  Q>

       370        380        390        400        410        420
         *          *          *          *          *          *
CAAGCAGTTTCCAATCGCTTGCGAGAGATGGGAAAGATTCAGAAGGTCGGTAGATGGGTG
 Q  A  V  S  N  R  L  R  E  M  G  K  I  Q  K  V  G  R  W  V>

       430        440        450        460        470        480
         *          *          *          *          *          *
CCACATGAGTTGAACGAGAGGCAGATGGAGAGGCGCAAAAACACATGCGAAATTTTGCTT
 P  H  E  L  N  E  R  Q  M  E  R  R  K  N  T  C  E  I  L  L>

       490        500        510
         *          *          *
TCACGATACAAAAGGAAGTCGTTTTTGCATTAA
 S  R  Y  K  R  K  S  F  L  H  *>
```

Figure 3.1 Mos1-N150/pET-15b fusion protein sequence.

The first 20 amino acids derived from pET-15b vector carries an N-terminal 6 His-tag sequence followed by N-terminal 150 amino acids (bold type) of Mos1 transposase.

3.2.2 Expression of Mos1-N150 protein in *E.coli*

The positive *Mos1*-N150/pET-15b construct was transformed into competent BL21(DE3) strain of *E.coli*. Transformed BL21(DE3) cells were incubated in LB containing ampicillin (100μg/ml) at 37°C for about 4 h until OD_{595} reaches 0.6 and then 1mM isopropyl-β-D-galactopyranoside (IPTG, Boehringer Mannheim) was added. The cells were incubated for an additional 4 h at 37°C. Cells were harvested by centrifugation and resuspended in 1 x Binding buffer (5mM imidazole, 500 mM NaCl, 20mM Tris). The cells were lysed by sonication, and the insoluble fraction of the cell lysate was precipitated by centrifugation. Protein expression was analysed by SDS-PAGE. The recombinant fusion protein was expressed at very high levels on induction with IPTG (Fig 3.2) and was expressed in both soluble and insoluble forms (Fig 3.3).

In order to increase the proportion of soluble protein available for use in biological function studies, several variables which might affect the solubility of the protein were considered. The effects of temperature, IPTG concentration, and length of induction time on the Mos1-N150 solubility were each investigated. The cells produced more soluble protein when expression was induced at 25-30°C than 37°C; changing the IPTG concentration from 0.5 mM to 2.0 mM had little effect on the amount of soluble protein. Increasing induction time from 3 h to 5 h, resulted in a small decrease in soluble protein obtained from the same amount of cells. Fig 3.3 shows that the minority of the recombinant Mos1-N150 was soluble while the remainder was insoluble.

3.2.3 Purification of Mos1-N150 protein

Purification of Mos1-N150 protein was first performed according to the Novagen pET System Manual. Pellets from 1000 ml cultures of expressed protein were lysed by sonication in Binding buffer. Following centrifugation, the cell-free extract containing soluble protein was applied to a Ni^{2+}-NTA affinity resin column which was then washed with Binding buffer and Wash buffer (60 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl, pH7.9). The bound Mos1-N150 protein was eluted with Elution buffer (1M imidazole, 500 mM NaCl, 20 mM Tris-HCl, pH7.9) and then dialysed against PBS or water overnight at 4°C to remove the imidazole. Unfortunately, almost all Mos1-N150 protein

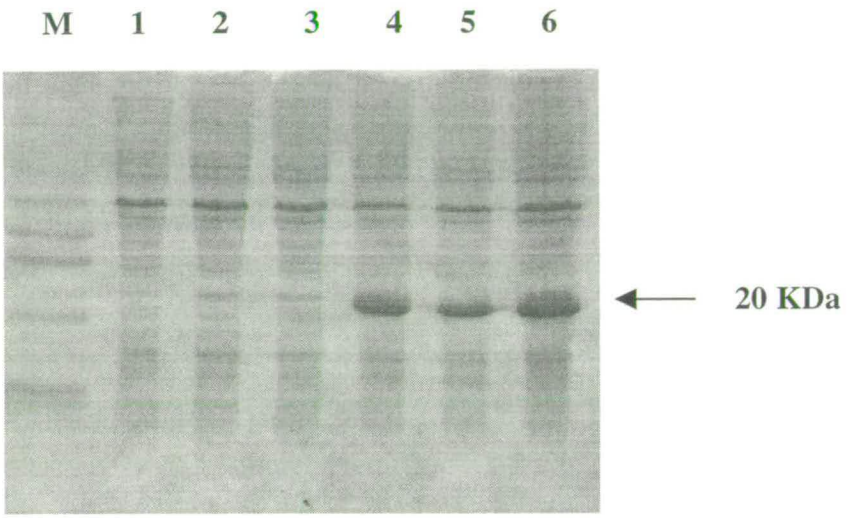


Figure 3.2 Expression of Mos1-N150 protein in different Mos1-N150/pET-15b constructs

Three positive colonies containing *Mos1*-N150/pET-15b constructs (No.1-3) were checked for their expression prior to (lanes 1-3) and following (lanes 4-6) IPTG induction by SDS-PAGE. Lane M, SDS-7 protein markers.

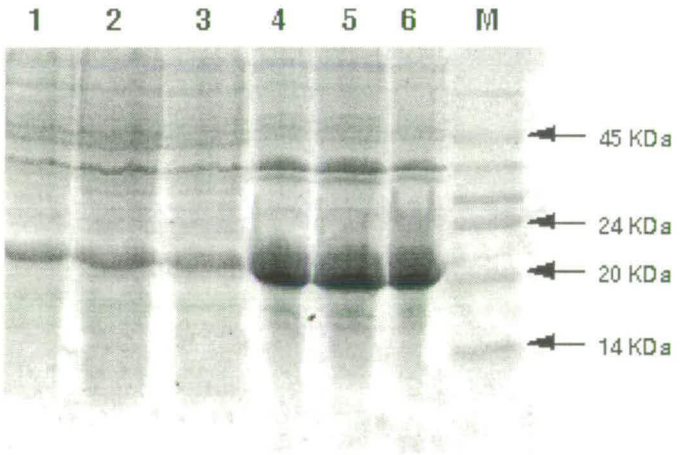


Figure 3.3 Soluble and insoluble Mos1-N150 protein expressed in *E.coli*

Three positive colonies containing Mos1-N150/pET-15b constructs (No.1-3) were induced and lysed by sonication in binding buffer. Following centrifugation, the cell-free extract containing soluble proteins (lanes 1-3) and insoluble pellet (lanes 4-6) were analysed by SDS-PAGE. Lane M contained molecular mass standards as indicated on the right.

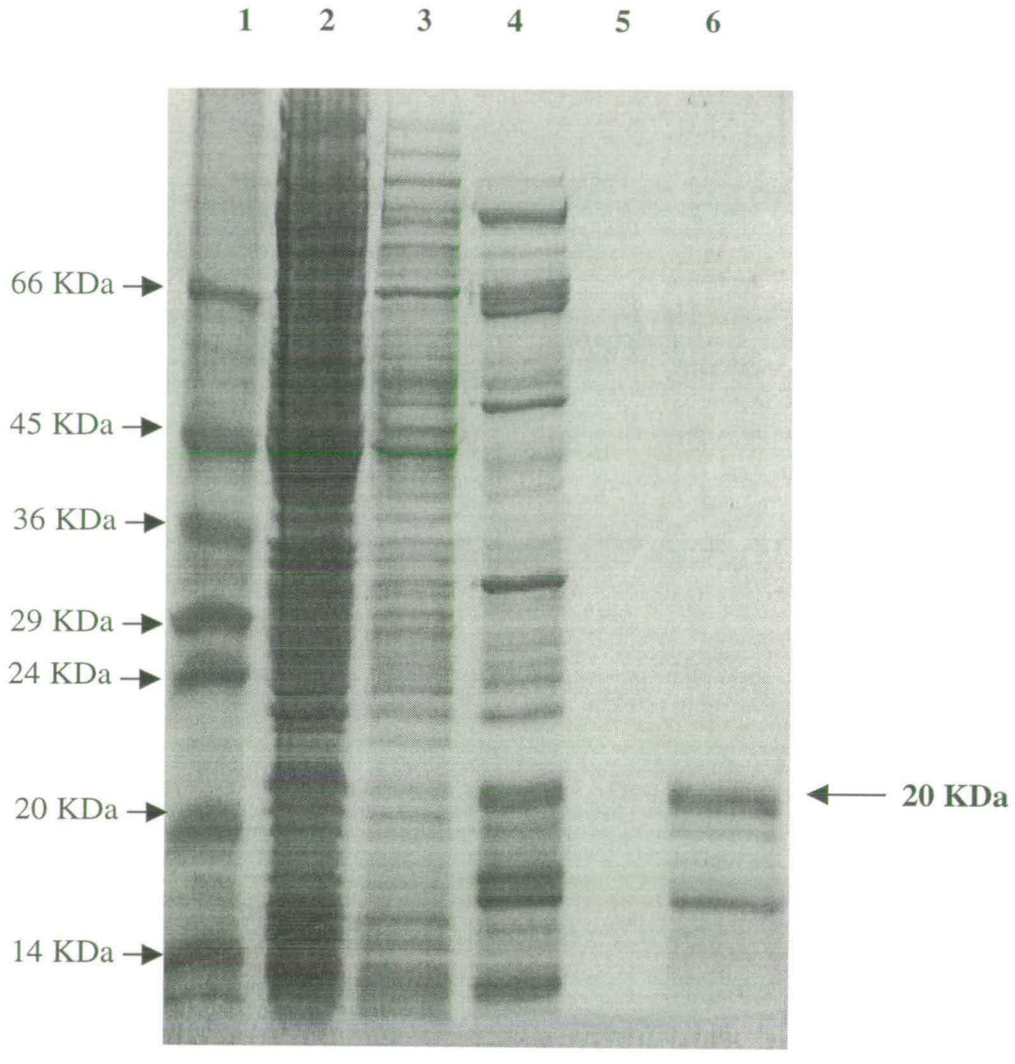


Figure 3.4 Purification of Mos1-N150 protein with His-tag system

The steps of purification of protein were shown by SDS-PAGE. Lane 1, protein markers; lane 2, soluble cell extract overexpressing Mos1-N150 after filtration; lane3, flow-through removed during Binding buffer wash; lane 4, flow-through removed during Wash buffer wash; lane 5, flow-through removed during TBS wash; lane6, purified Mos1-N150 protein after elution with TBSA and dialysis.

precipitated under these conditions, rendering the standard purification protocol unsuitable for purifying Mos1-N150.

A new protocol modified from Holzinger's protocol (Holzinger et al., 1996) was used to purify soluble Mos1-N150. A few ml of deionized water was added to a dry 8mm x 152mm column (PIERCE, polythylene end fittings with 40μm pore disc) to start the flow. The bottle containing His-Bind resin (Novagen) was mixed gently by inversion and 5 ml of resin was transferred to the column. The column was washed with 7.5 ml dH_2O, 12.5 ml of Charge Buffer (50 mM $NiSO_4$) and 7.5 ml of Binding buffer. The cell-free extract containing soluble protein was filtered through a 0.45 micron membrane before loading on a Ni^{2+}-NTA affinity resin column, and the column then washed with 25 ml of Binding buffer, 15 ml of Wash buffer and Tris-buffered saline (TBS, 150 mM NaCl; 20 mM Tris-HCl, pH 7.9). The bound Mos1-N150 protein was eluted with TBS containing 50 mM EDTA (TBSE) and then dialysed against TBS for 24-48 h at 4°C to remove the EDTA. Fig 3.4 depicts the different stages of protein purification, as visualised on a SDS-PAGE gel. The purified recombinant protein had an apparent molecular mass of 20 kDa. A smaller polypeptide was co-purified with the expected band. It was not seen immediately after sonication. This was presumably the result of proteolysis during the affinity purification stage although the protease inhibitor PMSF was added at the concentration of 0.1 mM to prevent against degradation.

The purified protein was dialysed in TBS (150mM NaCl, 20mM Tris-HCl) and concentrated using a 10K Microsep centrifugal concentrator (Filtron). The final concentration of 5mg/ml Mos1-N150 protein in 10% glycerol was aliquoted, frozen in liquid nitrogen and stored at -70°C.

3.3 Discussion

Expression and purification of recombinant proteins is a standard technique in molecular biology. It facilitates production of proteins and allows detailed characterisation of virtually any protein. Recombinant DNA techniques permit the construction of fusion proteins in which specific affinity tags are added to the protein sequence of interest; the use of these affinity tags simplifies the purification of the recombinant fusion proteins by employing affinity chromatography methods. One of the most effective expression and purification system is 6xHis tag expression, allowing the purification of proteins from less

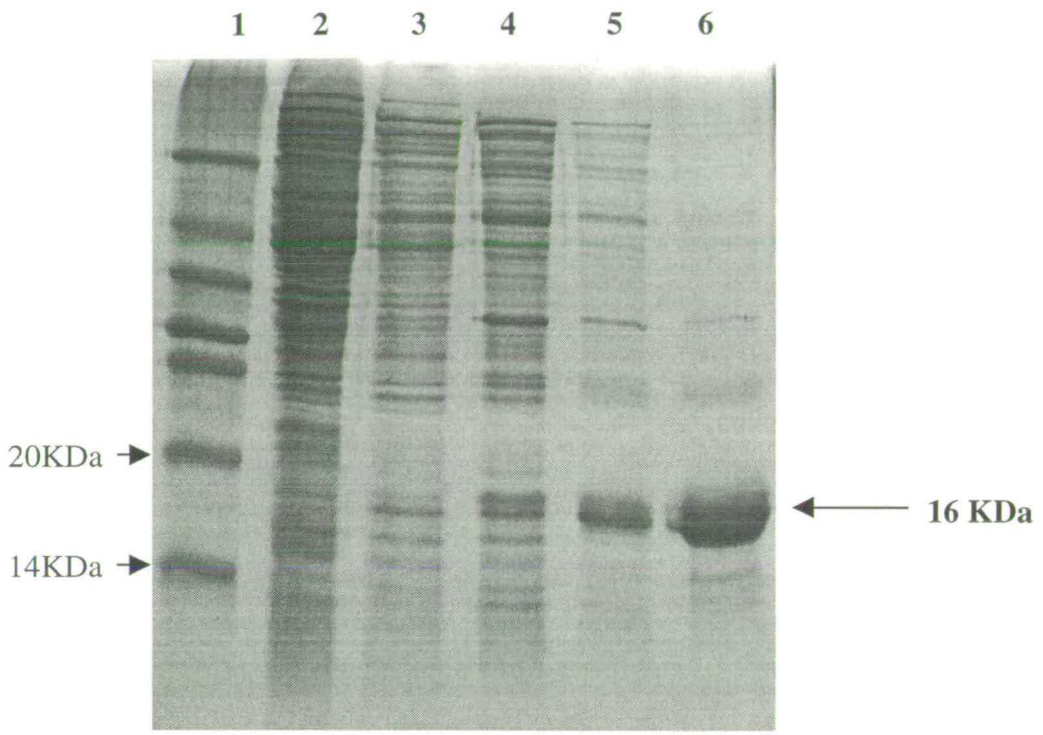


Figure 3.5 Purification of Mos1-N120 protein with His-tag system

Extract of cells containing soluble Mos1-N120 protein was applied to a Ni^{2+}-NTA affinity resin column (lane 2). The column was then washed with Binding buffer (lane 3), Wash buffer twice (lanes 4 and 5). The bound protein was eluted with TBS containing 50 mM EDTA (lane 6). Lane 1, protein markers.

than 1% of the total protein preparation to more than 95% homogeneity in just a single step (Janknecht et al., 1991). In this chapter Mos1-N150 was expressed and purified using a 6xHis tag system in the pET-15b vector.

Although a wide variety of heterologous expression systems, prokaryotic as well as eukaryotic, have been developed, the purification of the proteins obtained can still be problematic. Recombinant proteins expressed intracellularly in *E.coli* can be produced in soluble form, but in many cases, especially at high expression levels, are frequently aggregated and sequestered into insoluble inclusion bodies. It is believed that the intermolecular association of hydrophobic domains during folding plays a role in the formation of inclusion bodies. For proteins with cysteine residues, improper formation of disulphide bonds in the reducing environment of the *E.coli* cytoplasm may also contribute to incorrect folding and formation of inclusion bodies. Several factors, such as incubation temperature, expression level, IPTG concentrations, the genotype of E.*coli* strains and the carrier protein, are thought to affect the solubility of expressed fusion proteins. For the Mos1-N150 fusion protein, decreasing the induction temperature from 37°C to 25-30°C increased the amount of soluble protein recovered. Different concentrations of IPTG were used, but no obvious effect was observed.

The 6xHis tag protein purification system has been widely adopted due to its ease of use in rapidly preparing large amounts of purified protein. The 6xHis tag system takes advantage of the affinity of histidine residues for metal cations, usually divalent cations such as Zn^{2+}, Cu^{2+}, or Ni^{2+} (Yip and Hutchens, 1994). A crude cell extract is then applied to the column and only proteins with a high affinity for the divalent cation will bind to the column. After washing to remove other proteins, the bound protein is eluted from the column. Most commercial systems use imidazole as the elution buffer to recover bound proteins. However, this method posed major problems for Mos1-N150 purification. Although bound Mos1-N150 protein was eluted efficiently with imidazole elution buffer, it nearly all precipitated during dialysis against PBS or water overnight at 4°C to remove the imidazole. Similar observations have been reported for His-tag Bm-SPN-2, a serine proteinase inhibitor (Zang et al., 1999). To purify soluble Mos1-N150, a new protocol was developed, modified from Holzinger's protocol (Holzinger et al., 1996) which was initially used for refolding poorly soluble 6xHis tagged proteins. In this way sufficient amounts of purified Mos1-N150 soluble in TBS were obtained. The small polypeptide co-purified with the expected band was also observed during purification of Mos1-N120, a

smaller derivative of Mos1-N150. When protease inhibitor cocktail (Calbiochem-Novabiochem) was added during the whole procedure of purification, the degradation band can barely been seen (Fig 3.5). Because protease inhibitors may affect the activity of transposase, the protein used in this thesis was purified without this "cocktail".

Chapter 4

DNA binding activity of Mos1 transposase

4.1 Introduction

The first step in transposition is the recognition of the transposon DNA by a specific DNA-binding domain of the transposase protein. In order to carry out the transposition reaction, the transposase must first recognise and synapse two ends of the transposon. After synapsis of the ends, coordinated strand breakage and joining of the transposon ends to the host DNA follow. The transposase of phage *Mu* (MuA) binds to the ends of the *Mu* genome during the assembly of higher order nucleoprotein complexes. Each of the two subdomains at the N-terminus of MuA (Iß and Iγ) contains a helix-turn-helix (HTH) DNA-binding motif and binds to one half of the 22 bp recognition sequence (Clubb et al., 1997; Schumacher et al., 1997). The Drosophila *P*-element site-specific DNA-binding domain maps to the N-terminal 88 amino acids which contains a C_2HC putative metal-binding motif (Lee et al., 1998). *P* element transposase does not interact with the terminal 31 bp inverted repeats but instead interacts specifically with an internal 10 bp consensus sequence present at both the 5' and 3' ends of *P* element DNA (Kaufman et al., 1989). *Tc1*-like transposases contain bipartite DNA-binding domains and have been proposed to consist of two helix-turn-helix motifs (Pietrokovski and Henikoff, 1997). For Tc3A, the Tc3 transposase, it has been shown that the N-terminal domain of 65 amino acids is responsible for specific DNA binding. This domain binds two regions within the 462 bp *Tc3* inverted repeat. Between the N-terminal specific DNA-binding and the catalytic domain, another non-specific DNA binding domain recognises DNA sequences located more towards the cleavage site (Colloms et al., 1994; van Pouderoyen et al., 1997). Such a bipartite DNA binding has also been shown for the related Tc1 transposase Tc1A (Vos and Plasterk, 1994). Both the DNA binding domain of Tc1A and the DNA binding site in the inverted repeat of *Tc1* can be divided into two subdomains. The N-terminal 68 amino acids of Tc1A binds specifically to the sequence between 12-25 bp. The C-terminal part of the bipartite domain , Tc1 amino acids 69-142 , is essential for major groove contacts with base pairs 7-12. The N-terminal 74 amino acids of *pogo* transposase contains "solo" HTH motif that recognises a 12 bp sequence, two copies of which are present at each end of *pogo* DNA (Wang et al., 1999).

A helix-turn-helix DNA binding motif is predicted to lie between residues 87-108 of the Mos1 transposase (Pietrokovski and Henikoff, 1997). In order to determine whether this forms the DNA binding domain responsible for recognising the *Mos1* inverted repeats,

a protein (Mos1-N150) comprising the amino terminal 150 amino acids of the transposase was expressed and purified (see Chapter 3).

The gel retardation assay and DNase I footprinting experiments described in this chapter show that Mos1-N150 binds specifically to inverted repeats at both ends. The minimal DNA-binding domain is identified to be located between amino acid 1 and 120 which contains a helix-turn-helix motif. Site-directed mutagenesis indicates that the HTH is required for sequence-specific recognition of the terminal inverted repeats by Mos1 transposase. The transposase has 5-6 times higher affinity for the right end sequence of *Mos1* than for the left end.

4.2 Results

4.2.1 The sequence specific DNA binding activity of Mos1-N150 protein

A DNA fragment which extends from -58bp of *D.simulans* to +120bp of *Mos1* was end labelled and used as left-end probe, while the right-end probe contains 108 bp from the right-end of *mos1* and 50 bp of flanking *D.simulans*. The purified Mos1- N150 protein was allowed to bind to the probe in the presence of non-specific competitor poly dI-dC. The mobility of the labelled DNA was visualised by autoradiography. Fig 4.1 and Fig 4.2 show that the binding of DNA probe to the Mos1-N150 protein retarded the mobility of the fragment during electrophoresis, resulting in a discrete band corresponding to the protein-DNA complex.

In a competition experiment, unlabelled probe DNA was used as specific competitor, while the first 131 bp DNA sequence of the *pogo* element (Wang et al., 1999), which is of different sequence to that of *mos1*, was used as non-specific competitor. Fig 4.1 and Fig 4.2 show that the DNA-protein complex could be titrated away by a 100 fold molar excess of specific competitor, but not by the same molar excess of non-specific competitor. This suggests that the interaction of Mos1-N150 protein with the transposon end sequence is a specific reaction.

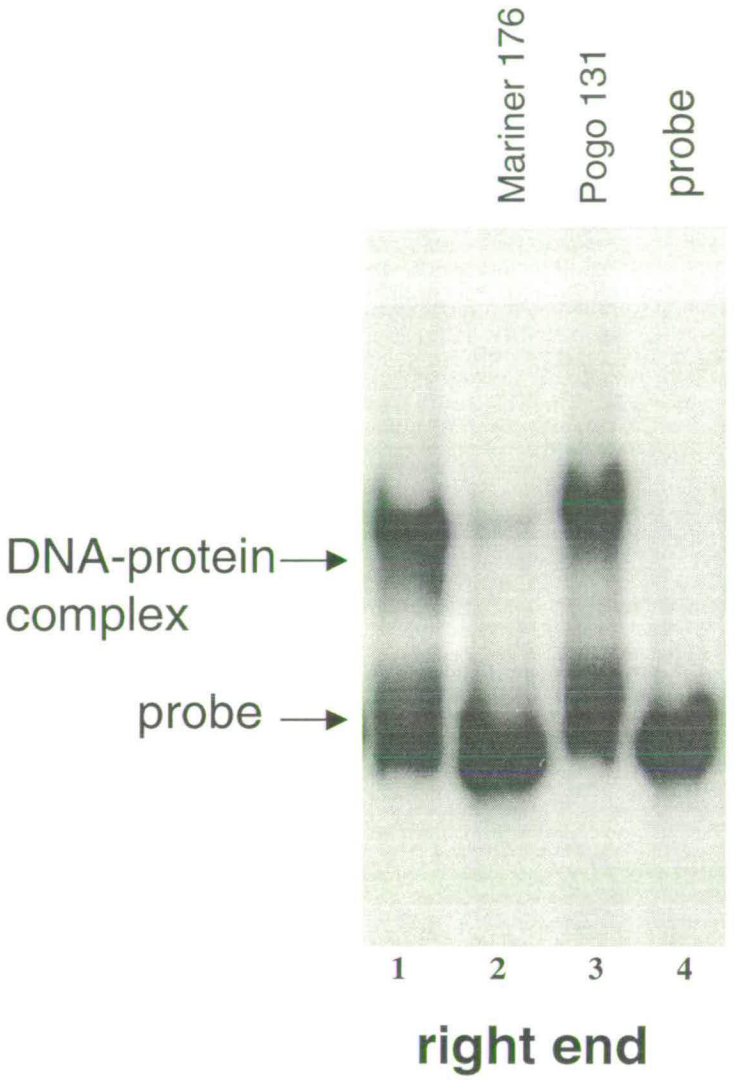


Figure 4.1 The specific binding of Mos1-N150 protein to right-end sequence of *Mos1* element. ^{32}P-labelled DNA fragment was incubated with 10 ng of purified Mos1-N150 protein in the presence or absence of unlabelled competitor DNA. Protein-DNA complex were resolved by native electrophoresis on a 8% polyacrylamide gel. Lanes 1-3, probe plus protein; lane 2, 100-fold molar excess of unlabelled probe was added as specific competitor; lane 3, 100-fold molar excess of a 131 bp fragment from the left end of the *pogo* transposable element was added as non-specific competitor; lane 4, right-end probe alone.

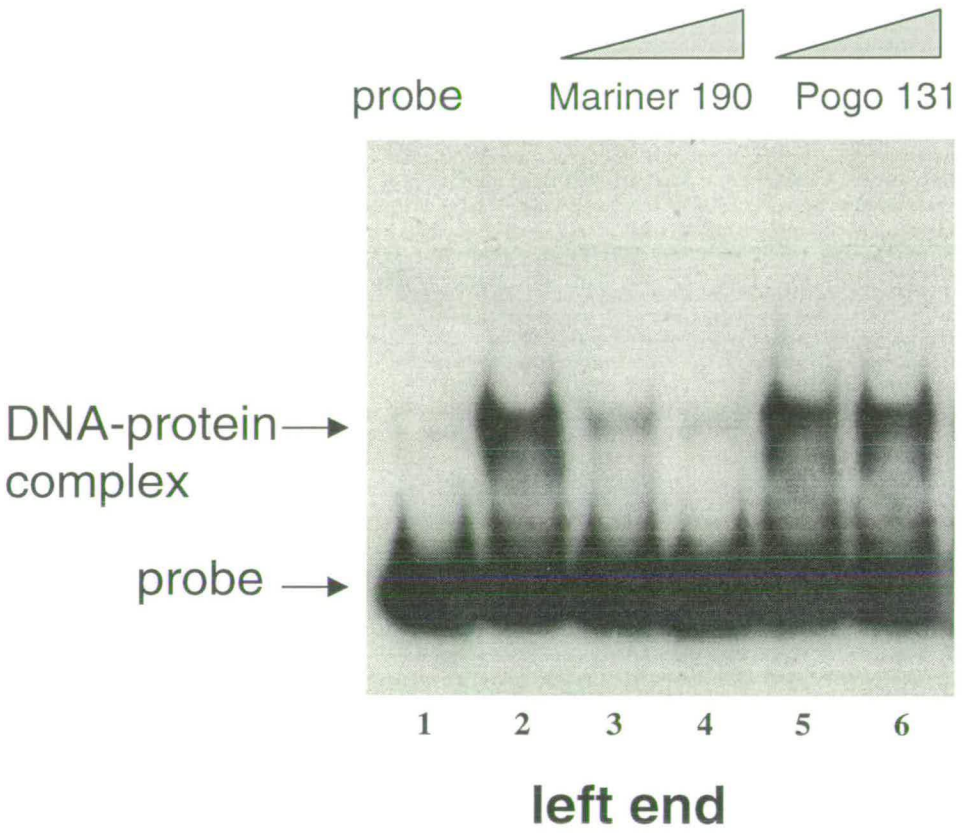


Figure 4.2 The sequence specific DNA binding activity of Mos1-N150 protein to the left-end sequence of *Mos1* element. ^{32}P-labelled DNA fragment was incubated with 10 ng of purified Mos1-N150 protein in the presence or absence of unlabelled competitor DNA. Protein-DNA complex were resolved by native electrophoresis on a 8% polyacrylamide gel. Lane 1, probe alone; lanes 2-6, probe incubated with Mos1-N150 fusion protein; lanes 3 and 4 are identical to lane 2 except for the addition of 10- or 100-fold molar excess of unlabelled probe as competitor; lanes 5 and 6 , incubation with 10- or 100-fold molar excess of a 131 bp fragment from the left end of the *pogo* transposable element as a non-specific competitor.

4.2.2 Mos1-N150 protein binds to *Mos1* terminal inverted repeat sequences

Tc1/mariner superfamily contain inverted repeats at their ends that can be bound by transposase which is thought to be essential for transposition (Vos et al., 1994 ; Colloms et al., 1994). DNase I footprinting was performed to locate the specific binding sites of Mos1-N150 on DNA. The 190 bp left-end DNA fragment and 176 bp right-end sequence, the same as those used for gel retardation, were end-labelled at the 3' end of *Xba*I site on the top strand or 5' end of the *Bam*HI site on the bottom strand. As expected, with the addition of increasing amount of Mos1-N150 protein resulted in the appearance of a protected region encompassing the *Mos1* inverted repeat. At the right end, protection from DNase I cleavage covered 22 nucleotides of inverted repeats on both strands. The top strand was protected from DNase I cleavage from positions 1263 to 1284, and the bottom strand from positions 1261 to 1282. So the protected region on the top strand is offset by 2bp relative to the protected region on the opposite strand (Fig 4.3). At the left end, the top strand was protected from DNase I cleavage between nucleotides 3 and 30, and the bottom strand between nucleotides 3 and 28. There were two DNase I hypersensitive sites located at position 29 and 36 on the bottom strand and one hypersensitive site at position 31 on the top strand (Fig 4.4).

4.2.3 The DNA binding domain of Mos1 transposase

In order to determine the region of Mos1 transposase that is responsible for DNA binding, various deletion derivatives of Mos1 transposase have been expressed in *E.coli* and their DNA binding properties were determined in a gel retardation assay (Fig 4.5A).

A set of expression constructs were prepared by inserting truncated versions of *Mos1* cDNA into the expression vector pET-15b. For in-frame cloning into pET-15b, primers were designed as the following: forward primer N30(+), bp 262-281 of *Mos1*; reverse primers N130(-), bp 561-543 of *Mos1*; N120(-), bp 531-514 of *Mos1* and N100(-), bp 471-454 of *Mos1*. All primers contain *Nde*I site and reverse primers contain stop codons. The PCR product using primers N6799 and N130(-) was subcloned into the *Nde*I site of pET-15b to express Mos1-N1-130. Likewise, N6799/N120(-) product encoded Mos1-N1-120, N6799/N100(-) encoded Mos1-N1-100 and N30(+)/N130(-) encoded Mos1-N30-130.

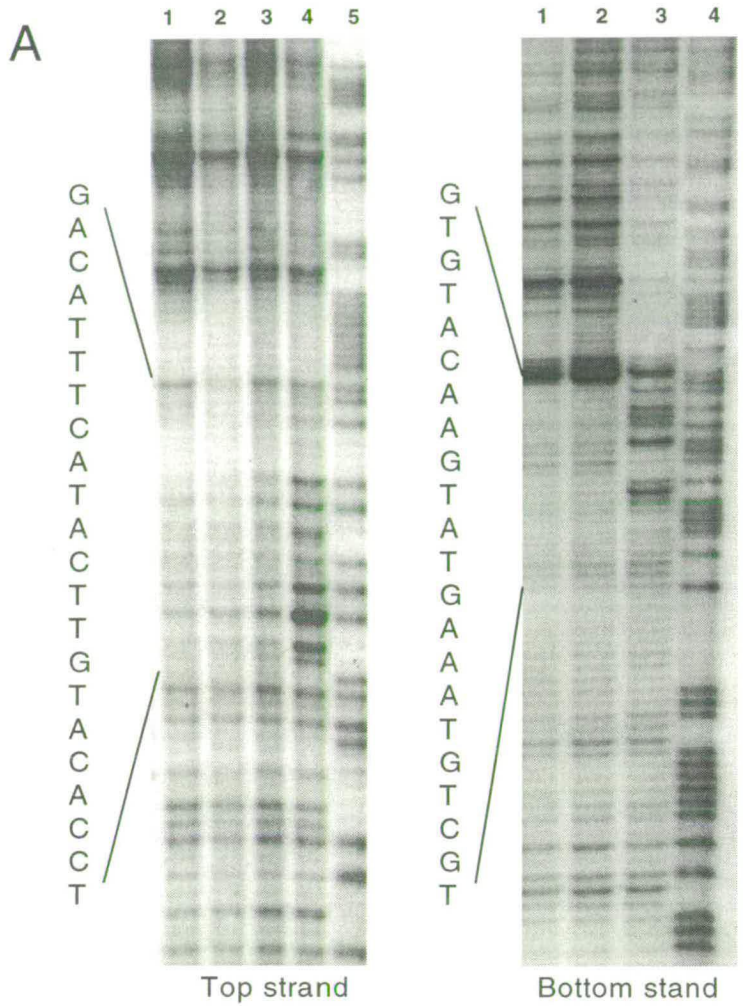


Figure 4.3 DNase I footprinting of the right-end sequence of *Mos1* element. (A) The top strand was labelled at the 3' end of *Xba*I site. Lanes 1-4 contain 10 μg, 2.5 μg, 150 ng, and 0 of Mos1-N150 protein; lane 5 is the G+A ladder. The bottom strand was labelled at the 5' end of *BamH*I site. Lanes 1-3 contain 2.5 μg, 150 ng and 0 of Mos1-N150; lane 5 is the G+A ladder. **(B)** Schematic summary of footprinting results. Brackets show the sequences protected by protein on both strands. Bold type indicates 28 bp inverted repeat of *Mos1* element.

A

B

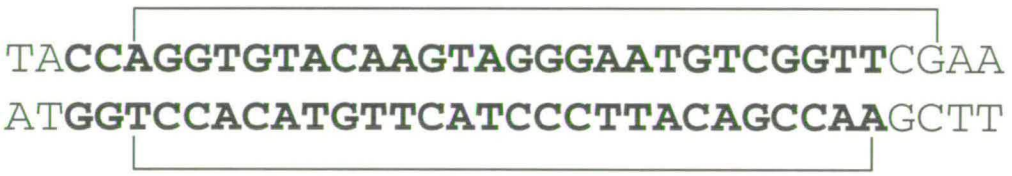


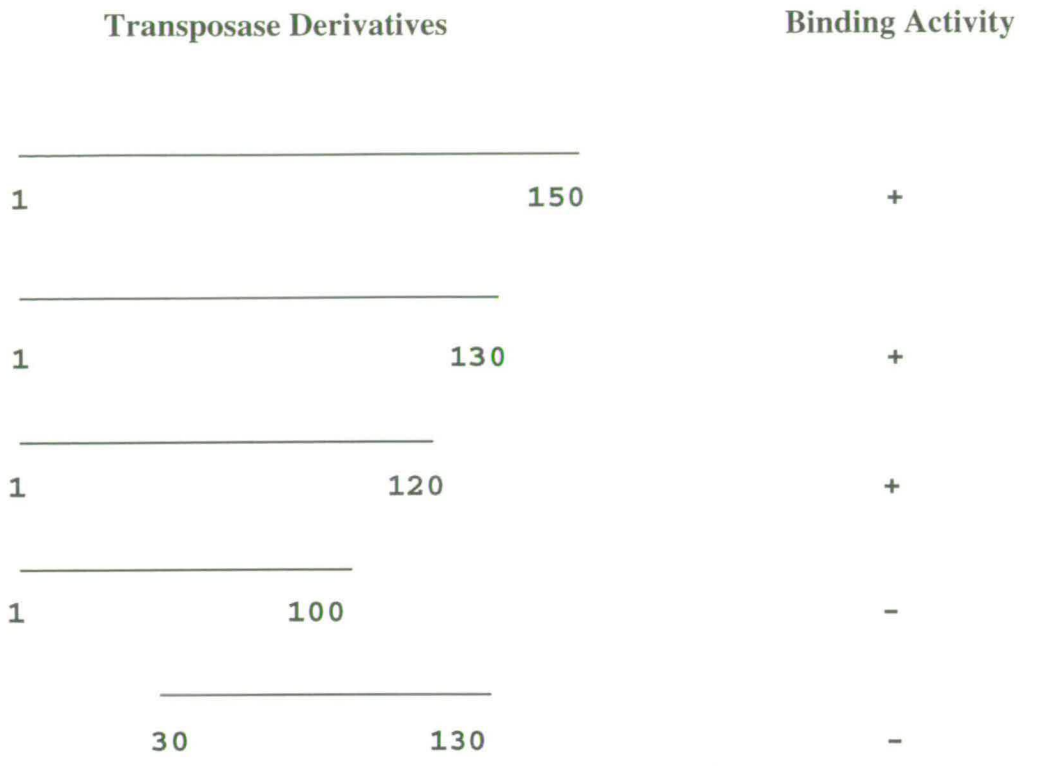
Figure 4.4 DNase I footprinting of the left-end sequence of *Mos1* element. (**A**) The 190 bp fragment of *Mos1* was labelled at the 3' end of *Xba*I site on the top strand. Lanes 1-3 contain 2.5 μg, 150 ng and 0 of Mos1-N150. The DNA fragment was labelled at the 5' end of *BamH*I site on the bottom strand. Lanes 1-3 contain 625 ng, 150 ng and 0 of Mos1-N150. **(B)** Schematic summary of **(A)**. Brackets show the sequences protected by protein on the top or bottom strands, respectively. Stars indicate several DNase I hypersensitive sites. Bold type indicates 28 bp inverted repeat of *Mos1* element.

The positive *Mos1*-N1-130/pET-15b, *Mos1*-N1-120/pET15b, *Mos1*-N1-100/pET15b and *Mos1*-N30-130 constructs were transformed into competent BL21(DE3) strain of *E.coli.* Transformed BL21(DE3) cells were incubated in LB containing ampicillin (100μg/ml) at 37°C for about 3 h until OD_{595} reaches 0.5 and then 1mM isopropyl-β-D-galactopyranoside (IPTG, Boehringer Mannheim) was added. The cells were incubated for an additional 3 h at 37°C. Upon induction, the fusion proteins were well expressed (Fig 4.5B). The Mos1-N1-100 always migrated slower than expected on SDS-PAGE. This abnormal migration may be due to the amino acid composition of this protein. Mos1-N1-100 is relatively rich in acidic amino acids compared to Mos1-N1-120. Higher contents of negatively charged amino acids may restrict the binding of SDS to protein molecules so that they migrate more slowly than expected.

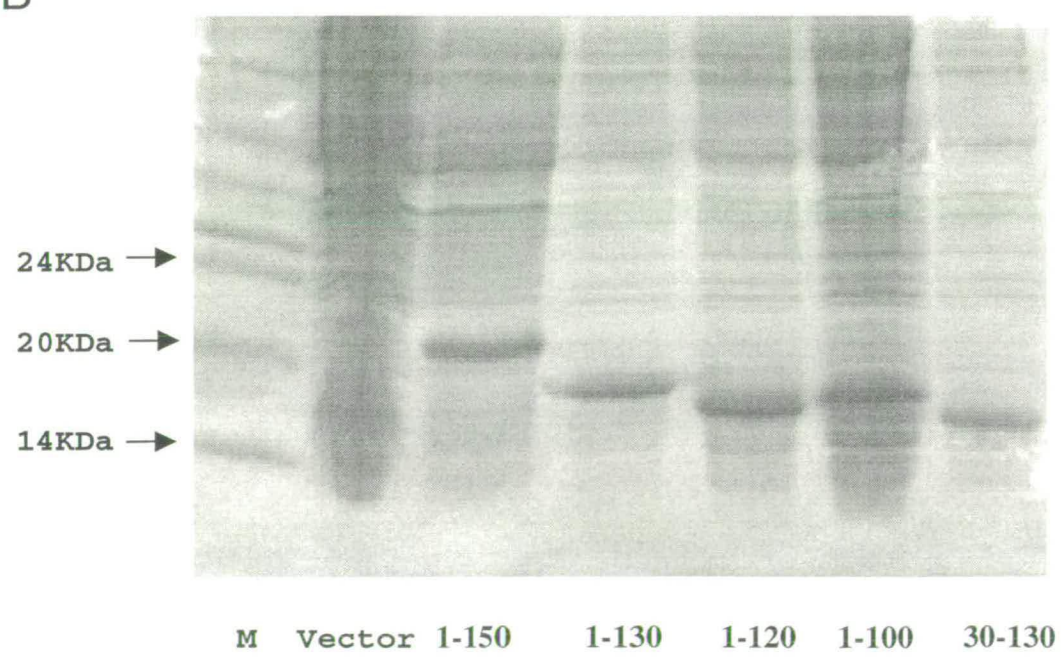
Bacteria were harvested by centrifugation and resuspended in a buffer containing 50 mM HEPES, pH 7.6; 50 mM NaCl; 1 mM EDTA and 1 mM DTT. Cells were broken by sonication and insoluble material was removed by centrifugation at 20,000 g for 30 mins. The crude protein extracts were used in gel retardation assay. Figure 4.5C showed that binding to right end of *Mos1* was unaffected by carboxyl terminal deletions of 30 amino acids of Mos1-N150, but truncation to amino acid 100 abolished binding. Protein Mos1-N30-130 which lacked the first 30 amino acid did not possess site-specific DNA binding activity. This indicates that a sequence specific DNA binding domain lies within the amino terminal 120 residues of Mos1 transposase including the predicted HTH motif which is a typical structure found in DNA binding proteins.

Pietrokovsky and Henikoff (Pietrokovsky and Henikoff, 1997) predicted that transposases of the *Tc1, mariner,* and *pogo* families have a helix-turn-helix DNA binding motif in their N-terminal domains. For the Mos1 protein, this runs from residues 87-108 (Fig. 4.8A). We have investigated whether residues within the putative HTH are essential for DNA binding by site-directed mutagenesis. Positively charged amino acids in the second helix are thought to be involved in specific DNA-protein contacts. The arginine (R) residue in position 18 of the second helix was therefore mutated to alanine (A) (Fig. 4.6). The lysine (K) residue four amino acids downstream was also mutated to alanine as a control (Fig. 4.7). Changing the arginine residue to alanine (R106A) abolished DNA binding activity of Mos1-N150, while changing the lysine 4 residues downstream of the HTH motif

A


Transposase Derivatives
Binding Activity
1
150
+
1
130
+
1
120
+
1
100
-
30
130
-

B


24KDa
20KDa
14KDa
M
Vector
1-150
1-130
1-120
1-100
30-130

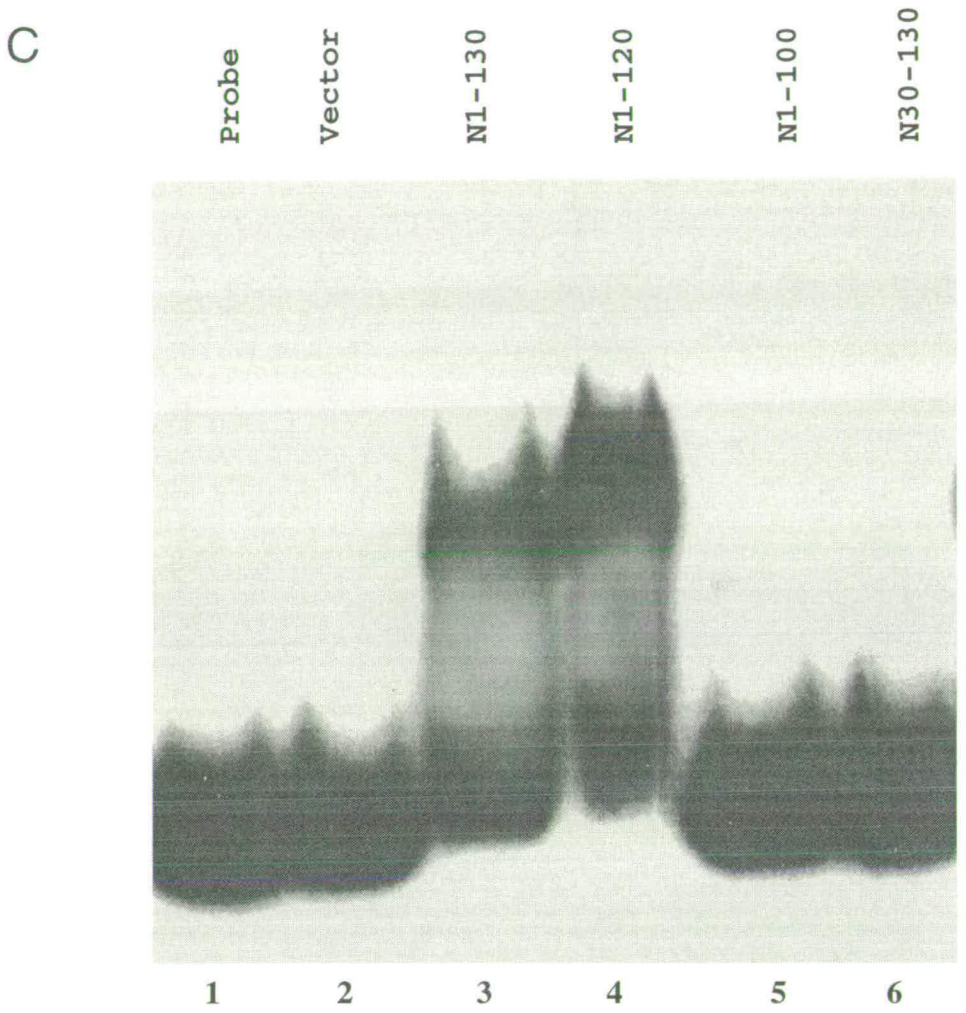


Figure 4.5 Expression and DNA binding of deleted Mos1-N150 molecules. (A) Schematic summary of deletions of Mos1-N150. Amino acids are numbered from the N-terminus to the C-terminus. DNA binding activity is denoted by +, and absence of binding by - symbols. **(B)** SDS-PAGE analysis of expressed His-tag fusion protein. The soluble crude extracts obtained after sonication were separated by SDS polyacrylamide gel electrophoresis and stained with commassie blue. Numbers of each lane correspond to the deletion numbers in **(A)**. **(C)** The DNA binding domain of *mariner* transposase is located within N-terminal 120 residues. *E.coli* extracts containing Mos1-N150 derivatives were incubated in the presence of 100 ng poly dI-dC with ^{32}P-labelled right-end sequence of *mariner*. The crude cell extracts used (0.1µl of each) contain Mos1-N1-130 (lane 3); Mos1-N1-120 (lane 4); Mos1-N1-100 (lane 5); or Mos1-N30-130 (lane 6). Lane 2 contains 0.1 µl of a control bacterial extract with no Mos1 derivatives and no protein was added in lane 1.

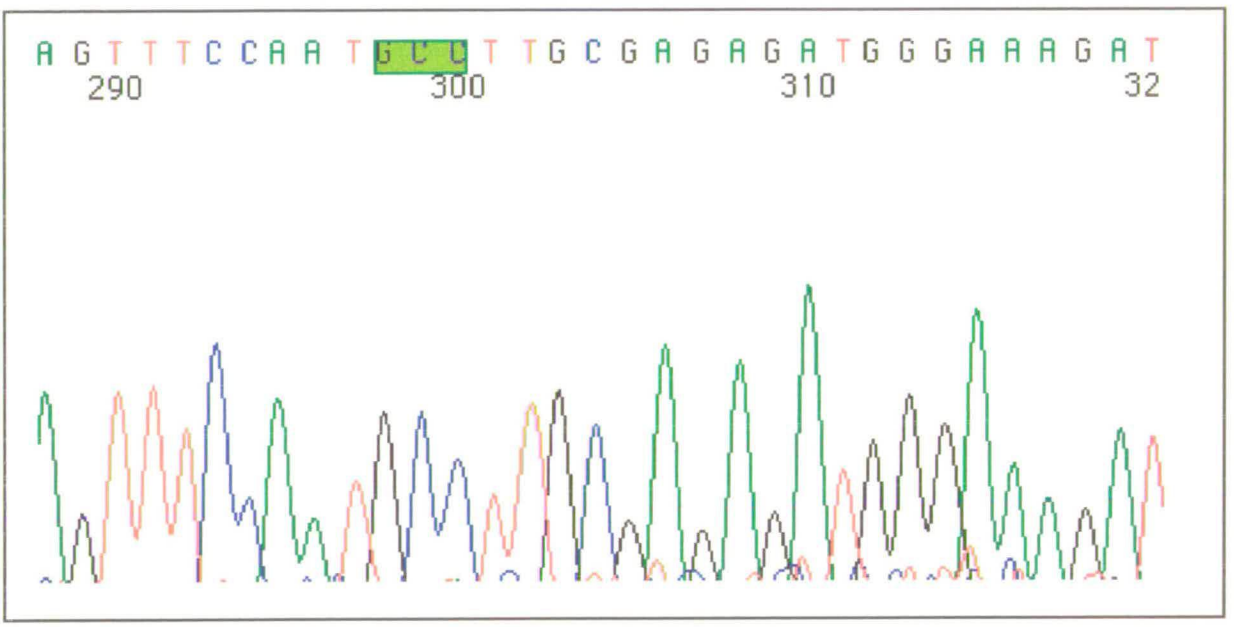


Figure 4.6 Sequence analysis of R106A mutant generated by site-directed mutagenesis. The arginine in position 106 of the Mos1-N150 was mutated to alanine by changing the codon from CGC to GCC. The complementary oligo pairs used to introduce the mutation (underlined) were as follow:

R106(+): 5'-GTT TCC AAT GCC TTG CGA GAA ATG-3'

R106(-): 5'-CAT CTC TCG CAA GCC ATT GGA AAC-3'

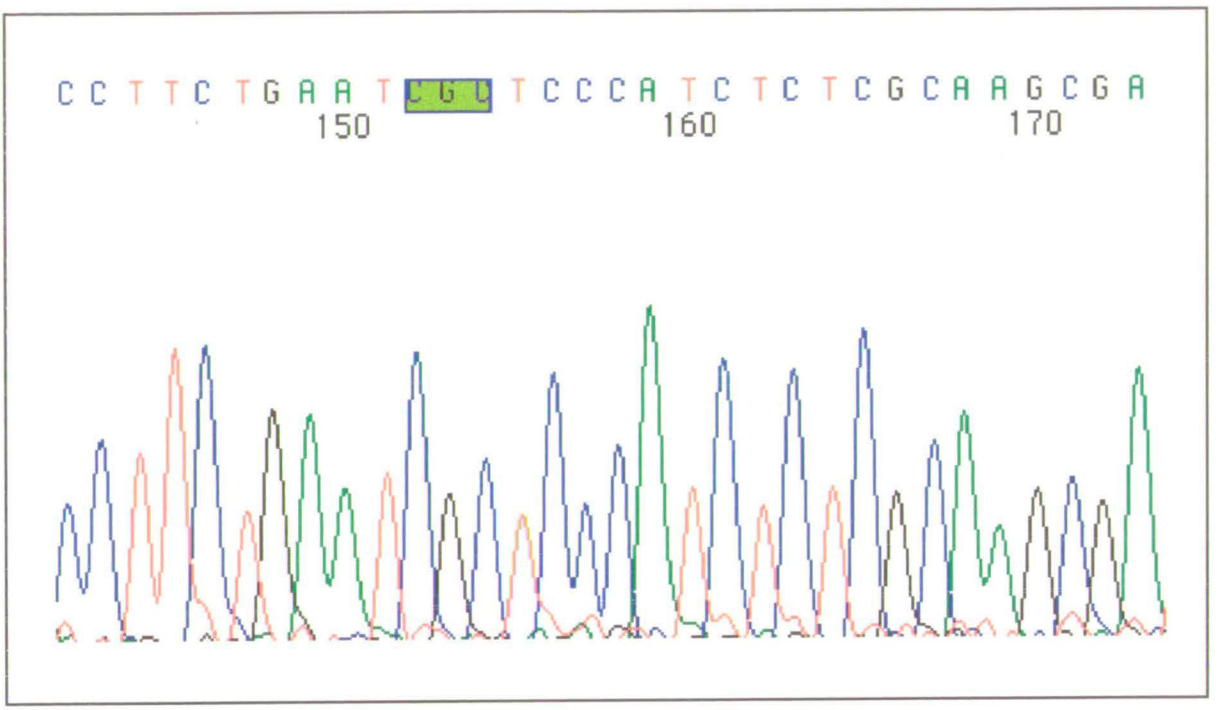


Figure 4.7 Sequence analysis of K112A mutant generated by site-directed mutagenesis. The lysine in position 112 of the Mos1-N150 was mutated to alanine by changing the codon from AAG to GCG. The complementary oligo pairs used to introduce the mutation (underlined) were as follow:

K112(+): 5'- GAG ATG GGA GCG ATT CAG AAG GTC-3'

K112(-): 5'- GAC CTT CTG AAT CGC TCC CAT CTC - 3'

The minus strand sequence was shown here.

(K112A) had no detectable effect on binding (Fig 4.8B).These results supported the prediction that HTH motif is responsible for sequence-specific binding.

One mutation S104P was unexpectedly recovered by PCR under standard condition. Sequencing data showed that the codon TCC which coded for S104 had been changed to the proline codon CCC (Fig. 4.9A). Introduction of proline into the second helix would be expected to disrupt the structure of the HTH motif itself. The DNA binding activity of Mos1-N150 was eliminated by the S104P mutation (Fig 4.9B). So the HTH structure itself contributes to the specific DNA-protein interactions.

4.2.4 The different affinity of Mos1-N150 for the two inverted repeats of *Mos1*

There are four mismatches between the two terminal inverted repeats of the *Mos1* element, including a difference in the first and last positions of the element. To check if these differences are critical for transposase binding, we performed competition experiments with either the left end or the right end as probe and various amounts of unlabelled left-end or right-end sequence as competitor assayed by gel retardation. Figure 4.10A shows that when the right-end probe was used, 35 ng (lane 4) of the right-end sequence had the same effect on the transposase-DNA complex as did 180 ng (between lanes 10 and 11) of the left-end sequence. Seventy nanograms (lane 5) of the right-end DNA was more effective as competitor than 337 ng (lane 12) of left-end DNA. Similar results were obtained when the left-end probed was used (Fig 4.10B). Twenty nanograms (lane 4) of the right-end sequence had the same effect as 100 ng (Lane 10) of the left end, while 50 ng (lane 5) of the right-end DNA and 300 ng (lane 12) of the left end were equally good as competitors. Taken together, these data indicated that Mos1-N150 had a 5 to 6 times higher affinity for the right-end inverted repeat of the Mos1 element than for the left end, implying that the left and right inverted repeats are not equivalent.

4.3 Discussion

Several classes of DNA-binding domains have been identified, which differ in their tertiary structures as well as in their way of interacting with DNA. These classes are referred to as helix-turn-helix motifs, helix-loop-helix motifs, zinc fingers, leucine zippers and ß-ribbons

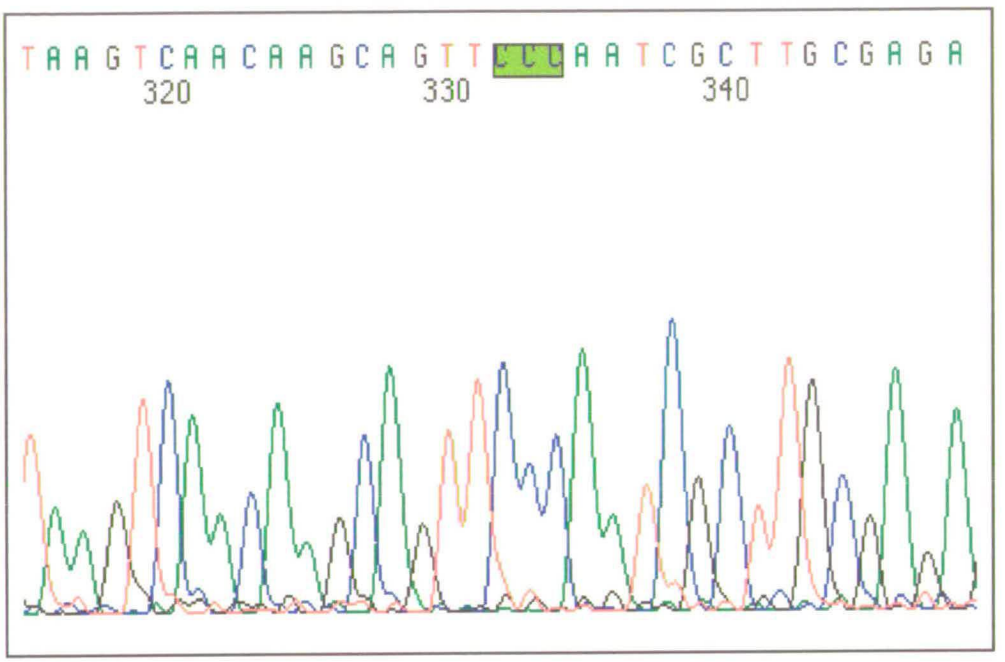
A

```
      HHHHHHHHttttHHHHHHHHHH
87 QTQKQLAEQLEVSQQAVSNRLREMGKIQKV 116
                  *       *
```

B

Figure 4.8 The HTH motif is responsible for specific DNA-binding activity of *mariner* transposase. **(A)** The N-terminus of *mariner* transposase contains an HTH motif with "H" corresponding to alpha helix and "t" to turn. The residues that have been mutated to alanine are underlined. **(B)** Gel retardation assay of mutated transposase. Purified Mos1-N120 (lane 2); Mos1-N150 [R106A] (lane 3); Mos1-N150 [K112A] (lane 4) proteins were tested for their specific binding to right-end probe. Lane 1 shows the free probe.

A



B

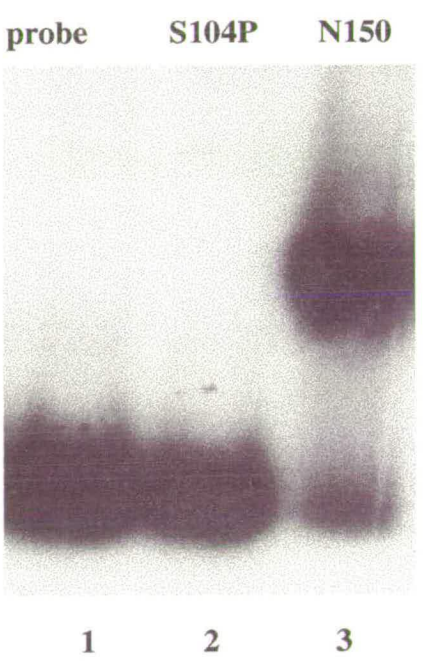
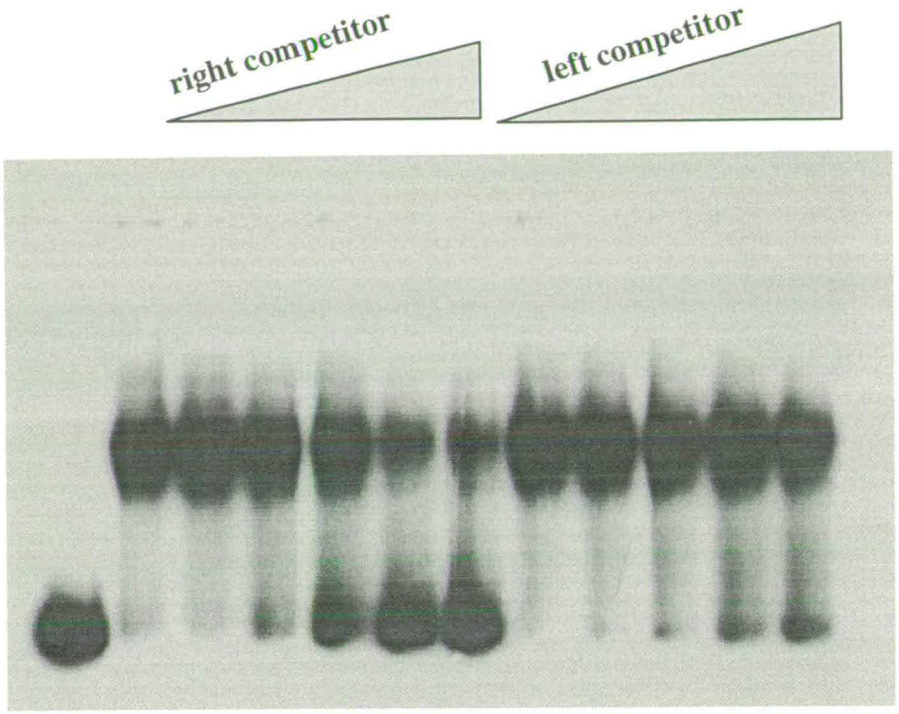


Figure 4.9 The DNA binding activity of S104P mutated Mos1-N150 protein.

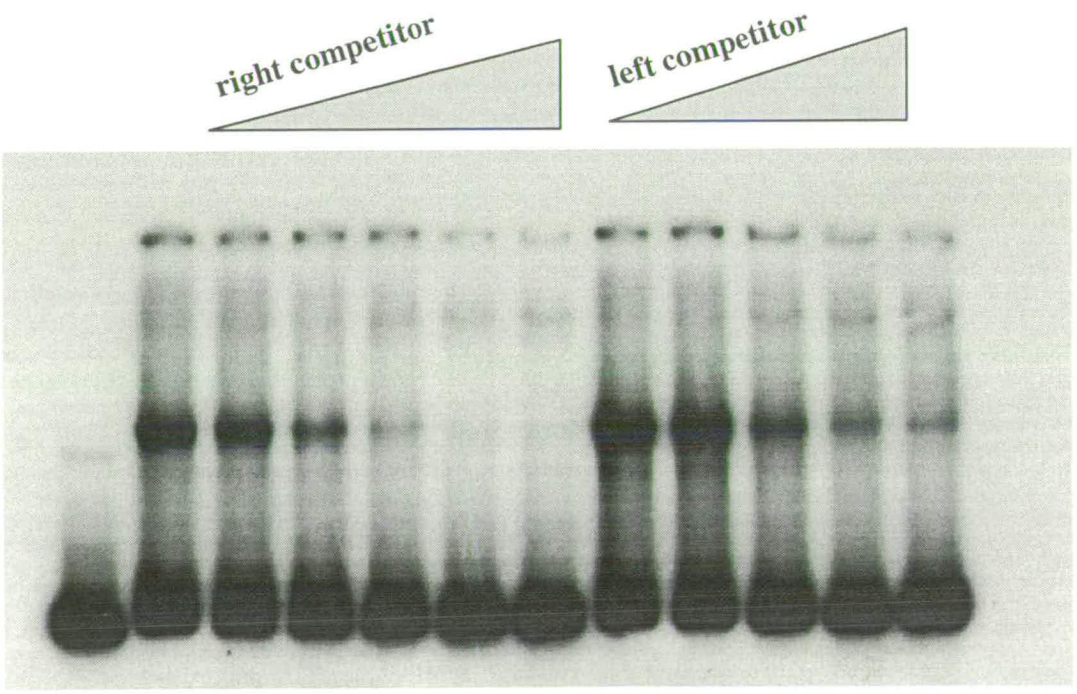
(A) Sequence analysis of S104P mutant generated by standard PCR. The serine in position 104 of Mos1-N150 was mutated to proline by changing the codon from TCC to CCC. **(B)** The DNA binding activity of Mos1-N150[S104P] (lane 2) was analysed by gel retardation assay and compared with wild-type Mos1-N150 (lane 3). Lane 1 shows the free probe.

A



Right-end probe

B



Left-end probe

Figure 4.10 Different affinity of Mos1-N150 for the left and the right end of Mos1 element. **(A)** The 176 bp right-end probe was incubated with 10 ng of purified Mos1-N150 fusion protein. Lane 1, probe alone; lanes 2-12, probe plus Mos1-N150; lanes 3-7 contain 17.5 ng, 35 ng, 70 ng, 105 ng and 187.5 ng of unlabelled right-end sequence respectively; lanes 8-12 contain 46 ng, 92 ng, 138 ng, 224 ng and 337 ng of the left-end sequence of *Mos1* as competitor. **(B)** The 190 bp left-end probe was used in gel retardation assay. Lane 1, probe alone; lanes 2-12, probe plus Mos1-N150; lanes 3-7 contain 10 ng, 20 ng, 50 ng, 100 ng, and 200 ng of right-end sequence as competitor; lanes 8-12 contain 20 ng, 50 ng, 100 ng, 200 ng and 300 ng of unlabelled left-end DNA fragment.

(Harrison, 1991). The best characterised of these classes is the helix-turn-helix (HTH) class. In its simplest form, the HTH motif consists of nearly perpendicular α-helices connected by a short linker (turn) of three amino acids. The second helix in the HTH motif is the "recognition" helix and penetrates into the major groove of the DNA target, allowing protein side-chains to make extensive base-specific hydrogen bonds and contacts (Brennan and Matthews, 1989). More recently, a systematic comparison has shown that all HTH-containing DNA binding proteins actually consist of at least three α-helices, the two helices of the HTH motif and a third α-helix that stabilised this unit (Suzuki and Brenner, 1995; Wintjens and Rooman, 1996).

There is little sequence conservation in the N-terminal regions of *Tc1/mariner* transposases and no apparent conservation of the inverted repeat sequence among elements in *Tc1/mariner* family. Despite this, computer analysis suggests that several members of the *Tc1/mariner* family contain a helix-turn-helix motif near their N-termini which may be responsible for sequence specific binding (Pietrokovovski and Henikoff, 1997). The crystal structure of the DNA-binding domain of *Tc3* transposase has shown that the HTH motif makes four base-specific contacts with the major groove (van Pouderoyen et al., 1997). For Mos1 protein, the HTH motif was predicted to run from residues 87 to 108. Mutations that changed basic residues within the second helix of the HTH of Mos1-N150 greatly reduced DNA binding indicating that the HTH is required for sequence-specific recognition of the terminal inverted repeats of *Mos1*. However, the HTH motif is necessary but not sufficient to obtain an efficient DNA-binding domain. Various C- or N-terminal deletions of the Mos1-N150 showed that the first 30 amino acids of Mos1-N150 are also necessary for its binding activity. These 30 amino acids might be directly involved in DNA binding or might be important for the correct folding of the region containing the HTH motif.

Few studies have compared the affinity of the transposase for the two inverted repeats of transposable elements. One study showed that the prokaryotic Tn522 transposase has the same affinity for both termini, the nucleotide sequences of which diverge by 20% (Rowland et al., 1995). There are four mismatches between two inverted repeats of *Mos1* element. The results described in section 4.2.4 demonstrate that the Mos1 transposase has a 5 to 6 times higher affinity for the right-end sequence than for the left end. The differential binding pattern of Mos1 transposase on its two ends may affect *Mos1* transposition,

creating a hierarchy in the first step of the excision or affecting a regulation mechanism. Differential transposition activity between the two ends has been observed with a number of transposable elements including *IS50* (Phadnis and Berg, 1987), *IS10* (Morisato and Kleckner, 1984) and *IS903* (Derbyshire et al., 1987). The two ends of *P* element are not equivalent. The 5' end can not function as a 3' end and induces aberrant transposition events. Thus, *P* element transposition must occur in an asymmetric fashion with respect to the functions of the two ends (Mullins et al., 1989). We predict that *Mos1* may also show a nonequivalence in the functions of its two ends.

Chapter 5

Subunit interactions of the Mos1 transposase

5.1 Introduction

All transposition reactions examined thus far, involve assembly of a stable protein-DNA transpososome, in which transposase -transposase interactions bridge transposon ends. Although the cleavage and strand transfer reactions of transposition appear to be catalysed within this multimeric, nucleoprotein complex, the active multimeric state of most transposase and integrases is not well understood. Several stable nucleoprotein complexes (transpososomes) that are intermediates in *Mu* transposition have been characterised. Each transpososome minimally contains a tetramer of transposase subunits bound to the two Mu DNA ends (Chaconas et al., 1996; Williams et al., 1999). Within the tetramer, the two subunits (MuA) bound to L1 and R1 sites at both ends of the *Mu* genome, donate all three active site DDE residues for 3' end cleavage and strand transfer of the partner *Mu* DNA end (catalysis in trans). In *Tn5* transposition, transposase binds to the transposon DNA at the end recognition sequences. Then, the end sequences are brought together via transposase oligomerization to form a synaptic complex (Reznikoff, 1993). Tn5 transposase is 476 amino acids in length and contains two distinct dimerization domains, one of which is located at the COOH terminus and the other near amino acids 114-314 (Braam and Reznikoff, 1998). The transposable element *En/Spm*-encoded TNPA protein consists of 621amino acids and contains a dimerization domain at the C-terminal, from residue 428 to 542. Dimers seems to occur in the free as well as in the DNA-bound state of TNPA (Trentmann et al., 1993). The regions in TNPA or Tn5 transposase that are required for dimer formation do not fall into any of the known classes of dimerization domains such as the leucine zipper.

Computer analysis (ProfileScan program) has failed to detect any known dimerization domains in Mos1 transposase. Two types of evidence indicate that Mos1 transposase functions as an oligomeric molecule. The first is the dominant-negative complementation of certain of the transposase mutants (Lohe et al., 1997). The second evidence emerged from the yeast two-hybrid system, which yields activation of the reporter gene when the GAL4 DNA-binding domain and the GAl4 activation domain were both fused with Mos1 transposase, implying that Mos1 subunits can interact physically (Lohe et al., 1996).

The yeast two-hybrid system is widely used to investigate protein-protein interaction studies (Fields and Song, 1989; White, 1996; Vidal and Legrain, 1999). Generally, it

requires the construction of hybrid genes to encode: (1) a DNA-binding domain fused to a protein "X", and (2) an activation domain fused to a protein "Y" (Fig 5.1). The DNA-binding domain targets the hybrid protein to its binding site, but, because most proteins lack an activation domain, this DNA-binding hybrid protein does not activate transcription of the reporter gene. Likewise, the hybrid protein that contains the activation domain does not activate expression of the reporter gene because it cannot bind the upstream activation sequence (UAS) in GAL4 system or LexA operator in LexA system. If X and Y can form a protein-protein complex and reconstitute proximity of the activation domain to the binding sites, transcription of a reporter gene occurs. The domains most commonly used are the DNA-binding domains of GAL4 and LexA, and the activation domain of GAL4. In this study, both the LexA DNA-binding domain and the GAL4 activation domain were fused to Mos1 protein.

In the experiments described in this chapter, the yeast two-hybrid system was used to confirm that Mos1 transposase monomers interact, then truncated polypeptides were made to try to locate the dimerization domain involved in this interaction. Deletion mutants suggest that both N- and C-termini are essential for Mos1 dimerization. Twelve single point mutants affecting protein-protein interactions have been isolated.

5.2 Results

5.2.1 Yeast two-hybrid system identified Mos1 subunit interactions

The Mos1 transposase was amplified by PCR using *pMos* as template, primer Y1(+) (CATG*CCATGG*ACATGTCGAGTTTCG TGCCG) encoding the first 6 N-terminal amino acids with a *Nco*I restriction site and primer Y345(-) (CCG*CTCGAG*TTATTCAAAGTAT TTGCCGTCG) encoding the last 7 C-terminal amino acids with a *Xho* I restriction site. The PCR product was gel purified before "TA" cloning into pGEM-T vector. The resulting plasmid pGEM-T-*Mos1* was digested with *Nco*I and *Xho*I, gel purified and ligated to *Nco*I/*Xho*I cut pACTIIst (containing the GAL4 activation domain and yeast *LEU2* gene for selection). Meanwhile, the *Nco*I-*Xho*I fragment containing Mos1 ORF was ligated to pAS-BC restricted with *Nco*I and *Sal* I (*Sal*I and *Xho*I have compatible ends). The *Eco*RI/ *Pst*I fragment digested from pAS-BC-Mos1 was ligated into pBTM116 (containing the

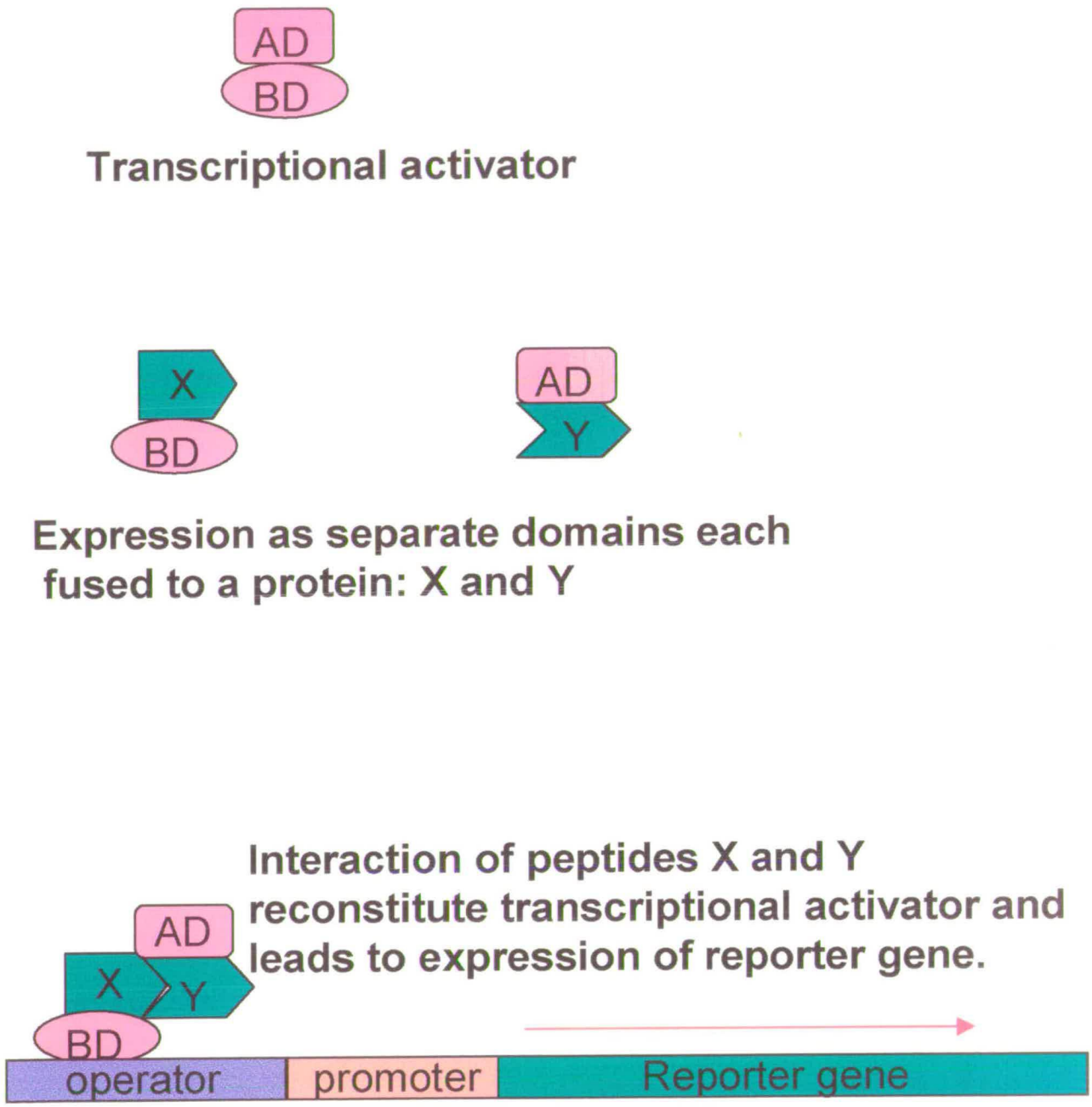


Figure 5.1 The yeast two-hybrid system to study protein-protein interactions

A hybrid protein is generated that includes a DNA-binding domain and a protein X, another hybrid protein is generated that fuses an activation domain to a protein Y. Both hybrid proteins are produced in the same transformant. The X and Y proteins bind noncovalently and activate transcription.

LexA DNA-binding domain and yeast *TRP1* gene as a selectable marker) (Fig 5.2). Both constructs were sequenced to confirm the fidelity of the wild type *Mos1*.

Yeast strain L40 with a chromosomal copy of the *E.coli lacZ* reporter gene preceded by four copies of the LexA binding site and the *HIS3* reporter gene downstream of eight copies of this sequence was used to test for interaction between Mos1 transposase monomers. These cells were co-transformed with two plasmids by the lithium acetate method and transformants were selected on medium lacking Tryptophan and Leucine (-LW). To test for expression of the reporter genes, cells were patched onto medium lacking Tryptophan, Leucine and Histidine (-LWH), containing 0, 0.5, 2, 10 mM 3-amino-1,2,4-triazole (3-AT; Sigma). *HIS3* encodes an enzymatic activity specifically inactivated by the competitive inhibitor 3-AT. The amount of 3-AT required to prevent growth by inhibition of His3 protein gave a quantitative measure of the interaction. X-gal overlay assay was performed to check expression of *lacZ* . After incubation at 30°C for 30 min, positive colonies turned blue.

The activation of the reporter gene was specific for the presence of Mos1 protein fused to both the DNA-binding domain and the activation domain. The activity was not induced by either two "empty plasmids" or one "empty plasmid" with one Mos1-containing vector (Fig 5.3). When *HIS3* was the reporter gene, its expression in the absence of an activator was modulated by 0.5 mM 3-AT. The maximum 3-AT concentration tolerated by yeast cells containing Mos1-Mos1 interaction was 20mM. Consequently, wild-type and mutant alleles of an interaction partner can be phenotypically discriminated on plates containing increasing concentrations of 3-AT (Fig 5.4).

5.2.2 Deletion mutants suggest that both N- and C- termini were essential for Mos1 dimerization

The amino acid sequence of Mos1 transposase does not contain any obvious protein-protein interaction motif, so various deletion derivatives of the protein were tested for their ability to interact with full-length transposase. Both N- and C-terminal deletions were engineered in the same way as full-length Mos1 described in 5.2.1. All forward primers contained *Nco*I sites and the reverse primers contained *Xho*I sites for ligation to two-hybrid

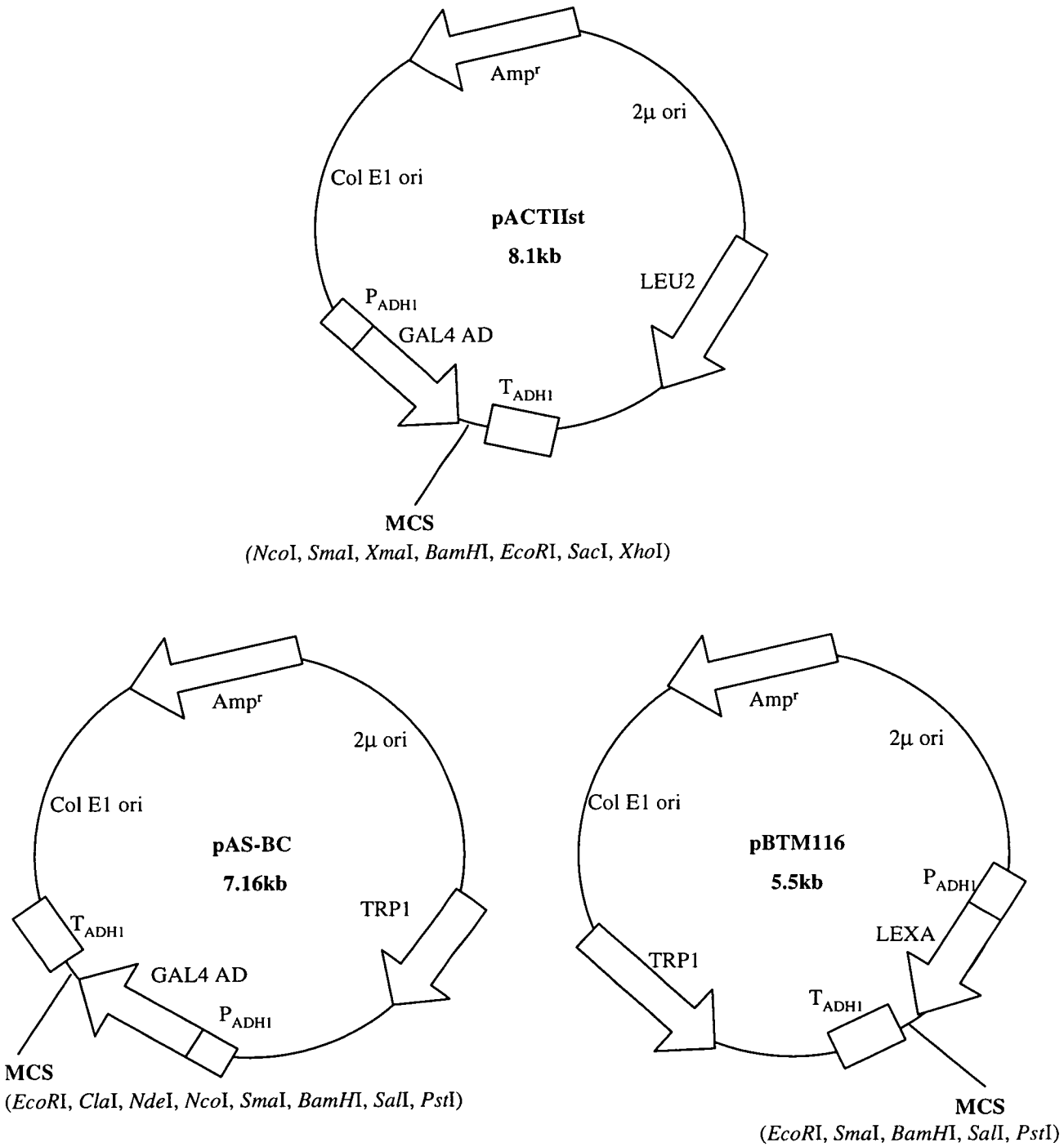
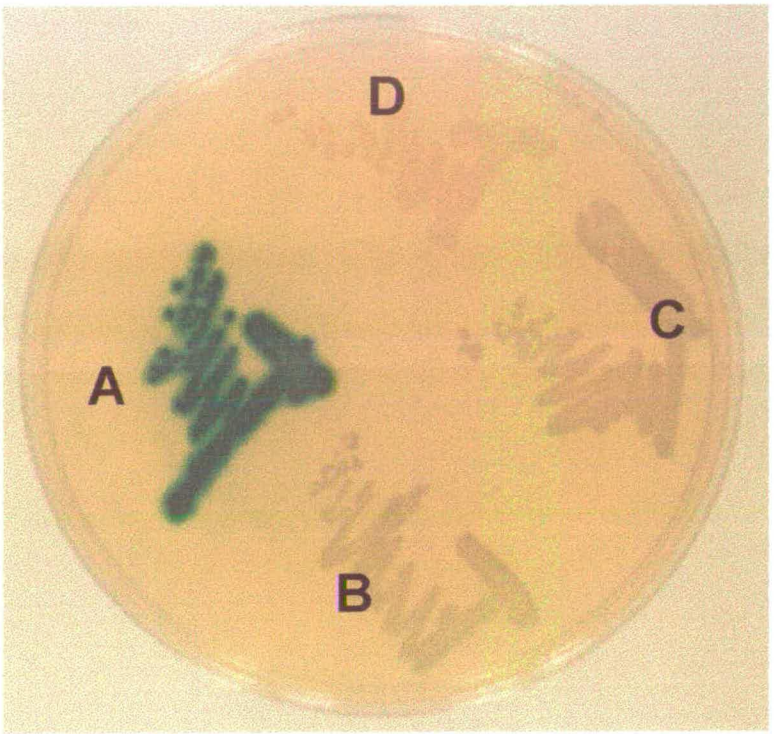


Figure 5.2 Maps of three plasmids for yeast two-hybrid system.

pACTIIst contains the GAL4 activating domain, pAS-BC contains the GAL4 DNA-binding domain and pBTM116 contains the complete LexA protein coding sequence (1-202). *Mos1* open reading frame were subcloned into the multiple cloning sites (MCS), resulting in pACTIIst-*Mos1*, pAS-BC-*Mos1* and pBTM116-*Mos1*.



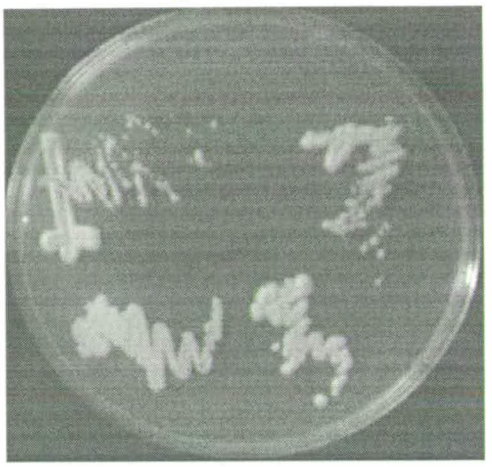
Section	Plasmid	Expression His3	Expression LacZ
A	pACTIIst-*Mos1* + pBTM116-*Mos1*	+	+
B	pACTIIst-*Mos1* + pBTM116	-	-
C	pACTIIst + pBTM116	-	-
D	pACTIIst + pBTM116-*Mos1*	-	-

Figure 5.3 Subunit interactions of Mos1 protein by two-hybrid system.

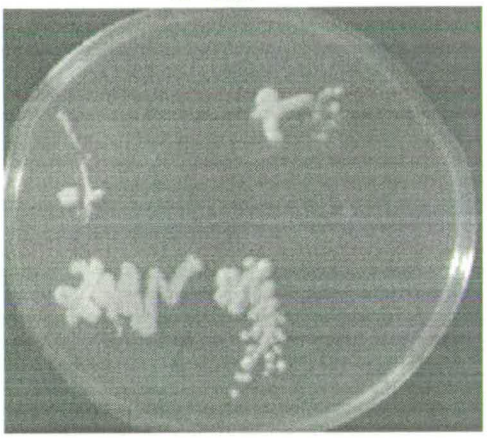
The activation of the reporter gene was specific for the presence of Mos1 protein fused to both the LexA DNA-binding domain and the Gal4 activation domain (Section A). The activation was not induced by either two "empty vector" (Section C) or one "empty vector" with one *Mos1*-containing vector (Sections B and D) .

0.5 mM

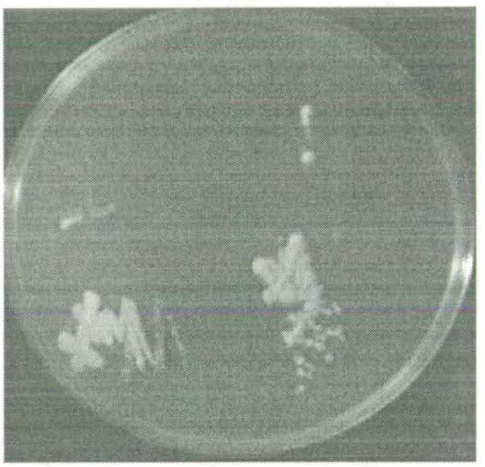
2 mM



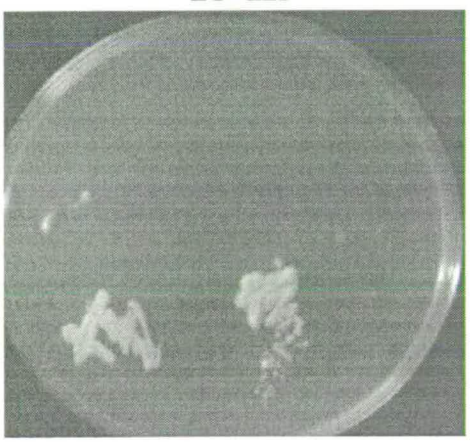
5 mM



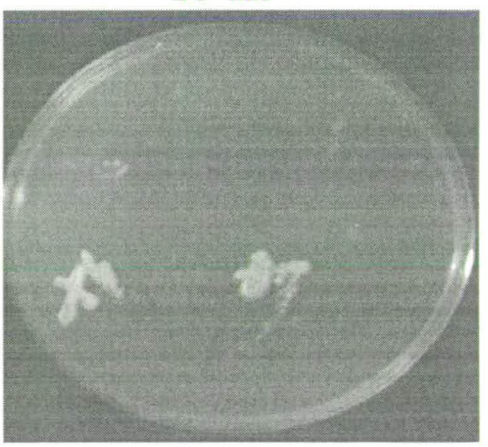
10 mM



15 mM



20 mM



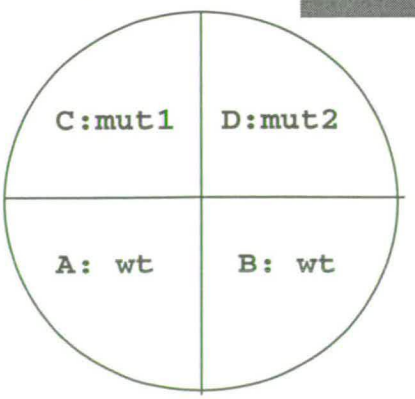


Figure 5.4 Interaction between two wild type Mos1 proteins or between mutant and wild type Mos1 protein: Assay for activation of the *HIS3* reporter gene.

The *S.cerevisiae* reporter strain L40 which contain pACTIIst-Mos1(wt) was transformed with pBTM116-Mos1(wt) (Sections A and B) or pBTM116-Mos1 (L124S) (Sections C) and pBTM116-Mos1(S302P) (Section D). Patches of Leu^+Trp^+ transformants were streaked to plates lacking tryptophan, histidine and leucine. The plates contained different concentrations of 3-AT between 0.5 mM and 20 mM to test for the ability to form colonies. The plates were photographed after incubation at 30°C for 3 days. The wild-type Mos1-Mos1 interaction induced His3 expression that allows growth at 3-AT concentration up to 20 mM, while Mos1 mutants allow growth at 3-AT level limited to 2 to 5 mM.

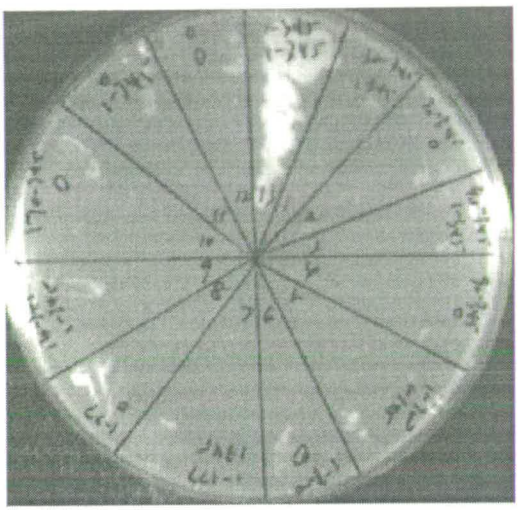
vectors. The PCR products using primers Y1(+) and Y177(-) encoded Mos1(1-177). Likewise, Y1(+)/Y300(-) encoded Mos1(1-300); Y30(+)/Y345(-) encoded Mos1(30-345); Y90(+)/Y300(-) encoded Mos1(90-345) and Y170(+) /Y345(-) encoded Mos1(170-345). Each of the deletions, residues 1-177, 1-300, 30-345, 90-345 and 170-345, was placed downstream of the LexA DNA binding domain in pBTM116 and introduced into L40 cells together with pACTIIst-*Mos1* but in no case was the expression of *lacZ* or *HIS3* above background (Fig 5.5). This was not due to lack of expression of the deleted transposase fusions as this was checked using antibodies to the LexA DNA binding domain. These results suggest that residues required for the interaction of transposase monomers are distributed along the length of the protein.

5.2.3 Single point mutations in Mos1 protein causing defective interactions

Since deletion mutants made in 5.2.2 might severely impair the three dimensional structure of the protein. The method described in this section was to isolate single point mutations that disrupt Mos1-Mos1 interactions. Large libraries of randomly generated mutant alleles of Mos1 protein were screened for interaction with wild-type Mos1 (Fig 5.6).

Mutagenic PCRs were performed using three methods. (1) Low-fidelity *Taq* polymerase lacking proof reading activity was used with 40 cycles amplifications; (2) PCR was performed with 20 μM of dATP, while the other dNTPs (dATP, dCTP and dGTP) were at a normal concentration (200μM); (3) $MnCl_2$ was added in reaction buffer at a final concentration of 0.5mM. All of the above three methods contained 100ng of pBTM116-Mos1(wt) as template, using pBTM(+) (50 bp upstream of cloning site of Mos1 in pBTM116 vector) as forward primer and pBTM(-) (50 bp downstream of cloning site of Mos1 in pBTM116 vector) as reverse primer. The amplified PCR products (containing mutated Mos1) were subsequently introduced into pACTIIst-Mos1-containing yeast cells (L40) with the linerized pBTM116 backbone digested with *Eco*RI/*Pst*I by gap-repair transformation.

The transformants were plated on medium lacking Leu, Trp and His to select for transformants in which pBTM116 had incorporated a molecule of amplified DNA by gap repair recombination. Transformants were streaked onto -LWH plates containing 0.5, 5, 10 mM 3-AT to test for their ability to form colonies. About 10% of the transformants displayed a growth defect at 5 mM 3-AT indicative of a defective interaction. The corresponding pBTM116-*Mos1*(mut) plasmids were isolated and amplified in *E.coli* using



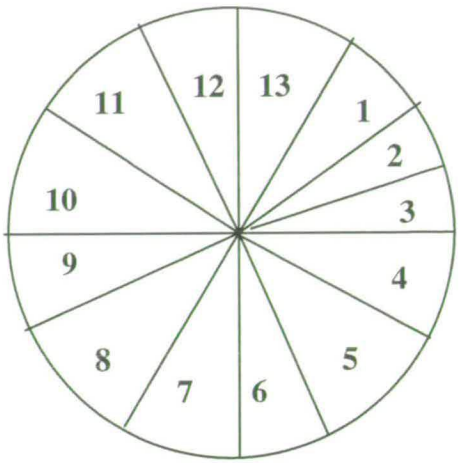


Figure 5.5 Absence of expression of the *HIS3* reporter gene in deletion mutants.

Yeast strain L40 was co-transformed with the following pairs of plasmids:

Section 1 : pBTM116-Mos1(30-345) + pACTIIst-Mos1(1-345)

Section 2 : pBTM116-Mos1(30-345) + pACTIIst

Section 3 : pBTM116-Mos1(90-345) + pACTIIst-Mos1(1-345)

Section 4 : pBTM116-Mos1(90-345) + pACTIIst

Section 5 : pBTM116-Mos1(1-300) + pACTIIst-Mos1(1-345)

Section 6 : pBTM116-Mos1(1-300) + pACTIIst

Section 7 : pBTM116-Mos1(1-177) + pACTIIst-Mos1(1-345)

Section 8 : pBTM116-Mos1(1-177) + pACTIIst

Section 9 : pBTM116-Mos1(170-345) + pACTIIst-Mos1(1-345)

Section 10 : pBTM116-Mos1(170-345) + pACTIIst

Section 11 : pBTM116 + pACTIIst-Mos1(1-345)

Section 12 : pBTM116 + pACTIIst

Section 13 : pBTM116-Mos1(1-345) + pACTIIst-Mos1(1-345)

Co-transformants were restreaked to -LWH + 0.5mM 3-AT plate to test for activation of the *HIS3* reporter gene. Only Section 13 contained interacting hybrid proteins.

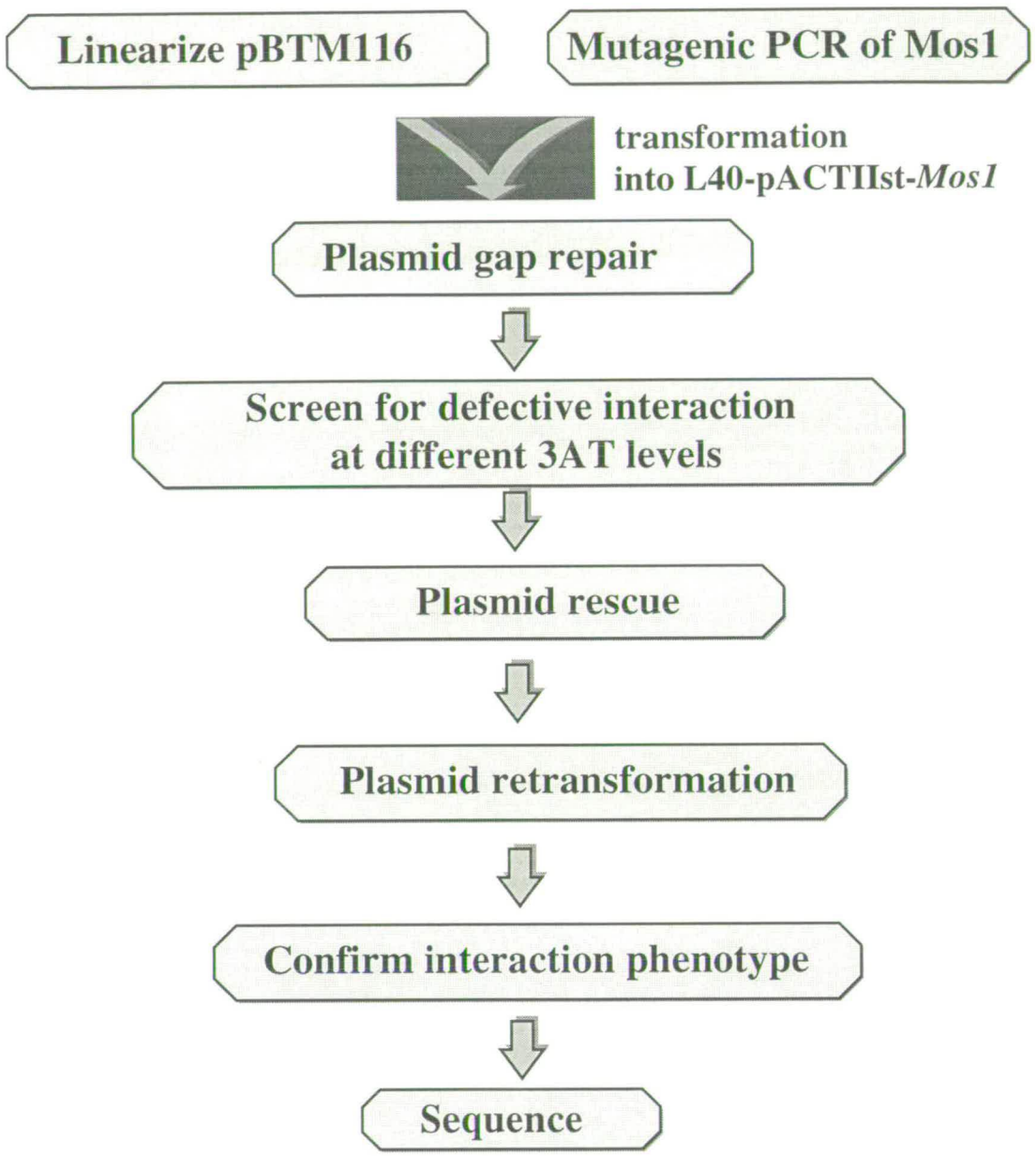


Figure 5.6 Reverse Two-hybrid system.

The "reverse" two-hybrid system was designed to identify events that dissociate protein-protein interactions. Transformants displayed a growth defect at 5 mM 3-AT indicating defective interactions. The rescued plasmids were retransformed into pACTIIst-Mos1-containing yeast cell to confirm the identical growth phenotypes.

the yeast plasmid rescue method. The rescued plasmids were reintroduced into pACTIIst-Mos1-containing yeast cells to confirm the identical phenotypes. Most of them were false positive and conferred wild-type phenotype. Only a small fraction of primary candidates proved to be real mutants (Fig 5.7).

Sequencing of the full-length Mos1 mutants revealed that mutagenic PCR(1) favoured single miscorporation, while mutagenic PCR(2) and PCR(3) usually produced mutants with more than one amino acid change (Tab 5.1). 12 single point mutations affecting protein-protein interactions have been isolated among 4000 transformants (Fig 5.8). Their growth phenotypes were summarised and compared with wild type Mos1 in Fig 5.9. The single amino acid changes affecting the interaction of tranposase monomers are distributed along the length of the protein but are concentrated in the N terminal region containing the putative HTH motif required for DNA binding.

The level of LexA-transposase fusion protein expressed from each the mutant plasmids was investigated as the reduced interaction seen in the two-hybrid assay could be due to instability of a mutant protein rather than a reduced ability to interact with a wild type monomer. This was done by Western blots of extracts of yeast cells carrying the appropriate plasmid using antibodies to LexA protein. The results indicate that each of the single point mutant proteins but one (D279G) was present at levels equivalent to that of the wild type fusion (Fig 5.10). The Mos1(S28P, V202A) double mutant exhibited strongest phenotype and could not grow on 0.5 mM 3-AT, while the Mos1(V202A) single-point mutant has no effect on interaction and could grow on 10 mM 3-AT. The severely impaired interaction of Mos1(S28P, V202A) with wild-type Mos1 may partially due to its low expression in yeast (Fig 5.10A, Lane2) and partially due to the S28P amino acid substitution.

There are 5 cysteine residues in Mos1 transposase including Cys336, ten amino acids upstream of the stop codon in Mos1 ORF. To investigate if this cysteine residue was involved in formation of Mos1 homodimers via disulfide bonds, site-directed mutation of C336A was generated by PCR using primer C336(-) with mutated bases. Plasmid pBTM116-*Mos1*(C336A) was transformed into pACTIIst-*Mos1*-containing yeast cells (L40) to test for interaction between wild-type Mos1 and Mos1(C336A). Changing cysteine 336 to alanine did not affect interaction of Mos1 subunit (Fig 5.11) indicating Cys336 of Mos1 protein is not responsible for such interactions.

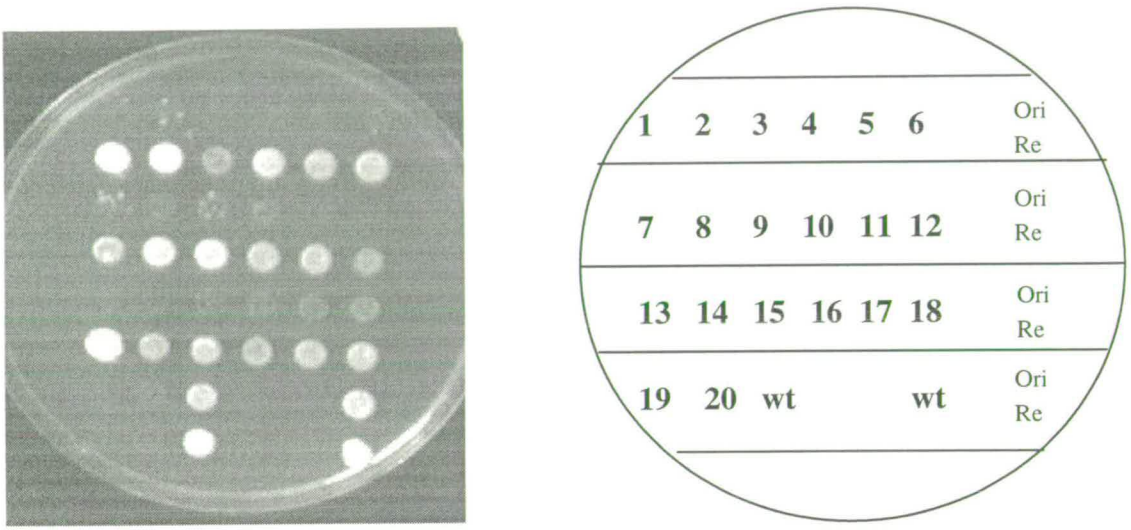


Figure 5.7 The comparison of growth phenotype between original and retransformed candidates.

The numbers are refer to different Mos1 (mut) alleles. Patches of original isolated (Ori) or retransformed (Re) cells were streaked onto -LWH plates containing 5mM 3-AT to test for the ability to form colonies. pACTIIst-Mos1(wt)-containing L40 was transformed with pBTM116-Mos1(wt) and shown as as positive control. The plates were photographed after incubation at 30°C for 3 days. Note that only No.19 and No.20 revealed identical phenotypes after originally isolated plasmids were reintroduced into the yeast cells in this growth assay. Other retransformants grew as normally as wild-type.

Table 5.1 Mutations obtained from random mutagenesis

Mutant Name	interactions *	Mutation(or mutations)
Taq(0)1	10mM	E9G
Taq(0)8	0.5mM	I160T
Taq(0)11	0.5mM	S104P
Taq(0)15	0.5mM	M192I, H273R, E312A
dATP9	0.5mM	S2P, F50L, M110V, N164S
dATP11	10mM	S2P, F50L
dATP12	10mM	K22E, S305G
dATP13	0.5mM	V116A, L149S, D279G
dATP14	10 mM	V202A
dATP15	0 mM	S28P, V202A
dATP17	0.5mM	V60A, D249G,N250S, T255A, H293R
$MnCl_2$2	0.5mM	M110V, R186H,T219C, E264D
$MnCl_2$20	10mM	K22R, H286Y
Taq(1)2	0.5mM	L92H
Taq(1)3	10mM	E123G, K133R
Taq(2)1	0.5mM	F53Y
Taq(2)5	0 mM	F36S, V43A, I113T
Taq(2)9	0.5mM	A93V
Taq(2)12	0.5mM	S302P
Taq(3)3	10mM	F344S
Taq(3)19	0.5mM	L124S
Taq(4)3	0.5mM	Q91R
Taq(4)5	0.5mM	D279G
Taq(4)8	0.5mM	W159R
Taq(4)12	0.5mM	V120A
Taq(4)13	0.5mM	Y237C

* As indicated by the ability to grow on different concentrations of 3-AT. Mutants couldn't grow at 5mM 3-AT indicating of defective interactions. Mutants with silent changes and chain termination codons are omitted from the table. Residues that appear more than once are underlined.

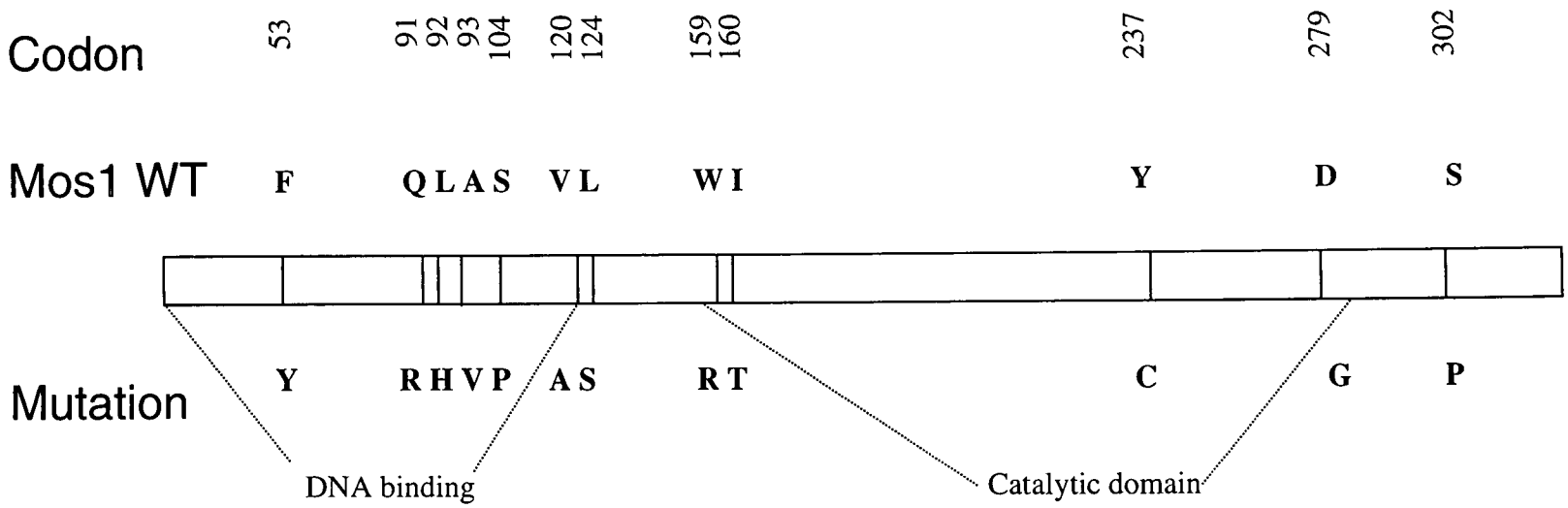
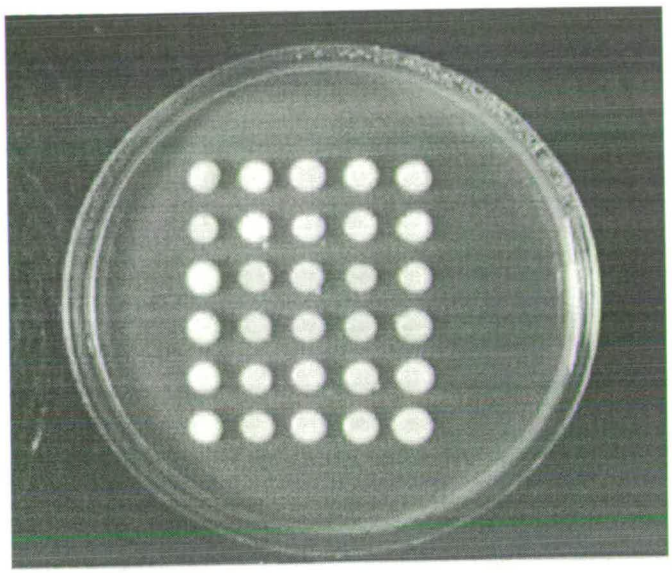
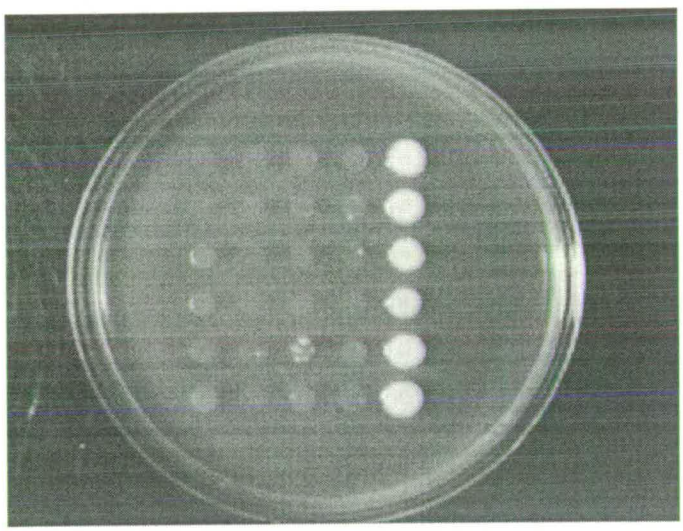


Figure 5.8 Single point mutations in Mos1 causing defective interactions.

The decreased level of interactions are listed in Table 1.



0.5mM 3-AT



5mM 3-AT

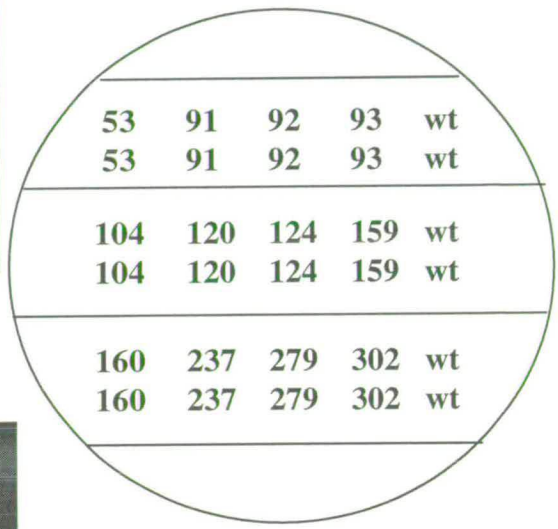
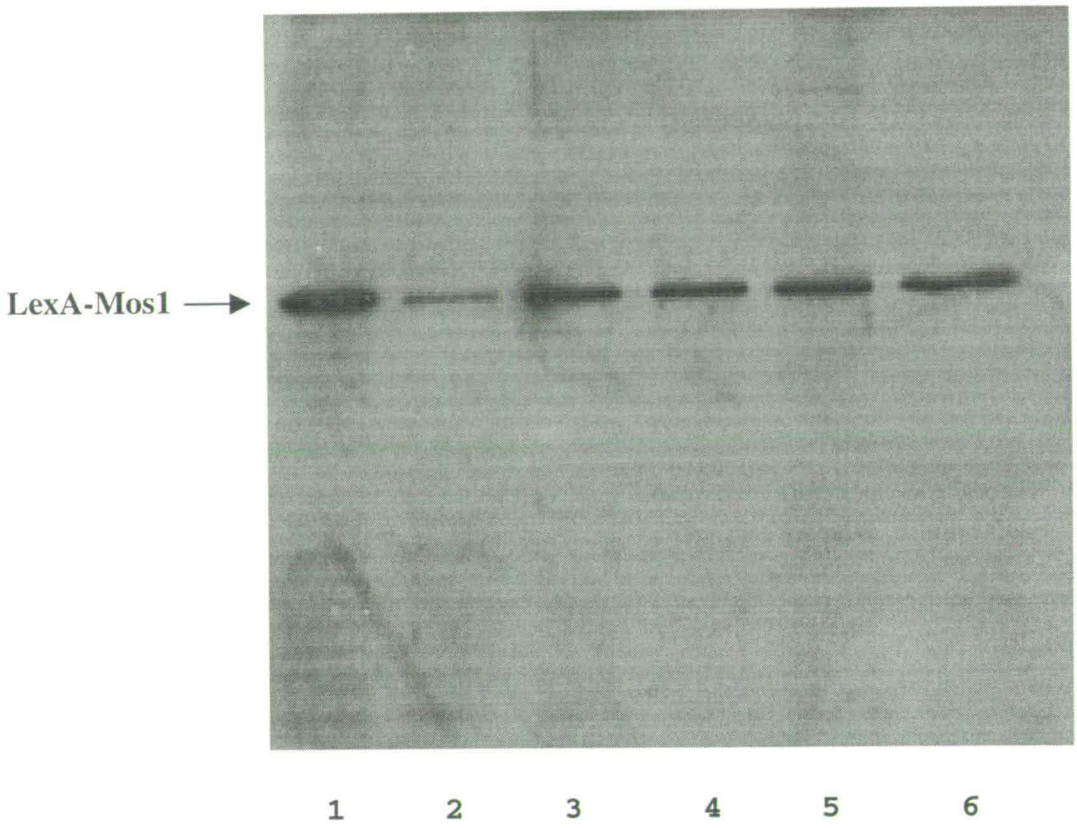


Figure 5.9 Growth phenotypes of 12 Mos1 mutant alleles.

The different Mos1 (mut) alleles are indicated with the codon number. Two individual Leu^+Trp^+ transformants were tested for each allele. Cells were streaked onto -LWH plates containing 0.5 mM or 5mM 3-AT to test for the ability to form colonies. pACTIIst-Mos1(wt)-containing L40 was transformed with pBTM116-Mos1(wt) and shown as positive control. The plates were photographed after incubation at 30°C for 3 days. Wild-type Mos1-Mos1 interaction induced His3 expression that allows growth at 3-AT concentration at 5 mM , while Mos1 mutants allow growth at 3-AT level limited to 0.5 mM.

A



B

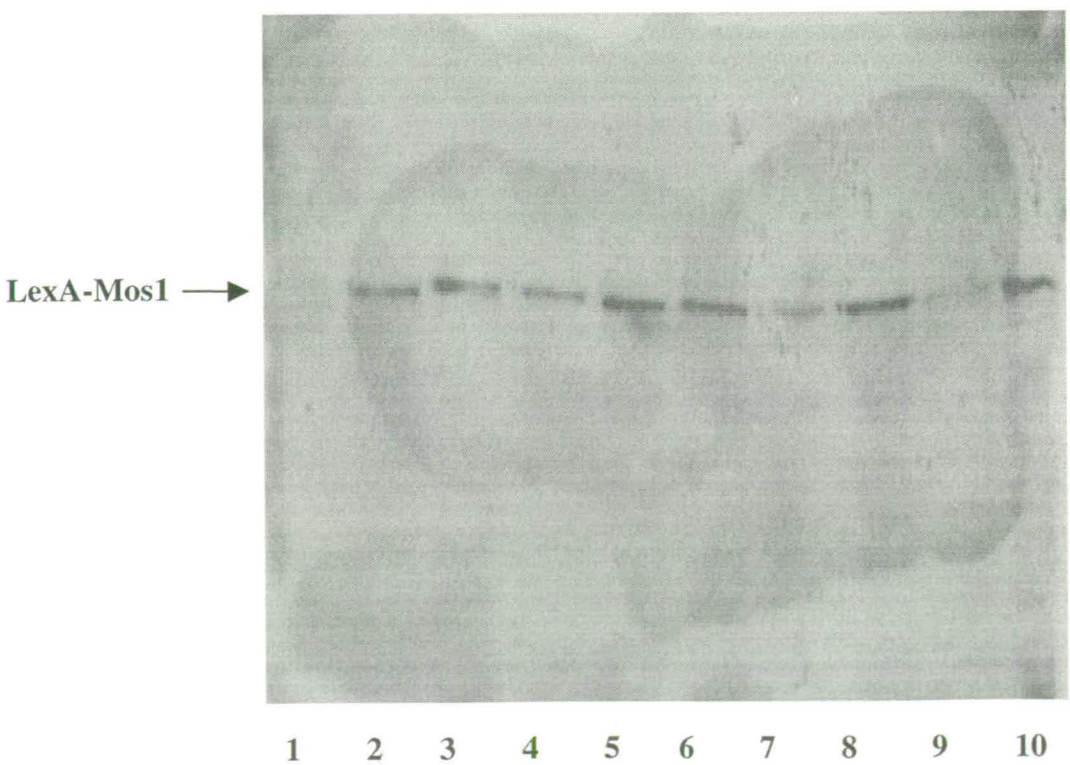
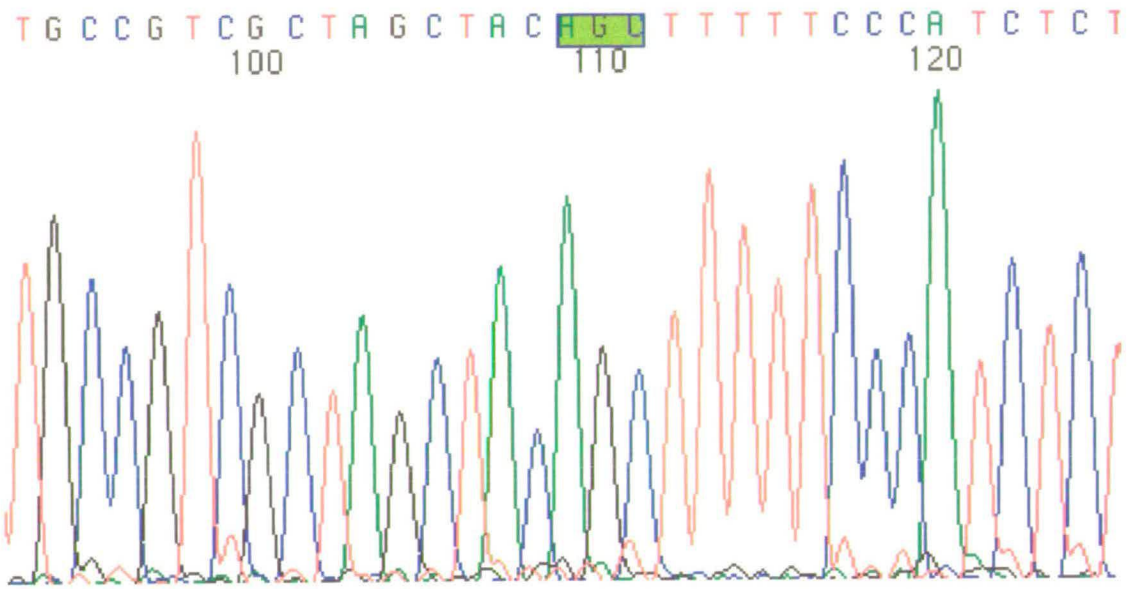


Figure 5.10 Expression levels of LexA-Mos1 mutant alleles.

Equivalent amounts of yeast extracts were run on a 10% SDS-PAGE gel and then electroblotted to PVDF membrane. The blot was probed with LexA mouse monoclonal IgG1 antibody (200ng/ml), followed by HRP-conjugated goat anti-mouse IgG(diluted 1:2000 to 200ng/ml). The signals were detected by chemiluminescence method. In most cases , the protein levels were comparable between samples. **(A)** Lane1: LexA-Mos1(wt); Lane2: LexA-Mos1(S28P, V202A); Lane3:LexA-Mos1(F53Y); Lane4:LexA-Mos1(A93V); Lane5: LexA-Mos1(L124S) and Lane6: LexA-Mos1(S302P). **(B)** Lane1:prestain marker; Lane2: LexA-Mos1(Q91R); Lane3:LexA-Mos1(L92H); Lane4: LexA-Mos1(S104P); Lane5: LexA-Mos1(V120A); Lane6: LexA-Mos1(W159R); Lane7: LexA-Mos1(I160T); Lane8: LexA-Mos1(Y237C); Lane9: LexA--Mos1(D279G); Lane10: LexA-Mos1(wt).

A



B

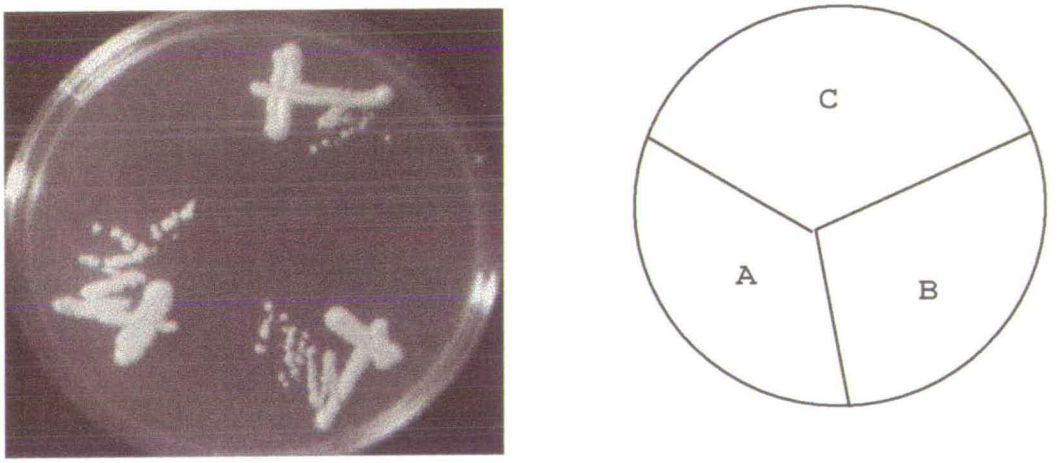


Figure 5.11 C336A mutation did not affect interaction of Mos1 subunit.

A. The cysteine in position 336 of Mos1 protein was mutated to alanine by changing the codon from TGT to GCT. The C336A mutation was generated by PCR using primer C336(-) as follow: 5'-3': CTC GAG TTA TTC AAA GTA TTT GCC GTC GCT AGC TAC AGC TTT TTC.

B. pACTIIst-Mos1(wt)-containing L40 was transformed with pBTM116-Mos1(wt) or pBTM116-Mos1(C336A) plasmids. Transformants were transferred to -LWH +10 mM 3-AT plate to test for His3 expression. C336A mutants (Sections A and B) had no effect on cell growth as compared with wild-type (Section C).

We conclude that the interaction of transposase monomers is not due to a single protein-protein interaction domain but is due to interacting sites along the length of the molecule.

5.3 Discussion

The yeast two-hybrid system presents several advantages over alternative assays for protein-protein interactions such as crosslinking, co-immunoprecipitation and co-fractionation by chromatography. First, since it is based on a powerful genetic selection scheme performed with a convenient microorganism (yeast), it allows very high numbers of potential coding sequences to be assayed in a relatively simple experiment. Second, it relies on assays performed *in vivo* , so it is not limited by the artificial conditions of *in vitro* assays and there is no need to purify protein. Finally, since it is based on a physical binding assay, a wide variety of protein-protein interactions can be detected and characterised following one single commonly used protocol.

In this chapter the "Reverse Two-hybrid System" was used to identify single amino acid changes that disrupt protein-protein interactions from large libraries of randomly generated mutant alleles. A "reverse" two-hybrid system was designed to identify events that dissociate protein-protein interactions. The isolated mutations may either affect a residue that directly participates in the interaction or disrupt local structural elements that help support the actual contact domains. The facility of the yeast two-hybrid system to identify proteins carrying dissociating mutations from randomly generated populations of mutations was first demonstrated by Lie and Fields in a screen for mutations in the tumour suppressor p53 that disrupt binding to simian virus 40 Large T antigen (Li and Fields, 1993). The library was screened for fusions that produced pale blue or white colonies when exposed to X-Gal. Although some of the dissociating events identified in this manner were due to a failure of yeast to produce stable full-length p53 (presumably due to frame shifts, nonsense mutations and destabilising mutations), 34 distinct p53 mutants were identified that were stably expressed in yeast but had an attenuated interaction, or no interaction with Large T antigen. About 40% (14) of the plasmids contained a single mutation.

To introduce point mutations into Mos1 transposase, polymerase chain reactions were performed under non-optimal conditions that reduce the fidelity of DNA synthesis. *Taq* DNA polymerase lacks a 3'-5' exonucleolytic editing activity and has a measured error rate of 10^{-5} to 10^{-4} error per nucleotide synthesised (Zhou et al., 1991). The frequency of mutation can be manipulated experimentally by increasing the number of cycles and /or

increasing the error rate of the polymerase. This latter factor can be adjusted by altering parameters such as the addition of transition metal ions, e.g. Mn^{2+} (Leung et al.,1989), the dNTP composition and concentration (Spee et al.,1993). These methods were used to generate random mutations as described in 5.3.3. Only method 1 favours single point mutations. Method(2) and (3) usually produce mutants with multiple changes that result in two or more amino acid substitutions (O'Neill et al., 1998).

Interestingly, those single point mutations with defective interactions were loosely clustered in the region encompassing residues 90-160. Amino acid changes at position 28 and 302 severely disrupted the interaction which were compatible with the results of the deletion mutations Mos1 (30-345) and Mos1(1-300). These two deletions were completely defective in dimerization. Loss of *His* expression could be due to either failure of the Mos1(wt)-Mos1(mut) interaction to occur or failure of the yeast to synthesise the intact Mos1(mut)-LexA hybrid. The latter situation would result from mutations that prematurely terminate the synthesis of the Mos1(mut)-LexA hybrid or lead to its instability in yeast. However, both the deletion mutations (data not shown) and most of the single point mutations analysed in western blot (5.3.3) resulted in expression of the intact hybrids.

The crystal structure of the complex between the N-terminal DNA-binding domain of Tc3 transposase and a oligomer of transposon DNA has demonstrated the specific DNA-binding domain contains three α-helices, of which two form a helix-turn-helix motif, while the first helix is involved in dimerization of the protein domains in the crystals, although it is not known if this is the case for the active transposition complex in solution. Using yeast two-hybrid system, we found that the full-length Mos1 transposase subunit could interact physically. However, the interaction between the N-terminal truncated polypeptide corresponding to Mos1-N1-177 and the full-length transposase reduced severely. This suggests the dimerization domain of Mos1 transposase is not within the N-terminal, implying the Mos1-N150 binds to the transposon as a monomer, not a dimer.

Chapter 6

In vitro transposition assays of Mos1 mutants

6.1 Introduction

The work described in chapter 5, 12 single point mutations which cause defective subunit interactions of Mos1 were isolated using the yeast two-hybrid system. To investigate whether these mutations affect transposition, the corresponding mutant proteins must be purified and their activity examined in *in vitro* excision and transposition assays.

In order to investigate the effect on transposition of mutations affecting protein-protein interactions, we must be sure that the mutations do not affect the other functions of transposases: DNA binding and DNA cleavage. Twelve single-point mutations are distributed over the whole open reading frame of *Mos1* . The F53Y, Q91R, L92H, A93V, S104P, V120A mutations occur in the DNA binding domain and four of them (Q91R, L92H, A93V, S104P) are within the HTH motif. These mutations are likely to have decreased DNA-binding activity and therefore would be expected to be inactive in transposition and excision assays. Indeed, the S104P mutation totally abolished the DNA binding activity of Mos1-N150 (Fig 4.9).

The catalytic domain of Mos1 transposase must include at least amino acids 156-284, the position of the first and the last amino acids of the D,D34D triad. Mutations W159R, I160T, Y237C, D279G which occur within this region are likely to affect the catalytic activity of the transposase and would also be unsuitable for *in vitro* assays.

The L124S and S302P mutations located in neither the DNA-binding nor the catalytic domain are therefore suitable candidates to be purified and studied further. The results described in this chapter show that the Mos1(L124S) mutant reduces transposase activity in both excision and transposition assays, indicating that subunit interactions are involved in the transposition reaction.

6.2 Results

6.2.1 Construction of plasmids

Expression vector pBCP378 is designed to allow the direct translation of an ORF from the *trc* promoter in *E.coli* (Velterop et al., 1995). The mutant *Mos1* ORF was subcloned into pBCP378 after amplification by PCR using either pBTM116-*Mos1*(L124S) or pBTM116-*Mos1*(S302P) as template. The forward primer N6799 is homologous to the first 17 bases of the Mos1 ORF and incorporates an *Nde*I site at the initiation codon ATG. The reverse primer T345(-) CGGATCCTAATTCAAAGTATTTGCCGTCG contains a site for the restriction enzyme *Bam*HI. 1 kb PCR products was analysed on an agarose gel before "TA" cloning into pGEM-T vector. The resulting plasmid pGEM-T-*Mos1*(mut) was digested with *Nde*I and *Bam*HI, gel purified and ligated to *Nde*I/*Bam*HI cut pBCP378. Positive clones of pBCP378-*Mos1*(mut) were detected by diagnostic restriction digestion and sequencing.

6.2.2 Expression and purification of Mos1 transposase mutants in *E.coli*

The positive pBCP378-*Mos1*(mut) constructs were transformed into competent BL21(DE3) strain of *E.coli*. Transformed BL21(DE) cells were incubated in LB containing ampicillin (100μg/ml) at 37°C for about 3 h until OD_{595} reached 0.5 and then isopropyl-β-D-galactopyranoside (IPTG, Boehringer Mannheim) was added to 0.5 mM . The cells were incubated for an additional 2 h at 37°C. Cells were harvested by centrifugation and resuspended in 5 ml of 20 mM Tris-HCl (pH7.5), 10% glycerol, 2 mM $MgCl_2$, 1 mM DTT. Lysozyme was added to a concentration of 0.1 mg/ml and the cells incubated for 5 minutes at room temperature, after which 5 ml of Detergent Buffer [20 mM Tris-HCl (pH7.5), 4 mM EDTA, 0.2 M NaCl, 1% deoxycholate, 1% IGEPAL, 2 mM DTT] was added and left at room temperature for an additional 15 min. $MgCl_2$ was then added to a final concentration of 10 mM followed by the addition of 100 units of DNaseI. The mixture was pipetted up and down until it was no longer viscous and was left at room temperature for a further 15 minutes. The lysate was then centrifuged at 12,500 rpm for 30 min and the supernatant withdrawn and discarded. Mos1 transposase was present in inclusion bodies. The pellet containing inclusion bodies was washed three times in 0.5% IGEPAL, 1 mM EDTA followed by one wash in 6 M urea before finally being resuspended in 0.5 ml of Guanidine Buffer [20 mM Tris-HCl (pH7.5), 6 M guanidine hydrochloride, 5 mM DTT]. After centrifugation at 12,500 rpm for 15 min, the supernatant was diluted into 40 ml of Dilution Buffer [25 mM Tris-HCl (pH7.5), 8 M urea, 5 mM DTT, 10% glycerol] and incubated with pre-equilibrated CM sepharose resin at room temperature for 1.5 hours.

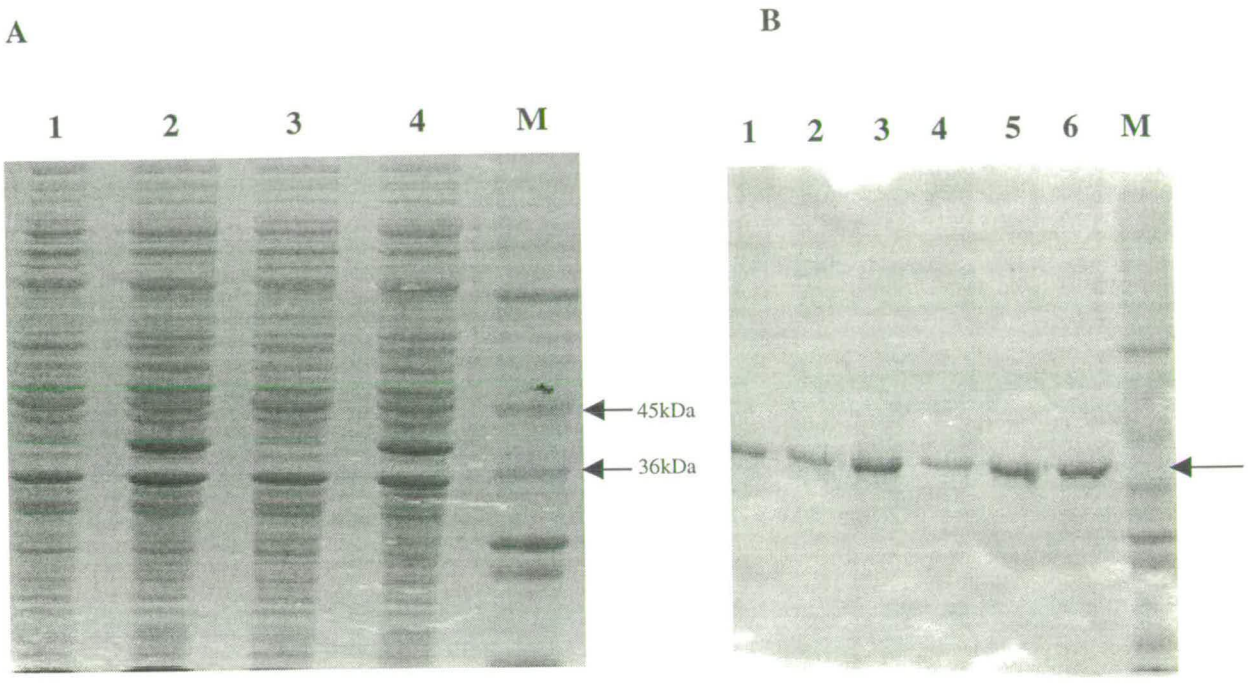


Figure 6.1 Expression and purification of Mos1 transposase mutants.

A. SDS-PAGE analysis of the transposase expression. Lane 1, cells containing pBCP378/*Mos1*(L124S) before induction with IPTG; lane 2, cells containing pBCP378/*Mos1*(L124S) induced with IPTG after 2 hours; lane 3, cells containing pBCP378/*Mos1*(S302P) before IPTG induction; lane 4, cells containing pBCP378/*Mos1*(S302P) after induction; lane M contained molecular mass standards as indicated on the right. **B.** The arrow indicates the purified transposase after washing, solubilization, chromotography and refolding steps. Lanes 1-3, purified Mos1(L124S); lanes 3-6, purified Mos1(S302P).

The resin was washed 4 times with 40 ml of Washing Buffer [25 mM Tris-HCl (pH7.5), 8 M urea, 5 mM DTT, 10% glycerol , 50 mM NaCl], after which protein was eluted from resin by incubation with 3 x 8 ml of Elution Buffer [25 mM Tris-HCl, 8 M urea, 500 mM NaCl, 10% glycerol, 5 mM DTT] for 20 min. Eluted protein was placed in dialysis tubing (Cellulose membrane, 12,000 MWCO) and dialysed against 500 ml of Dialysis Buffer I [50 mM Tris-HCl (pH7.5), 4 M urea, 5 mM DTT, 10% glycerol, 500 mM NaCl, 100 mM triethanolamine] for 4 hours at room temperature. A second dialysis was performed against 500 ml of Dialysis Buffer II [Dialysis Buffer I except containing 2 M urea] for 4 h at 4°C. A third dialysis was carried against 500 ml of Dialysis Buffer III [Dialysis Buffer I except containing 1 M urea] overnight at 4°C. The last dialysis was performed twice against 500 ml of Dialysis Buffer 4 [50 mM Tris-HCl (pH7.5), 5 mM DTT, 10% glycerol, 500 mM NaCl, 100 mM triethanolamine, 30 μM PMSF, 15 μM benzamidine] for 3 hours at 4°C. The samples were then removed from the dialysis tubing and filtered through 0.2 μm filter to remove any aggregates. The protein was aliquoted and stored frozen at -70°C. The concentration of protein was measured by means of the Bradford assay and typically was 250-350 μg/ml. Protein expression and purification were analysed by SDS-PAGE. The native , unfused Mos1 transposase mutants Mos1(L124S) and Mos1(S302P) lacking any tag were expressed well on induction with IPTG and the purified protein had the expected molecular weight of 40.8 kDa (Fig 6.1).

6.2.3 DNA binding activity of Mos1 transposase mutants

To assess whether the DNA binding activity of Mos1 transposase mutants is affected, gel retardation assays were performed using purified full-length Mos1 protein, both wild-type and mutants. Fig 6.2 shows that Mos1(L124S) mutant had the same DNA binding activities compared with wild-type protein, while the Mos1(S302P) mutant showed reduced DNA-binding activity. Although the DNA-binding domain of Mos1 transposase has been characterised within the N-terminal 120 amino acids (Chapter 3), the substitution of serine to proline at position 302 might affect the correct folding of the whole transposase and subsequently block the ability of the DNA binding domain to recognise the ends of the *Mos1* element. The alternative possibility could be that the Mos1 (S302P) mutant did not fold well during renaturing dialysis, thus the activity of protein decreased as obvious aggregation can be seen in the wells. In view of these results, only Mos1(L124S) mutant was used to do following *in vitro* experiments.

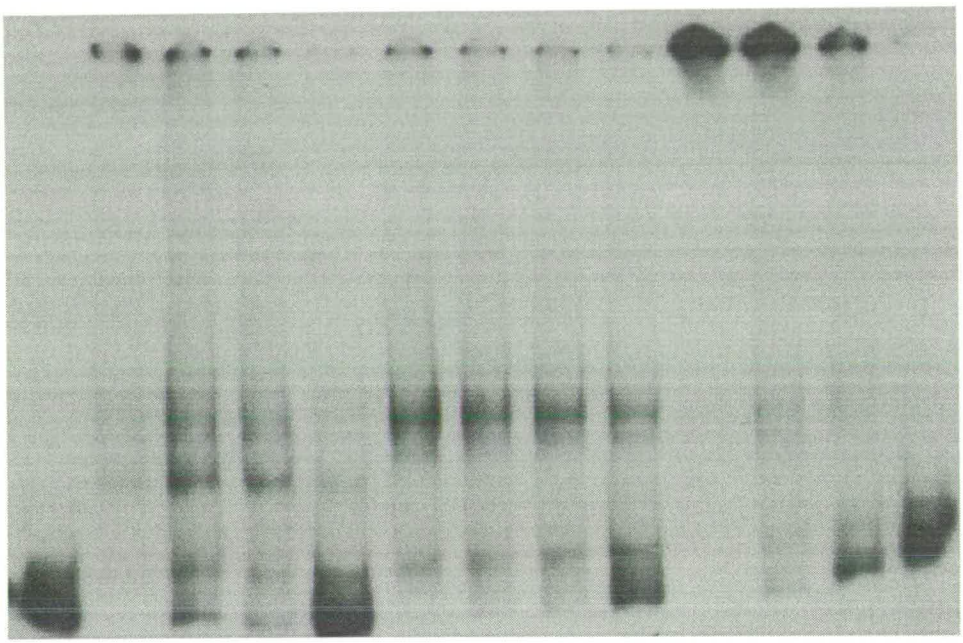


Figure 6.2 Gel retardation assay by Mos1 transposase mutants.

The gel retardation assay was carried out as described in chapter 3. The 176 bp right-end sequence of *Mos1* was labelled with ^{32}P and incubated with wild-type and mutant Mos1 transposase before being separated on a 5% polyacrylamide gel. The position of the probe fragment is indicated by an arrow. Lane 1, probe alone; lanes 2-5, incubation with 224 nM, 56 nM, 28 nM, 2.8 nM of wild-type Mos1; lanes 6-9 show the probe incubated with 28 nM, 14 nM, 7 nM 2.8 nM of Mos1(L124S) mutant; lanes 10-13, incubation with 448 nM, 112 nM, 28 nM and 2.8 nM of Mos1(S302P) mutant.

6.2.4 Impaired excision activity of Mos1(L124S) mutant

Purified transposase protein (both wild-type and mutant) was incubated with supercoiled *pMos* which contains the *Mos1* element. The products were examined by agarose gel electrophoresis and Southern blotting using *pMos* as probe. 1.3 kb DNA fragments were generated that were of the size for linear *Mos1* element (Fig 6.3A). However, the intensity of the excised bands were much weaker when Mos1(L124S) was used rather than wild-type protein. Quantification of the product using Molecular Dynamics Image Quant^{TM} software demonstrated that there was about a 20-fold reduction in excision product in the presence of Mos1(L124S) as compared with equal molar of wild-type Mos1.

A range of dilutions of wild-type Mos1 and Mos1(L124S) mutant were included in the excision assay. A band of approximately 8 kb could be detected in each lane. This is the mobility of linear *pMos* corresponding to single end cleavage. When wild type Mos1 protein was substituted by Mos1(L124S) the 6.7 kb band corresponding to the plasmid vector was reduced in intensity as was the 1.3 kb excised *Mos1* fragment (Fig 6.3B). As shown in Chapter 5, Mos1(L124S) mutant caused defective Mos1 subunit interaction. This implies the second end cleavage requires Mos1-Mos1 interaction.

6.2.5 Cleavage at the right end is about 10 times more efficient than at the left end of *Mos1*

The 8kb band is presumably to correspond to linear plasmid in which cleavage at either right or left end has occurred. To investigate whether each end is cleaved with equal efficiency. This band was cut from the gel and purified using Qiagen Gel Extraction Kit. DNA was digested with *Sal* I and separated by 1.3% agarose gel. Products were identified by Southern blotting analysis by probing with the 1.3 kb *Mos1*. Cleavage at the right end would be revealed by the appearance of a 937 bp fragment and cleavage at the left end would produce a 349 bp fragment (Fig 6.4). Quantification of the products using Molecular Dynamics Image Quan^{TM} software demonstrated that the level of left-end cleavage products is about 10-fold less abundant than the level of the right-end cleavage mediated by either wild-type transposase or Mos1(L124S) mutant. However, in the presence of mutant (lanes 1 and 2), the cleavage at the left-end seems imprecise. Several bands slightly

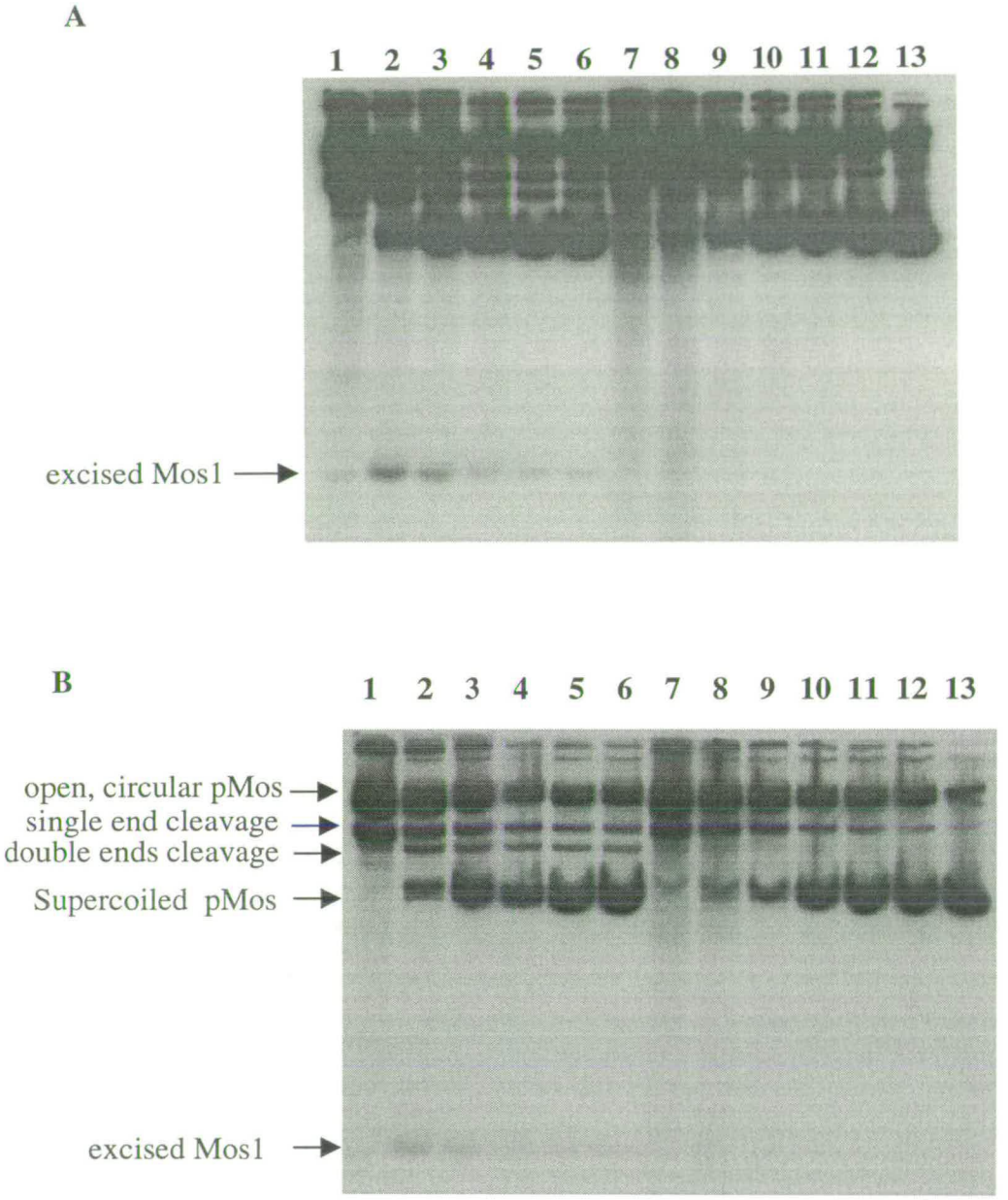
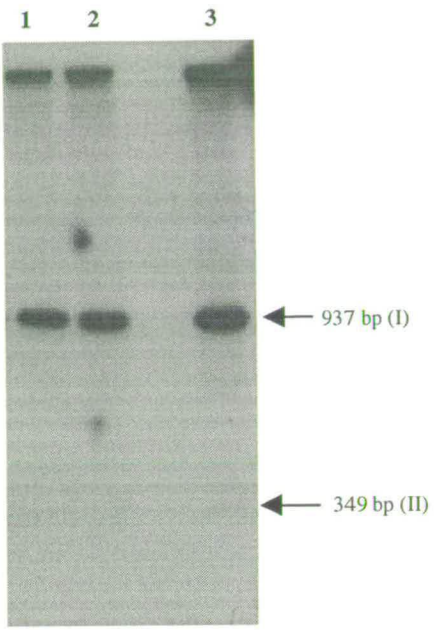


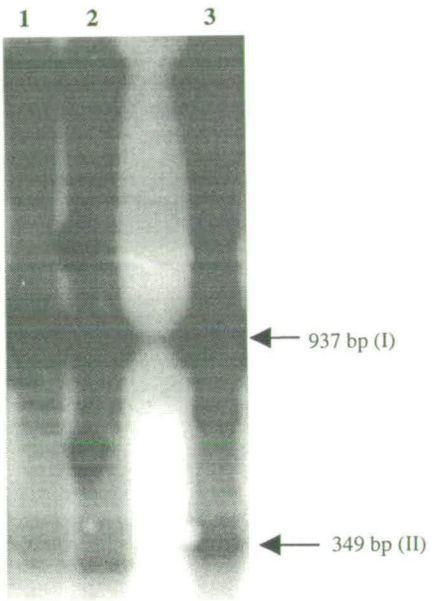
Figure 6.3 ***In vitro*** **excision of Mos1 transposable element from plasmid pMos.**

A. Products of in vitro transposition reactions were separated on a 1% agarose gel, transferred to nitrocellulose, and probed with radiolabelled *pMos*. Lanes 1-6 show reaction products using 1200 nM, 600 nM, 300 nM, 150 nM, 75 nM, 37.5 nM of wild type Mos1 transposase; lanes 7-12 contain 1200 nM, 600 mM, 300 nM, 150 nM, 75 nM, 37.5 nM of Mos1(L124S) mutant. **B**. Same as **A** with short-time X-ray exposure. The arrows indicate various reaction intermediates and products.

A



B



C

	lane1	lane 2	lane 3
I	353034	447745	916195
II	9814	15894	40008
R=I/II	36	28	23
R/2.68*	13.4	10.4	8.6

* the ratio of 937bp/349 bp, correction factor for the different length of DNA.

Figure 6.4 Different efficiency of cleavage at two transposon ends.

A. 8 kb single-end cleavage product was digested with *Sal* I and separated by 1.3% agarose gel. Southern blots were developed with 1.3 kb *Mos1* probe. Lanes 1 and 2 show products from reaction with 300 nM, 600 nM of Mos1(L124S) mutant; lane 3 shows reaction products using wild-type Mos1 transposase. 937 bp and 349bp indicate the positions of the right and the left end cleavage, respectively. **B.** Same as A with long exposure of the autoradiogram to show the left-end cleavage more clearly. **C**.The reaction products were quantified by phosphorimaging.

above or below those resulting from authentic cleavage (lane 3) might arise from imprecise cleavage at the left-end.

6.2.6 Impaired transposition activity of Mos1(L124S) mutant

Transposition activity was examined in an *in vitro* plasmid based genetic assay (Fig 6.5). The donor plasmid pRJM345Mos1Tet made by Dr. Angela Dawson contains the complete *Mos1* element into which a tetracycline resistance cassette had been inserted. Target plasmid was a tetramer of pBSKS+ (Kaufman and Rio, 1992). The donor plasmid was incubated with the target plasmid in the presence of transposase for 2 hours at 30^{o}C. The products were electroporated into DH10B *E.coli* cells, allowing selection against the transposon donor and detection of transpositions by the occurrence of doubly resistant (Tet^{R} and Amp^{R}) colonies. The efficiency of transposition was scored by counting the number of colonies that grew in the presence of both ampicillin and tetracycline, relative to the total number of ampicillin resistance colonies, which served as a control for the amount of plasmid DNA recovered from the *in vitro* reaction (Tab 6.1). In the presence of 40 mM wild-type Mos1 transposase, approximately 2.2 x 10^{-3} transposition events per target molecule were obtained. When wild-type transposase was replaced by Mos1(L124S) mutant, the transposition frequency dropped about 20-fold under otherwise identical conditions. No Amp^{R},Tet^{R} colonies were obtained in the absence of transposase. *BamH*I restriction analysis of recovered plasmids confirmed the presence of transposition in all cases (Fig 6.6). In each case the transposition products produce three *BamH*I restriction fragments: a prominent fragment of 2.9 kb (the monomer length of pBSKS+), and two fragments whose lengths vary depending upon the site of the *Mos1* element insertion, and whose sizes sum to 5.6 kb.

6.2.7 Randomness of *Mos1* insertion *in vitro*

Sequence analysis of the transposon termini and the insertion site of Amp^{R}-Tet^{R} products confirmed that they were *bona fide* transposition events. The target site of 14 insertions obtained in the presence of Mn^{2+} are summarised in Figure 6.7A. Only 4 of them (28.6%) inserted into a TA dinucleotide, while 50% of the transposition products in the presence of wild type transposase inserted into the TA dinucleotides. The other target site duplications such as CA, GA, TG, TC obtained in the presence of Mos1(L124S) mutant were never

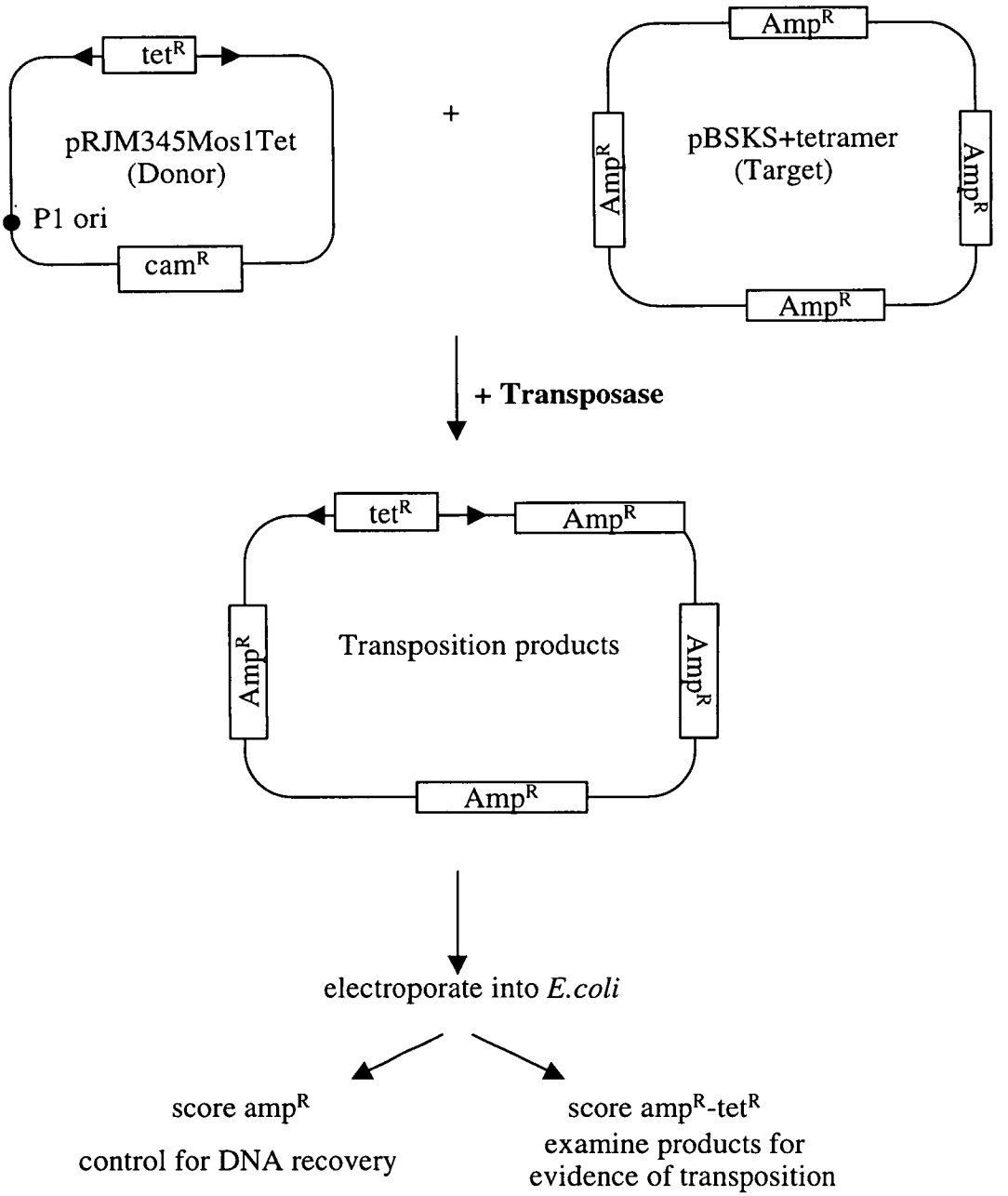


Figure 6.5 Schematic diagram of *In vitro mariner* transposition assay.

The chloramphenicol(cam), tetracycline(tet) and ampicillin(amp) drug resistance genes on the donor and target DNA are indicated. The donor plasmid with bacteriophage P1 origin of replication could be grown in *E.coli* strain RR4478 that contains P1 rep protein but not in strains that lack this protein. The arrows represent the inverted terminal repeats of *Mos1*.

Table 6.1 *In vitro* transposition assay with wild-type Mos1 and Mos1(L124S) mutant in the presence of Mn^{2+}.

Transposase concentration	Transposition efficiency[a] ($x10^{-4}$)	Relative efficiency[b] (%)
40 nM (wt[c])	22	100
20 nM (wt)	12	54.5
10 nM (wt)	5.4	24.5
5 nM (wt)	5	22.5
50 nM (mut[d])	0.65	3.0
25 nM (mut)	0.95	4.3
12.5 nM (mut)	0.7	3.2
6 nM (mut)	1.2	5.5

[a]The ratio of amp^R,tet^R to amp^R gives a measure of the transposition efficiency.

[b]Relative efficiencies are calculated by dividing reach of the absolute transposition efficiencies by the absolute maximum transposition efficiency obtained with 40 nM of wild-type Mos1.

[c]wt is referred to wild-type Mos1 transposase.

[d]mut is referred to Mos1(L124S)mutant.

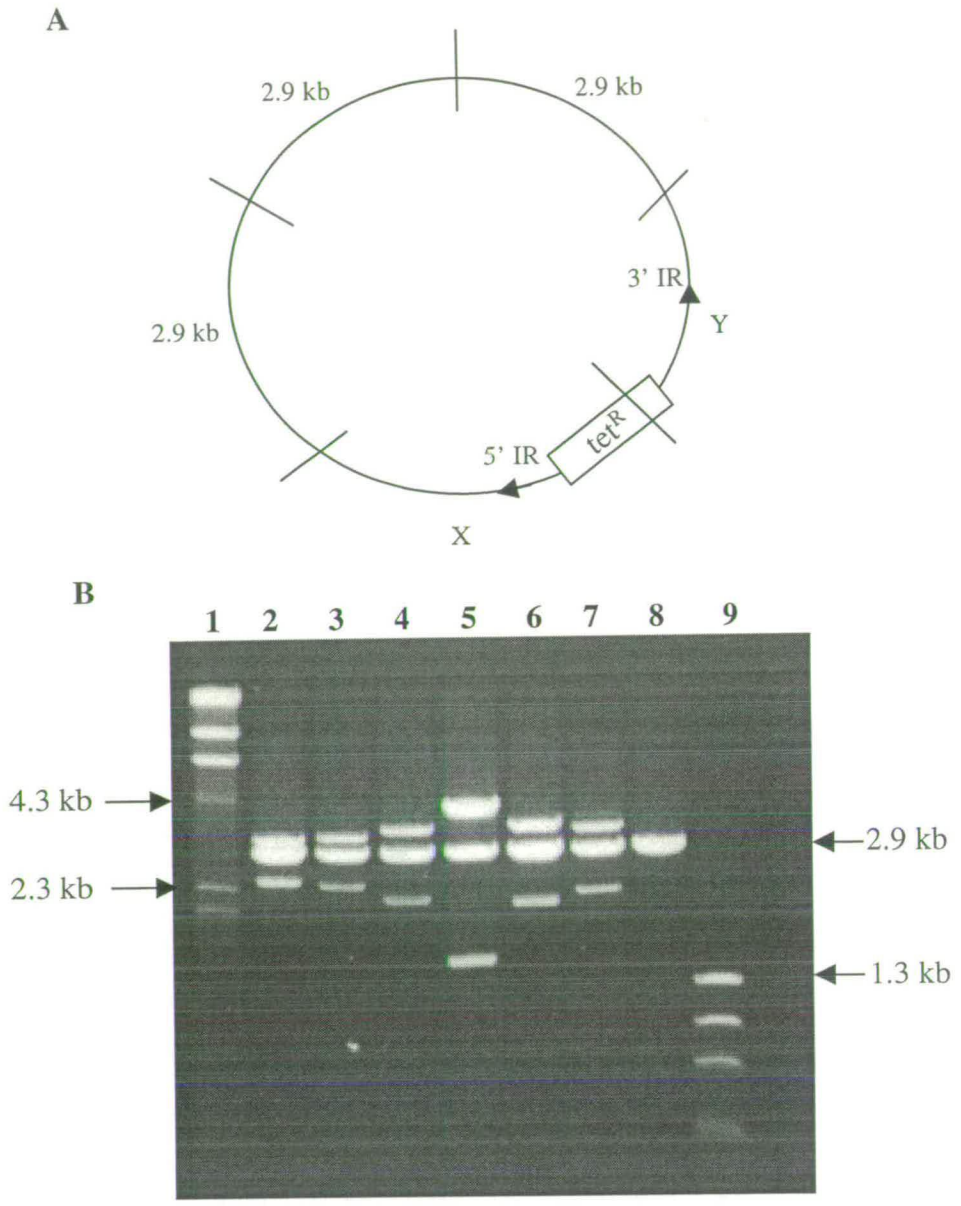


Figure 6.6 Structure of *in vitro* transposition products

A. Schematic diagram of a simple Mos1-tet^R transposon insertion into the pBSKS+tetramer target plasmid. Two arrows indicate the 5' and 3' terminal IRs of the *Mos1* element. Line segments perpendicular to the plasmid represent *Bam*HI restriction sites, which occur once per monomer length of the plasmid, and also once in the tet^R gene within the *Mos1* element. X and Y refer to the length of two fragments whose length will vary depending on the site of insertion, but whose combined length equals 5.6 kb. **B**. *Bam*HI digestion analysis of plasmids from amp^R-tet^R colonies. Lane1, Lambda DNA-*Hind*III marker; lanes 2-7, 6 different amp^R-tet^R plasmids; lane 8, pBSKS+tetramer target plasmid; lane9, ϕX174 DNA-*Hae*III marker.

observed in the presence of wild type transposase in which only TA, TT, AT duplications occurred (Dawson, personal communications). In one case (No.14), the target site duplication was changed from 2 nucleotides to 3 nucleotides, as seen with the *Himar1* transposase (Lampe et al., 1996). 14 insertions in a 2.9 kb pBSKS+ target is shown in Figure 6.8 and revealed little regional specificity. Similar results were obtained with wild-type transposase (Dawson, personal communications).

Transposition in the presence of Mg^{2+} yields a different spectrum than observed with Mn^{2+}. All insertions generated in the presence of wild type transposase with Mg^{2+} as the divalent cation were flanked by TA target site duplications (Tosi and Beverley, 2000; Dawson, personal communications) as is the case for *mariner* elements transposing *in vivo* (Bryan et al., 1990). We were interested in determing whether substitution of Mg^{2+} for Mn^{2+} might restore the requirement for insertion into a TA dinucleotide in the presence of Mos1(L124S) mutant. Plasmid DNAs from 8 Amp^R,Tet^R colonies derived from transposition in the presence of Mos1(L124S) with Mg^{2+} as divalent cation were analysed by restriction digestion and agarose gel electrophoresis. We determined the junction sequence between the *Mos1* element and plasmid DNA. While only 3 of 8 insertions were flanked by TA dinucleotides, the remainder proved to have inserted at other target sites such as TC, CA. The sequence TC at position 1798 of pBSKS+ was used a total of 4 times (Fig 6.7B).

6.3 Discussion

Transposases are multifunctional enzymes that must accomplish a series of tasks which including recognition and binding to the ends of the element, bringing them together into a synaptic complex and catalysing DNA cleavage and strand transfer. These functions have been shown to be shared by different domains of the transposases: DNA binding domain, catalytic domain and dimerization domain. In Tn5 transposase, these various regions are known to overlap extensively (Braam et al., 1999). The L124S mutant occurs outside of the regions of DNA-binding domain and catalytic domain of Mos1 transposase. The impaired activity of Mos1(L124S) transposase mutant is most likely due to a defective dimerization domain.

The results of an excision assay showed that single-end cleavage was not disrupted in the presence of Mos1(L124S), while double-end cleavage was significantly impaired. This

No.	Target	dupl	5'ITR	3'ITR	dupl	Target
1	AAAGTATA	**TA**	CCAGGTGTA...TACACCTGA		**TA**	TGAGTAAA
2	TCTGAGAA	**TA**	CCAGGTGTA...TACACCTGA		**TA**	GTGTATGC
3	TCGCCGCA	**TA**	CCAGGTGTA...TACACCTGA		**TA**	CACTATTC
4	CCAGGCTT	**TA**	CCAGGTGTA...TACACCTGA		**TA**	CACTTTAT
5	ATGCTCGT	**CA**	CCAGGTGTA...TACACCTGA		**CA**	GGGGGGCG
6	TTGGAGTC	**CA**	CCAGGTGTA...TACACCTGA		**CA**	CGTTCTTT
7	CTTGAGTC	**CA**	CCAGGTGTA...TACACCTGA		**CA**	ACCCGGTA
8	TAAGACAC	**GA**	CCAGGTGTA...TACACCTGA		**GA**	CTTATCGC
9	ACGGTATC	**GA**	CCAGGTGTA...TACACCTGA		**GA**	TAAGCTTG
10	TCGATTTT	**TG**	CCAGGTGTA...TACACCTGA		**TG**	TGATGCTC
11	GGAAATGT	**TG**	CCAGGTGTA...TACACCTGA		**TG**	AATACTCA
12	TAGTTCTT	**TC**	CCAGGTGTA...TACACCTGA		**TC**	CTGCGTTA
13	AAAGGATC	**TC**	CCAGGTGTA...TACACCTGA		**TC**	AAGAAGAT
14	CATGAGTG	**ATA**	CCAGGTGTA...TACACCTGA		**ATA**	ACACTGCG

Figure 6.7A Target site selection of Tet^R-Amp^R products in the presence of Mn^{2+}.

Sequences of the termini of the *Mos1*-Tet element (terminal-most nine nucleotides only) integrated into the pBSKS+ target DNA. Dots represent sequences between the termini of the element removed for this figure. Target site duplications are in bold type.

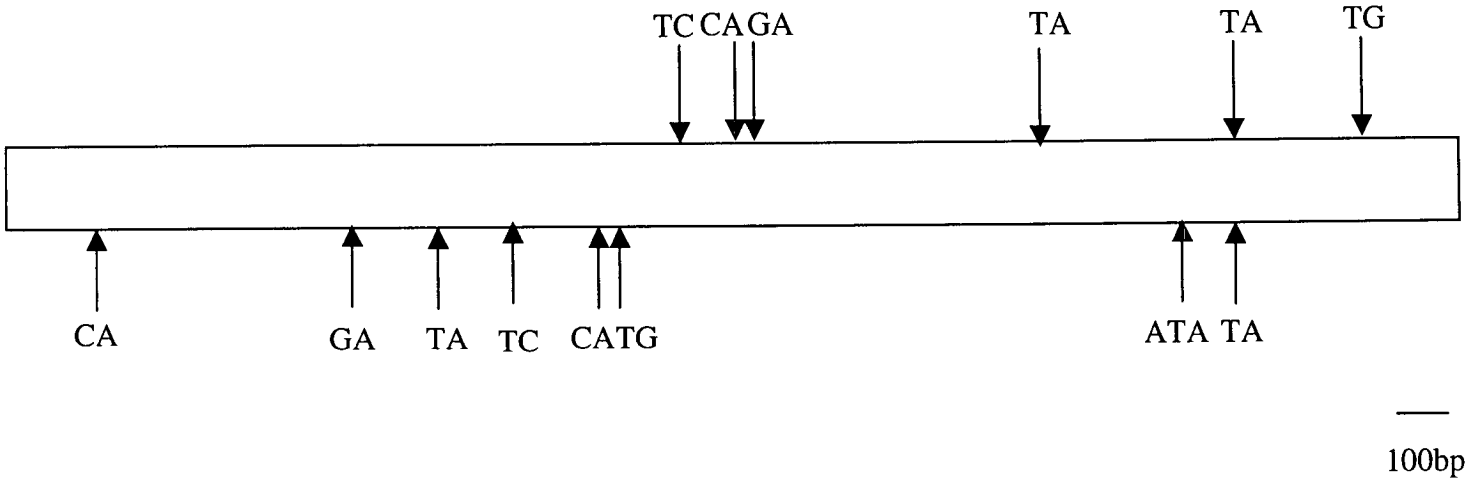


Figure 6.8 Randomness of *Mos1* insertion mediated by Mos1(L124S) mutant.

Insertion of *Mos1* into 2.9 kb pBSKS+ target in reaction containing Mn^{2+}. The vertical arrows represent individual insertions. Those above or below the map represent insertions with the Tet^R marker in the forward or reverse orientation, respectively.

No.	Target	dupl	5'ITR	3'ITR	dupl	Target
1	TAATCCGG	**TA**	CCAGGTGTA...TACACCTGA		**TA**	ACTATCGT
2	ATTCGCCC	**TA**	CCAGGTGTA...TACACCTGA		**TA**	TAGTGAGT
3	TGCTTCAA	**TA**	CCAGGTGTA...TACACCTGA		**TA**	ATATTGAA
4	AAAGGATC	**TC**	CCAGGTGTA...TACACCTGA		**TC**	AAGAAGAT
5	AAAGGATC	**TC**	CCAGGTGTA...TACACCTGA		**TC**	AAGAAGAT
6	AAAGGATC	**TC**	CCAGGTGTA...TACACCTGA		**TC**	AAGAAGAT
7	AAAGGATC	**TC**	CCAGGTGTA...TACACCTGA		**TC**	AAGAAGAT
8	ATGCTCGT	**CA**	CCAGGTGTA...TACACCTGA		**CA**	GGGGGGCG

Figure 6.7B Target site selection of Tet^R-Amp^R products in the presence of Mg^{2+}.

Sequences of the termini of the *Mos1*-Tet element (terminal-most nine nucleotides only) integrated into the pBSKS+ target DNA. Dots represent sequences between the termini of the element removed for this figure. Target site duplications are in bold type.

confirms that Mos1(L124S) mutant has both DNA binding and catalytic activities and suggests that excision of *Mos1* from the vector by wild type enzyme is due to co-ordinated cleavage at both ends due to interaction between monomers bound to both ends, an interaction which is reduced in Mos1(L124S).

Cleavage was about 10 fold less efficient at the left end than at the right . This is consistent with our observation that transposase binds preferentially to the right hand end (Chapter 4). The imprecise cleavage mediated by Mos1(L124S) mutant suggests the correctly formed transpososome might be required for correct cutting and those imprecise cleavage products were not capable of further transposition because Figure 6.7 shows all transposition products contain the precise termini of *Mos1*.

Despite the ability of Mos1(L124S) protein to cleave the left and right ends of *Mos1* with a similar efficiency to wild type enzyme the frequency of transposition stimulated by the mutant enzyme *in vitro* was about 20 fold lower again suggesting that the ability of transposase monomers to bind to each other is essential for normal transposition.

The requirement of Mos1 transposase for a divalent cation was analysed *in vitro*. It has been demonstrated that there is an 8-fold decrease of excised *Mos1* bands in excision assay (Smith, 1997) or 10-fold drop of transposition frequency in transposition assay (Dawson, personal communications) when Mg^{2+} was substituted for Mn^{2+}. Although Mn^{2+} is unlikely to be the cation associated with the active site of transposition *in vivo* it has been shown to reduce the stringency of nuclease reactions and to permit reactions which would otherwise fail to proceed *in vitro* (van Gent et al., 1996; Junop and Haniford, 1996). The transposition frequency of the Mos1(L124S) was reduced about 40-fold in the presence of Mg^{2+} as compared with Mn^{2+} and was below the level of detection in most cases. The highest transposition efficiency (1.6×10^{-6}) was obtained with Mos1(L124S) at a concentration of 12.5 nM in the presence of Mg^{2+} and only 8 colonies were obtained for further analysis (Fig 6.7B).

The dinucleotide insertion sites were altered when Mos1(L124S) mutant was substituted for wild-type transposases in the presence of either Mg^{2+} or Mn^{2+}. While 30%-40% of insertions were flanked by the dinucleotide TA other target sites were found including some which have not been reported for previously for *mariner* or any other element of this type (van Luenen et al., 1994; Ketting et al., 1997; Lampe et al., 1998; Tosi and Beverley, 2000). This may reflect a change in target site selection due to a change in the

conformation of the transpososome resulting from altered protein-protein interaction or a direct effect on transposase activity resulting from L124S mutation.

Chapter 7

Promoter activity of *mariner* transposable element

7.1 Introduction

Due to its mutagenic effects, uncontrolled transposition would be deleterious for the host organism and therefore for the survival of the transposable element itself. A variety of mechanisms have evolved that constrain transposition including antoregulation by element-encoded molecules as well as host defence system that protects the integrity of the genome. Knowledge of the regulation of *mariner*–like elements is also of practical importance in learning how best to use these elements in the germline transformation of other organisms. One type of regulatory phenomenon exhibited by *mariner* elements is called 'overproduction inhibition' in which overproduction of the wild-type transposase appears to reduce the overall level of transposase activity as assayed by the excision of a nonautonomous target element (Lohe and Hartl, 1996a). Defining the molecular basis of overproduction inhibition is a priority because it may restrict the level of activity of *mariner* elements. One interesting possibility is that there is a feedback mechanism in which the transposase binds its own promoter and inhibits transcription. Where is the promoter of *mariner* element located? Are there any other regulatory sequences involved in *mariner* transcription? None of these questions has been answered. The primary aim of the work described in this chapter is to identify those sequences within *Mos1* which regulate expression of this element.

The sequences upstream from the presumed start of transcription of many eukaryotic genes transcribed by RNA polymerase II have been determined. An AT-rich region or "TATA" box which play a role in promotion of transcription by RNA polymerase II has been observed about 30 base pairs (bp) upsteam from this point (Breathnach and Chambon, 1981). *P* element transcripts have their 5' ends approximately at nucleotide 87 (Karess and Rubin, 1984). This region is just 28 bases from a TATACA sequence (O'Hare and Rubin, 1983). The rice blast fungus transposon *Pot2* shares structural features with the *Tc1-mariner* superfamily (Robertson, 1995). It is 1857 bp long with 43 bp inverted repeats and also duplicates the dinucleotide TA at the target site (Kachroo et al., 1994). Canonical TATA box motifs are present at both ends of *Pot2* and prove to be functional as promoters (Kimura and Yamaguchi, 1998).

The *mariner* elements have also been predicted to contain such "TATA" box (CATAA) at -29, relative to the potential initiation codon (Jacobson et al., 1986). The *Mos1* promoter is

believed to be relatively weak (Maruyama and Hartl, 1991), although there is no direct evidence. In this chapter, the promoter activity of *Mos1* was detected directly and quantitatively in transgenic flies by *P*-element-mediated germline transformation.

The use of *P* transposable elements as vectors to introduce specific DNA segments into the *Drosophila* germline revolutionised the study of gene regulation and function in *Drosophila* (Rubin and Spradling, 1982). The standard method of germline transformation in *Drosophila melanogaster* involves the co-injection of two plasmids based on the *P* element into fly embryos. One is the "helper" plasmid that provides a source of *P* transposase but cannot transpose and integrate into the genomic DNA of the recipient embryo because of a specific deletion in one of its terminal repeats. The other is the construct to be integrated that carries the inverted terminal repeats of the *P* element and transposes with the aid of the helper. This construct also contains a selectable transformation marker and a few unique restriction sites for cloning the desired DNA fragment. In this chapter, the *lacZ P*-element vectors were used to study the regulatory elements that control the spatial distribution of gene expression. These vectors allows fusion of *Drosophila* promoters to the genes encoding *E.coli* ß-galactosidase and can be used for germline transformation or transfecting tissue culture cells. The presence of ß-galactosidase activity in a tissue can be detected simply by its conversion of the chromogenic substrate X-gal (5-bromo-4-chloro-3-indolyl-ß-D-galactoside) or ONPG (*o*-nitrophenyl-ß-D-galactopyranoside).

7.2 Results

7.2.1 *P*-element mediated germline transformation

The PCR product containing the first 171 bp of the *Mos1* element was inserted upstream from *lacZ* reporter gene of pCaSpeR-AUG-ßgal. The Adh sequence of pCaSpeR-AUG-ßgal contains the AUG start codon which allows translation of ß-galactosidase (Thummel et al., 1988). The PCR products of the first 1206 bp, 861 bp and 516 bp of *Mos1* were subcloned into an upstream polylinker of pCaSpeR-ßgal respectively. Insertions of these three DNA fragments must be in-frame with the *lacZ* gene which results in the expression of a ß-galactosidase fusion protein (Thummel et al., 1988). Both of these transformation

vectors contains a *Drosophila white*$^{+}$ gene and transformants can be identified by screening for restoration of eye pigmentation in *white*$^{-}$ flies.

The plasmid DNA was purified by the "Bulk prep" protocol (see Chapter 2) kindly provided by Ruth Kirby. Flies of genotype w^{1118} were used as recipients. The w^{1118} allele included a deletion of part of the *white* gene. The injected DNA solution contained 500 µg/ml of the vector of interest along with 250 µg/ml of "wings-clipped" helper plasmid pπ25.7wc. Transformations were carried out essentially as described in Karess (1985). Embryos were injected and incubated under oil at 18°C. Emerging larvae were transferred to standard Drosophila medium and maintained at 25°C. G_0 flies were collected and mated individually to w^{1118} virgins. The resulting G_1 flies were assayed for a coloured-eye phenotype, which ranges from pale yellow to wild type brick red (Fig 7.1). Transformants were crossed to w^{1118} flies to produce males and females with the same insertion. Homozygous lines were then established by sibling matings. The transformation efficiency of four constructs are summarised in Table 7.1.

Transgenic *Drosophila melanogaster* lines were characterised by Southern blotting to confirm that each line was a real transformed line and to establish how many copies of the transgene were present in each line. In each case the enzyme cuts once within the inserted *P* element and the probe hybridises to only one of the two products of this cleavage. The mobility of the hybridising band depends on the distance from the end of the *P* element to the nearest recognition site for the restriction enzyme in the flanking genomic DNA. Southern blots are shown in Figures 7.2-7.5 and demonstrated that the majority of the lines have only one copy of the transgene.

7.2.2 Expression of ß-galactosidase in transgenic flies

The *Mos1* promoter was presumed to lie between nucleotides 1 and 171 before the ATG start codon of *Mos1* ORF. The P171 construct carrying bases 1 to 171 of *Mos1* was injected and 9 independent lines were generated. Ovaries from these lines were dissected and stained for ß-galactosidase activity. No staining was visible in each case. This could be due to low expression of ß-galactosidase. RT-PCR was then done to detect low level of mRNA. Total RNA was isolated from ovaries using Qiagen RNeasy Mini Kit. Reverse transcription was performed using oligo(dT) as the primer and 1 µg of total RNA according

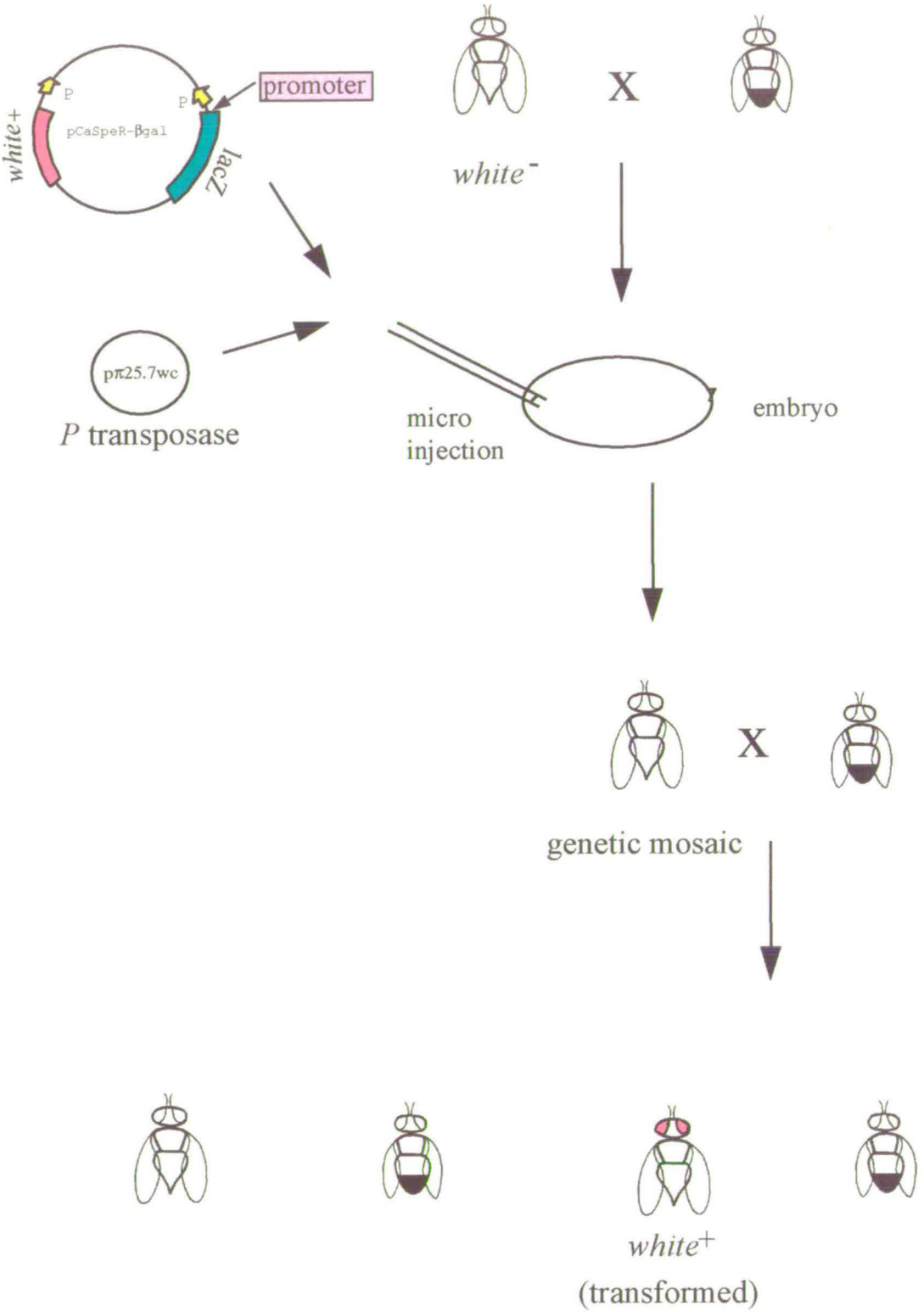


Figure 7.1 *P* element-mediated germline transformation

DNA injected at the posterior pole prior to cellularization will become incorporated into germline precursors. Adults that develop from injected embryos (G_0) are genetic mosaics with respect to the presence of transposon in their germline. In the next generation (G_1) the transformed individuals can be selected.

Table 7.1 Germline transformation data

Construct	No. of embryos injected	No. of G_0	No. of fertile G_0	No. of G_0 with transformed G_1	efficiency (%)
P171	471	68	60	9	15
P516	330	96	90	11	12.2
P861	370	83	78	8	10.2
P1206	374	85	76	12	15.7

Transformation efficiency is the percentage of fertile G_0 adults giving transformed G_1. P171 is referred to the first 171 bp of *Mos1* inserted into pCaSpeR-AUG-ßgal; P516, P861, P1206 referred to the first 516 bp, 861 bp and 1206 bp of *Mos1* inserted into pCaSpeR-ßgal respectively.

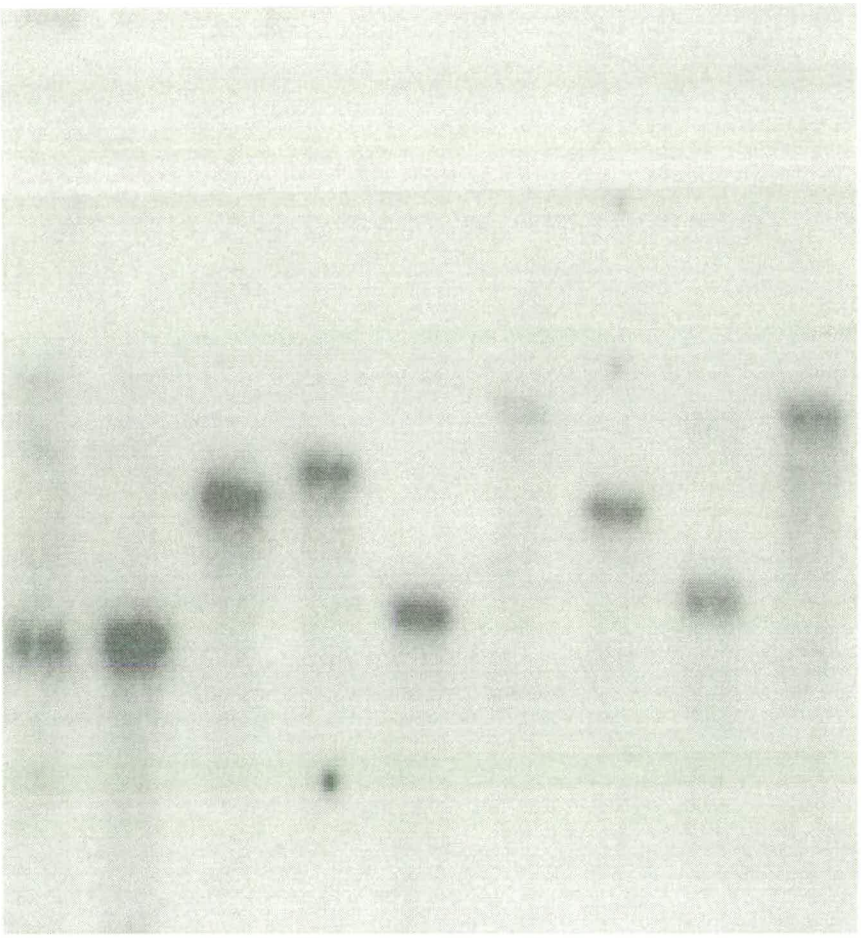


Figure 7.2 Southern blotting analysis of P171 transgenic lines.

Genomic DNA from 10 flies was digested with *Bam*HI overnight and run on a 0.8% agarose gel at 20 mV. Following Southern transfer the blots were probed with the 3.5 kb *Xba*I/*Bam*HI fragment from the pCaSpeR-AUG-ßgal vector. The numbers referred to different transgenic lines.

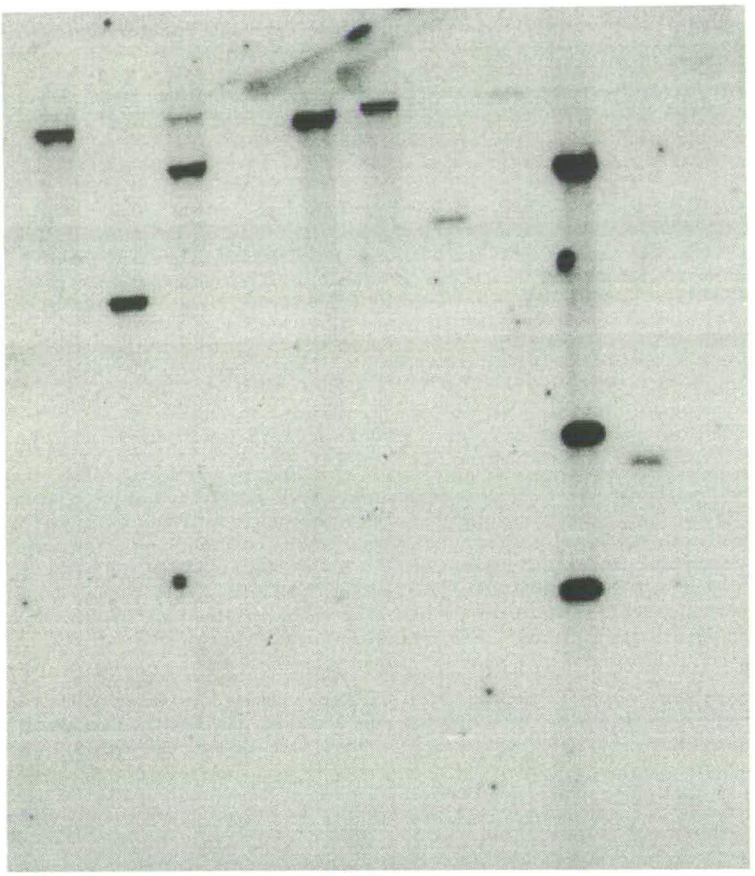


Figure 7.3 Southern blotting of P516 transgenic flies.

Genomic DNA from 10 flies was digested with *Bam*HI overnight and run on a 0.8% agarose gel at 20 mV. Following Southern transfer, the blots were probed with the PCR product containing 1.3 kb Mos1 sequences. The numbers referred to different transgenic lines.

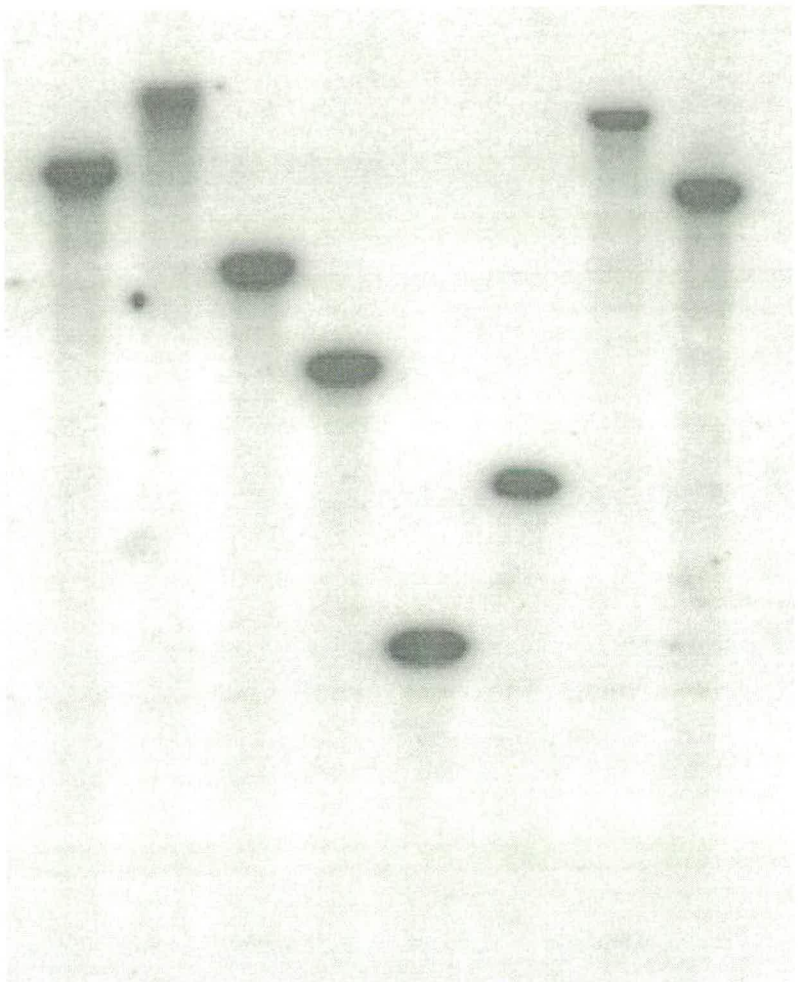


Figure 7.4 Southern blotting of P861 transgenic flies.

Genomic DNA from 10 flies was digested with *BamH*I overnight and run on a 0.8% agarose gel at 20 mV. Following Southern transfer, the blots were probed with the PCR product containing 1.3 kb of Mos1 sequence. The numbers referred to different transgenic lines.

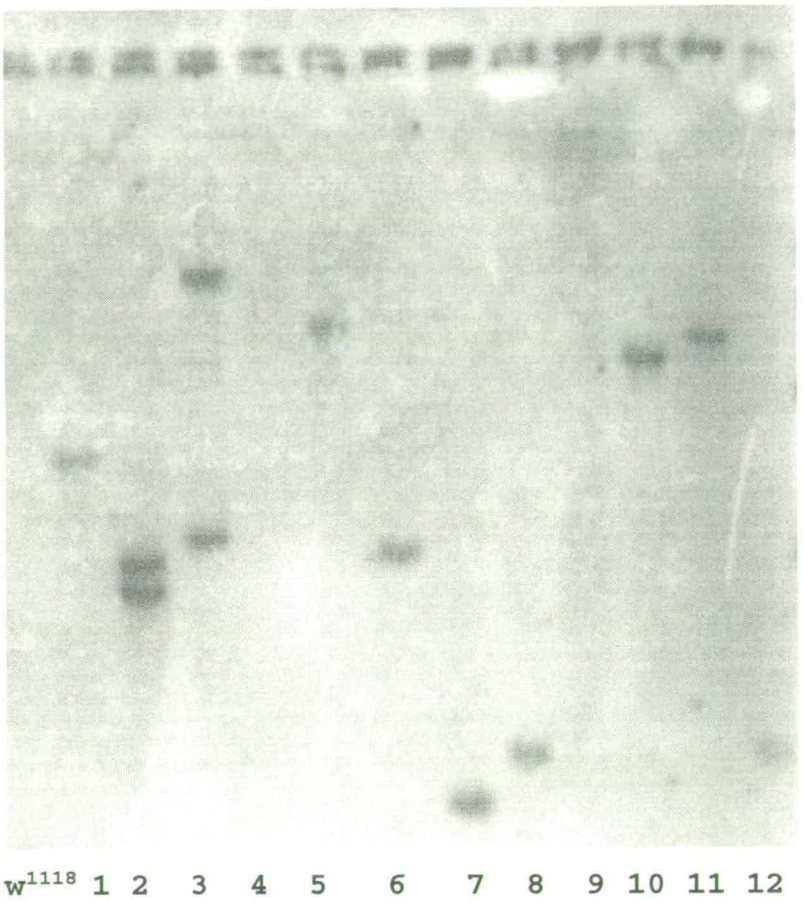


Figure 7.5 Southern blotting of P1206 transgenic flies.

Genomic DNA from 10 flies was digested with *Sph*I overnight and run on a 0.8% agarose gel at 20 mV. Following Southern transfer, the blots were probed with the PCR fragment containing the first 700 bp of Mos1. W^{118} is indicated as negative control. The numbers referred to different transgenic lines.

to the GeneAmp RT-PCR kit protocol (Perkin-Elmer). Amplification of *lacZ* cDNA was performed using 1 µl of first-strand cDNA, with the following *lacZ* gene-specific primers: LacZ(+) (GCCGGTCTGGGAGGCATTGGTCTG); LacZ(-) (GGCCTCAGGAAGATCG CACTCCAG). This generated a product of the expected length of 270 bp. Figure 7.6 shows that *lacZ* is expressed in P171 transgenic lines, while the expression level is markedly lower than the positive control (line 274) which contains the 41 to 186 sequences of the *I* factor upstream of *hsp70* promoter. This implies the activity of 171 bp of *Mos1* promoter is relatively weak.

However, significant maternal effects of Mos1 transposase strongly indicate high expression of *Mos1* in the female germline (Bryan and Hartl, 1988; Lohe et al., 1995). When *Mos*/+ females, homozygous for w^{pch}, were crossed with w^{pch} males from a non-mosaic strain, all offspring of both sexes were found to be mosaic. This maternal effect in which the non-*Mos* offspring were mosaic was mediated by the mariner transposase (or its mRNA) transmitted to the egg maternally. Therefore other regulatory sequences downstream of the first 171 bp region may also be required for *Mos1* transcription. The P1206 construct containing both the first 171 bp and the whole ORF of *Mos1* was injected and 12 independent lines were obtained. ß-galactosidase staining of ovaries was seen in half of them indicating a strong position effect on the expression of *lacZ* gene. This result suggests there is an enhancer within Mos1 ORF which upregulates the activity of *Mos1* promoter. To map the sequences responsible for this enhancer activity, two truncated constructs P861 and P516 were injected. 8 and 11 different lines were generated respectively. ß-galactosidase staining of ovaries showed that 75% lines of P861 and 90% lines of P516 stained blue (Figure 7.7). The pattern of staining varied from very strong staining (++++) to no staining (-) and shown in Figure 7.8.

7.2.3 ß-galactosidase activity assay of transgenic flies

To quantitatively measure and compare the levels of active ß-galactosisase expressed in ovaries of transgenic flies, ß-galactosidase activity assays were performed. ß-galactosidase catalyses the hydrolysis of ONPG to the ONP anion which produces a bright yellow colour with a peak absorbance at 420 nm that can be quantified using a spectrophotometer. The slope of the line obtained by plotting OD_{420} against time was taken as the measure of activity after corrected for the concentration of protein in the extract. The units are OD420/min/g protein. Low levels of activities were observed in w^{1118} flies and w^{1118} flies

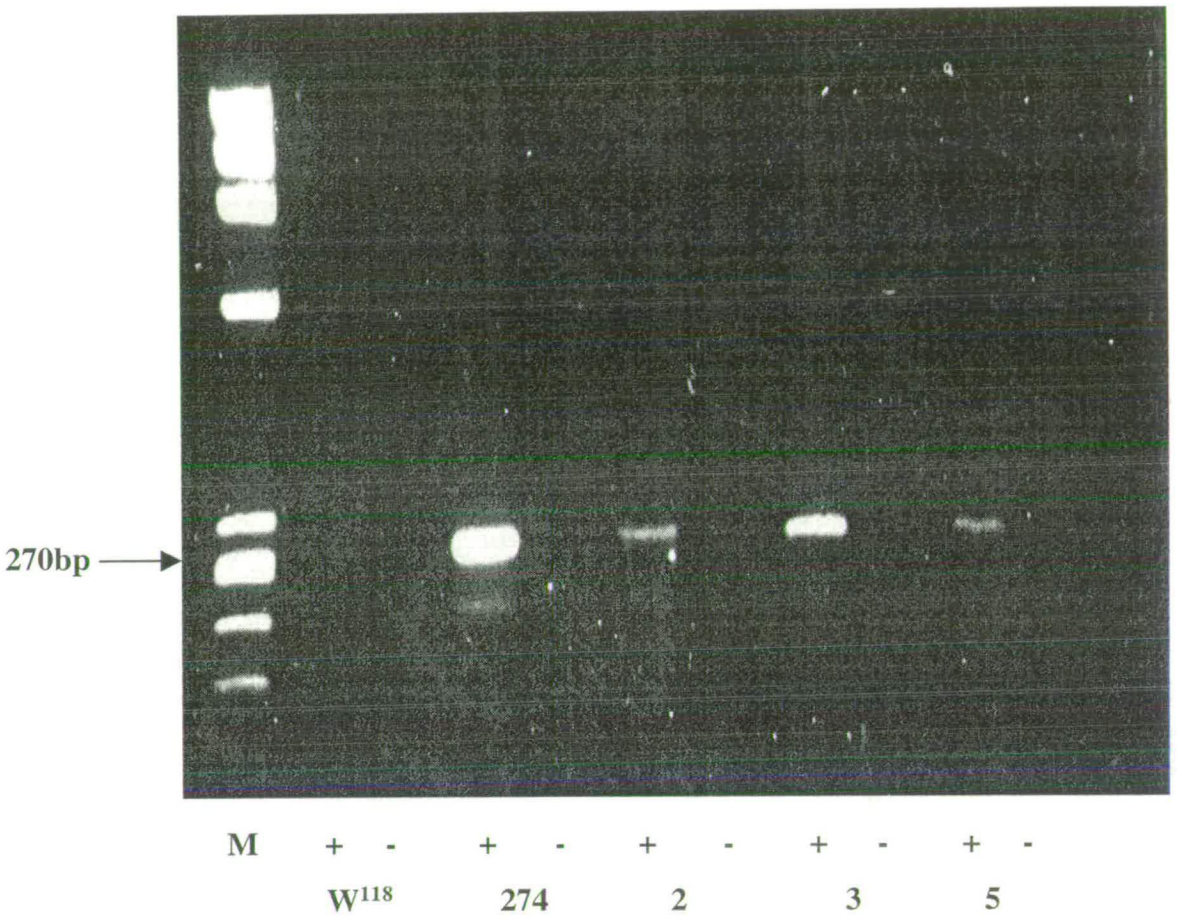


Figure 7.6 RT-PCR of *lacZ* mRNA in ovaries of flies transformed with P171.

RNAs isolated from ovaries were processed to obtain first-strand cDNA. These cDNAs were used as templates for PCR amplification of a 270 bp fragment from *lacZ* gene. w^{1118} is a negative control; 274 as positive control which contains the 41 to 186 sequences of the *I* factor upstream of *hsp70* promoter; 2 , 3, 5 contain products from the P171 transgenic lines 2, 3 and 5 respectively. +: with reverse transcriptase; -: without reverse transcriptase. M: ϕX174 DNA-*Hae*III marker.

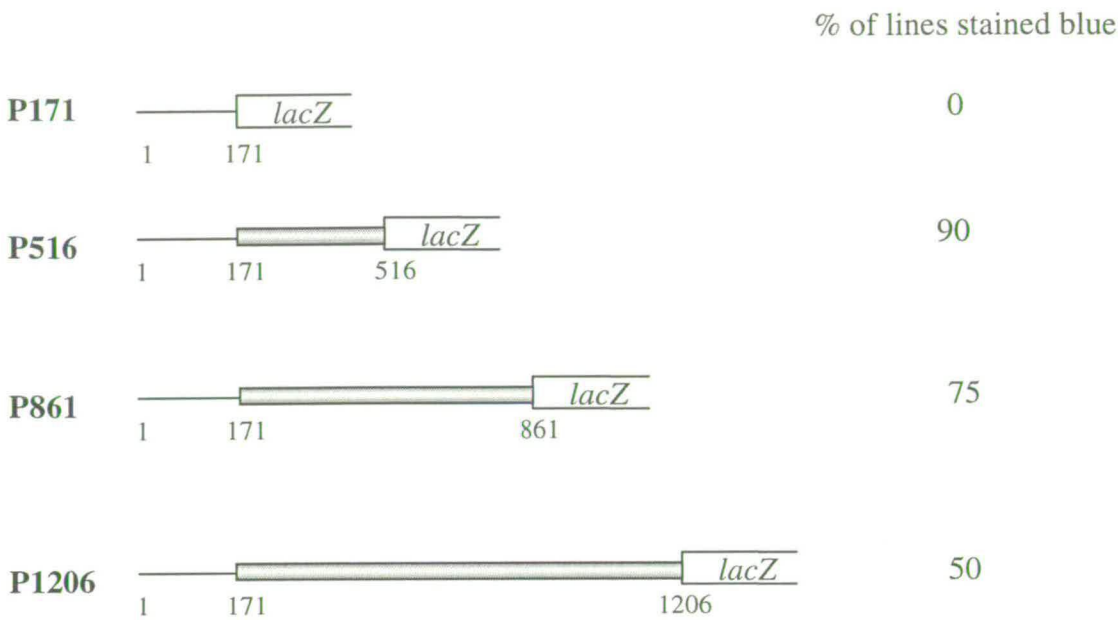
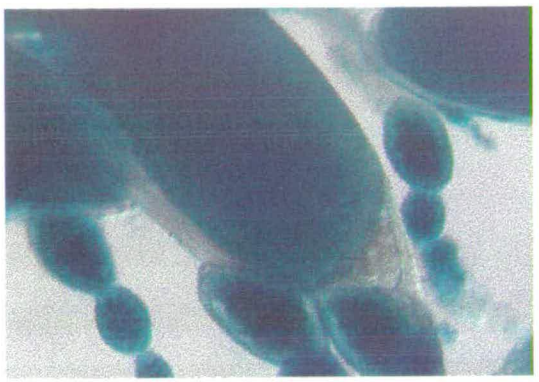
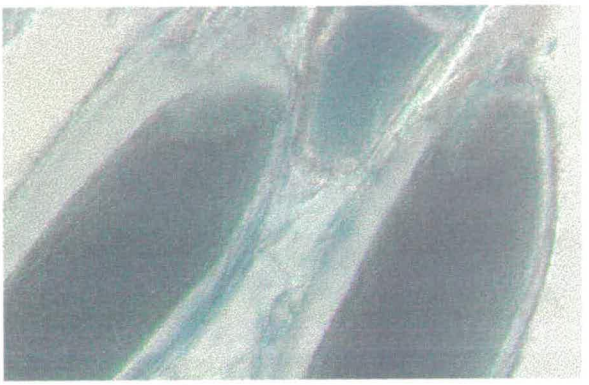


Figure 7.7 ß-galactosidase staining in ovaries of transgenic lines.

Maps of the four constructs used to demonstrate the promoter activity of *Mos1* is shown. The numbers indicate the nucleotide sites of Mos1. Nucleotides +1 to +171of *Mos1* were selected as a putative promoter region, as they make up the entire 5' non-coding region. The P1206 construct contains both 5' non-coding region and the intact ORF of *Mos1*. The P516 construct contains 5' non-coding region and 1/3 of the ORF of *Mos1*, while the P861 construct contains 5' non-coding region and 2/3 of the ORF of *Mos1*. These three DNA fragments were inserted in-frame with the *lacZ* gene which results in the expression of a ß-galactosidase fusion protein. Ovaries of females from lines transformed with those constructs were stained for ß-galactosidase activity. The results were expressed as the percentage of lines which carry the same construct stained blue.

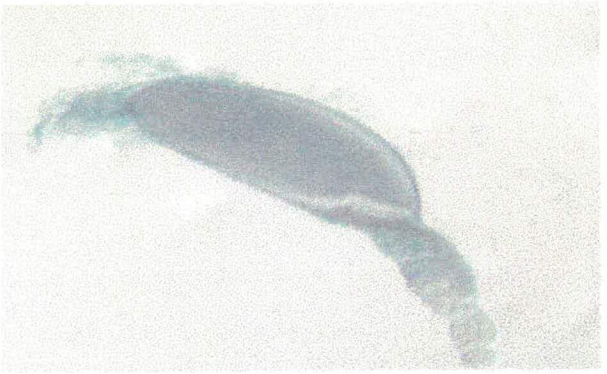


P516-line7 (++++)



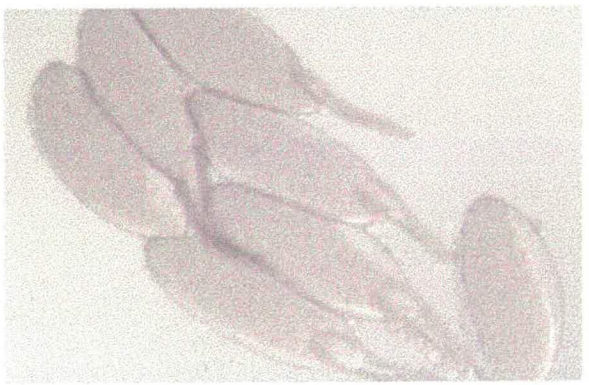
P516-line10 (+++)

P516-line3 (++)



P516-line5 (+)

P516-line4 (-)



w^{1118} (-)

Figure 7.8 Expression of ß-galactosidase in transgenic flies.

ß-galactosidase expression varies in the different lines of same construct, presumably as a result of position effect. *lacZ* expression in heterozygous rather than homozygous flies. w^{1118} as a negative control. (-) indicates no staining; (+): weak staining; (++): moderate staining; (+++): strong staining; (++++): very strong staining.

transformed with the pCaSpeR-ßgal vector whereas several lines of flies transformed with P516, P1206 and P861 show higher ß-galactosidase activity. Line 7 obtained from transformation with P516 expressed about 40 times more ß-galactosidase activity than that of control flies. Most of the lines obtained from P171 exhibited more or less similar levels of ß-galactosidase activity to that of the control (Table 7.2 and Figure 7.9). Statistical analysis was carried out to compare the ß-galactosidase activity between these constructs (Table 7.3) and the results showed there was no significant difference between them.

The P1206 construct encodes the full length transposase of Mos1. The decreased promoter activity of P1206 construct compared to P516 might be due to autorepression of Mos1 transposase on its own promoter. To test this hypothesis, flies homozygous for the P516 (Line1 or Line7) were crossed to flies homozygous for *P[ry⁺, hsp70:Mos1]* in which the *Mos1* coding region is driven by the *hsp70* heat-shock promoter (Lohe et al., 1995). This construct can function as a source of transposase to excise the *peach* element, yielding very high levels of somatic mosacism of w^{pch} even in the absence of heat shock. This phenomonon had been confirmed before the above crosses were carried out. The ovaries of female progeny were dissected and ß-galactosidase activity assay were performed. In each case, the activity obtained from these progeny is almost the same as that from their P516 parents (Table 7.4). The similar expression level of P516 construct in the presence or in the absence of wild-type transposase indicating the decreased promoter activity of P1206 construct is not due to autoregulatory inhibition of Mos1 transposase but most likely due to the inhibitory sequences present downstream of the nucleotide 516.

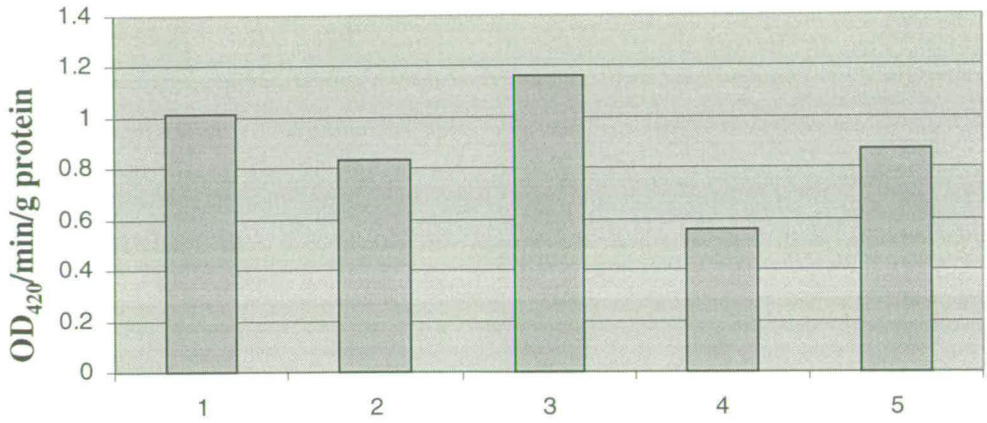
7.2.4 Expression of ß-galactosidase in tissue culture cells

Promoters inserted into plasmids pCaSpeR-ßgal or pCaSpeR-AUG-ßgal polylinker can be tested for *lacZ* expression in transfected *Drosophila* tissue culture cells. All four constructs (P171, P516, P861 and P1206) were introduced into Schneider line 2 *Drosophila* tissue culture cells by calcium phosphate transfection. After transfection, calcium phosphate solution was removed and the cells were washed and incubated for 2 days with fresh medium. 50 µl of resupended cells were diluted in 250 µl fresh medium and transferred into 24 well plates for ß-galactosidase staining. The cells were fixed by glutaraldehyde-formaldehyde for 10 minutes at room temperature and then stained in solution containing X-gal at 37°C for 2 hours. The plates were analysed under a microscope and the total stained cells were counted for each well. As expected, transfection with P171 resulted in

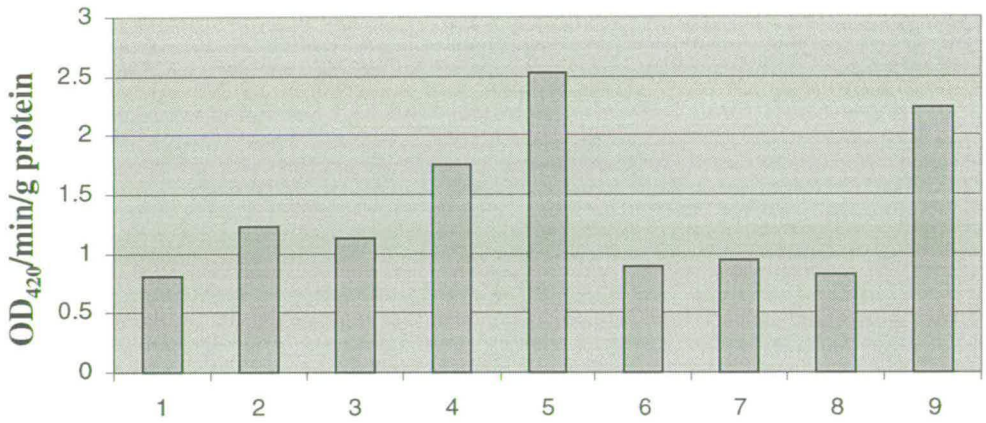
Table 7.2 The measurements of ß-galactosidase activity in ovaries from lines transformed with *Mos1* promoter fragments controlling expression of the *E.coli lacZ* gene.

Line	control	P171	P516	P861	P1206
1	1.02	0.83	13.67	1.22	1.76
2	0.83	1.23	8.76	1.86	11.26
3	1.16	1.12	7.54	3.67	14.9
4	0.56	1.75	1.10	8.63	0.57
5	0.87	2.54	1.44	0.93	1.73
6		0.89	1.54	1.80	0.48
7		0.94	36.24	3.46	0.56
8		0.83	0.18	3.98	
9		2.24	1.79		1.21
10			2.56		0.79
11			0.61		15.02
12					4.15

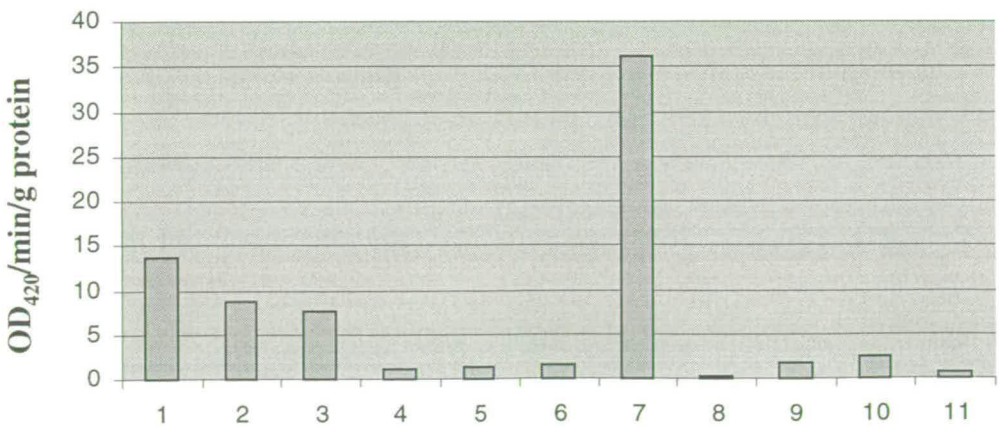
The slope of the line obtained by plotting OD_{420} against time was taken as the measure of activity after correction for the concentration of protein in the extract. The units are OD_{420}/min/g protein. Line 8 of P1206 was lost because of low viability. Line 1 of the control refers to w^{1118} flies. Lines 2-5 of the control refer to flies transformed with pCaSpeR- ßgal vector.



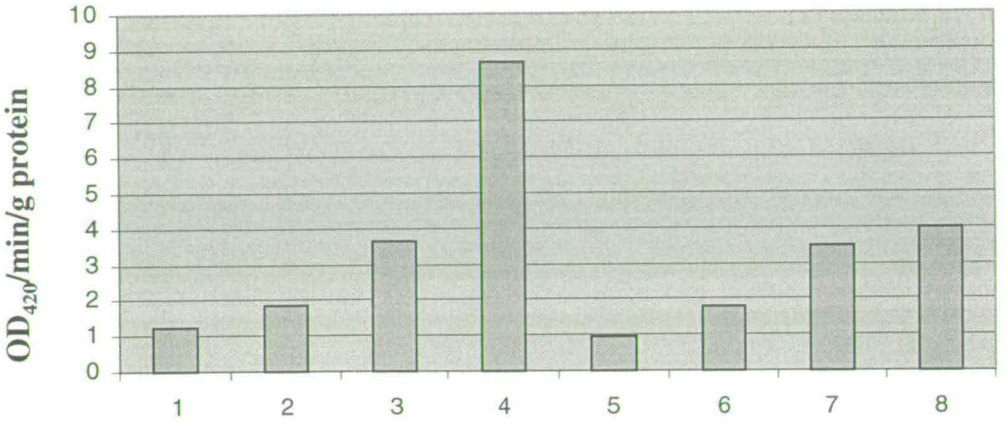
Control (0.888 ± 0.226)



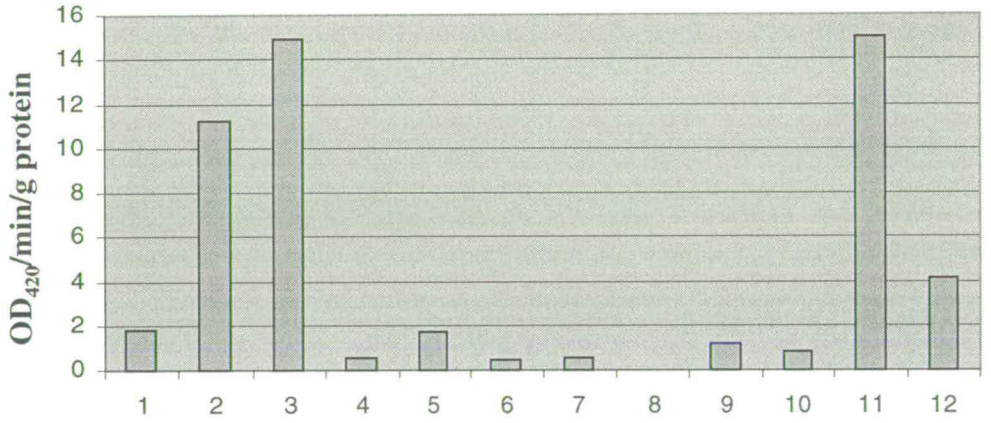
P171 (1.37 ± 0.648)



P516 (6.86 ± 10.6)



P861 (3.19 ± 2.49)



P1206 (4.77 ± 5.93)

Figure 7.9 The levels of ß-galactosidase activity in the ovaries of control and transgenic flies.

The activities were measured as a slope of the graphs plotted between OD420 and the time in the unit of OD_{420}/min/g protein. Data shown in brackets are the means ±SD of different lines of the same insertion. Line 8 from P1206 was lost because of low viability. Line 1 of control is referred to w^{1118} flies. Lines 2-5 of control are referred to flies transformed with pCaSpeR-ßgal vector.

Table 7.3 Calculation of *t* statistics for the pairwise comparison between different groups.

	P171	P516	P861	P1206
control	0.154	0.092	0.054	0.055
P171		0.116	0.074	0.084
P516			0.293	0.575
P861				0.442

The numbers indicate the *P* value calculated by Microsoft Excel Program. In each case, the *P* value exceeds 0.05 and therefore there is no significant difference between these groups.

Table 7.4 Comparisons of ß-galactosidase activity in ovary of P516 transgenic flies in the presence or in the absence of Mos1 transposase.

Line	P516(1)	P516(1)/*hsp:Mos1*	P516(7)	P516(7)/*hsp:Mos1*
Activity (OD_{420}/min/g protein)	13.67	20.02	36.24	27.28

Flies homozygous for the P516 (Line1 and Line7) were crossed to flies homozygous for *P*[*ry*$^+$, *hsp70:Mos1*] in which the *Mos1* coding region is driven by the *hsp70* heat-shock promoter. The ovaries of female progeny were dissected and ß-galactosidase activity assays were performed.

few blue cells whereas transfection of P516 generated many blue cells indicating efficient ß-galactosidase synthesis (Figure 7.10 and Table 7.5). These results obtained from transient expression in transfected *Drosophila* tissue culture cells are consistent with those observed in intact transformed flies.

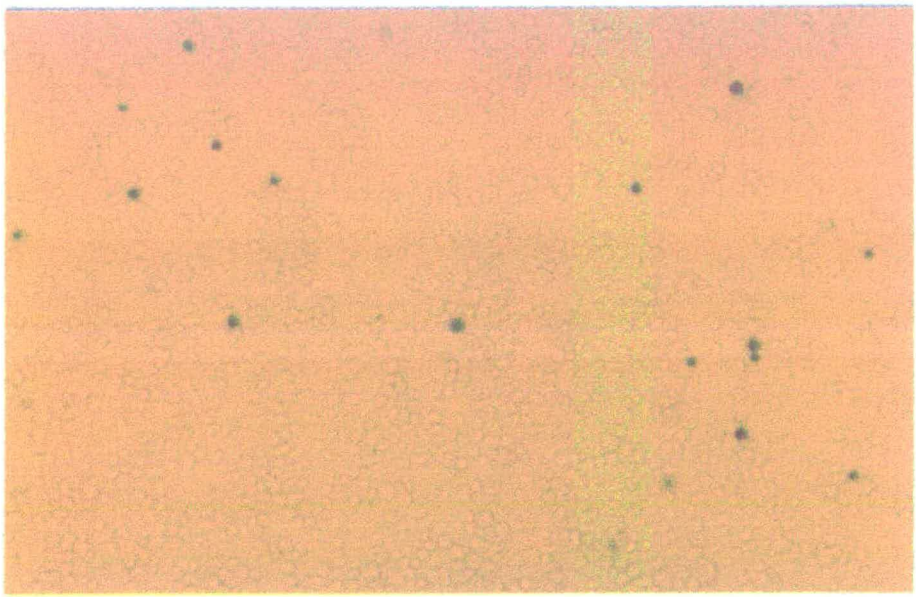
7.3 Discussion

The results from transgenic flies demonstrate the first 516 bp of *Mos1* contain the sequences required for high ß-galactosidase expression in the female germ line. The promoter lies within nucleotide 1-171 which shows low activity. There is enhancer-like element in the region of nucleotide 172-516 that stimulates ß-galactosidase expression.

P transposase binds specifically to a 10 bp DNA sequence that overlaps the TATA box of the *P* promoter (Kaufman et al., 1989). The full-length transposase can repress transcription in vitro by interfering with TFIID-TATA box interactions, thereby blocking the assembly of an RNA polymerase II transcription complex at the *P* element promoter (Kaufman and Rio, 1991). TnsB is a sequence-specific DNA binding protein that recognises multiple sites in both ends of the transposon. One of the TnsB binding sites overlaps a proposed promoter for the transposition genes of *Tn7* and may also regulate the expression of *Tn7*-encoded transposition genes (Arciszewska et al., 1991). Mos1 transposase is also a sequence-specific DNA binding protein that binds specifically to the inverted repeats of *Mos1* (Chapter 4). The autoregulatory inhibition of Mos1 transposase on its own promoter was tested in 7.2.3. However, the similar expression level of P516 construct in the presence or in the absence of wild-type transposase does not support such hypothesis. This result also suggests that the 'overproduction inhibition' may not act at the level of transcription through autorepression of the *Mos1* promoter.

The expression of a transgene is affected by its location in the genome as a result of local differences in chromatin structure, and in some cases the proximity of enhancer or repressor elements that regulate transcription (Spradling and Rubin, 1983; Levis et al., 1985). It thus is advisable to generate a number of lines containing different insertions. More than 8 independent lines have been obtained from each construct described in this Chapter, and the average activity of different lines would tend to minimise such position effect. Statistical analysis was carried out to compare the ß-galactosidase activity between these constructs. Since the standard deviation within each group is relatively high,

A



B

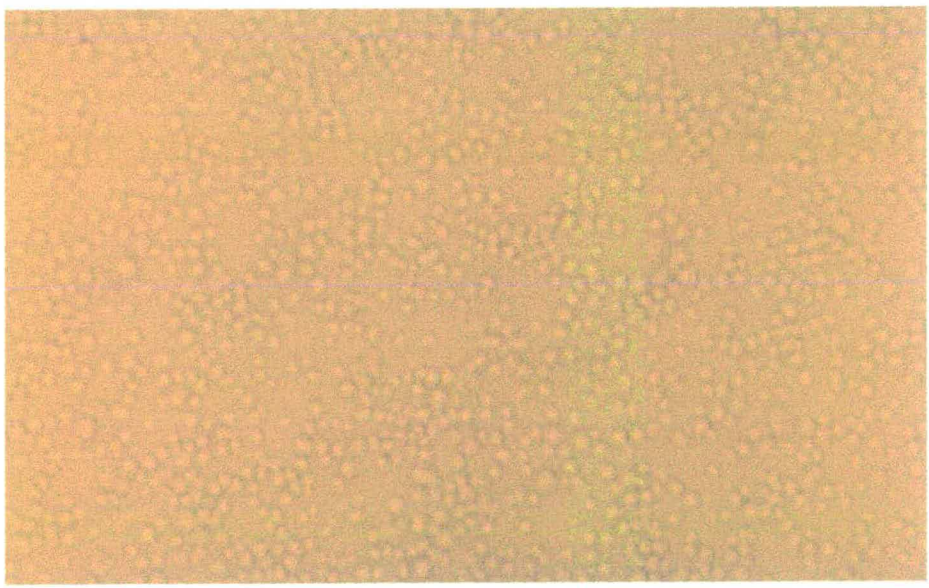


Figure 7.10 Staining for ß -galactosidase activity in transfected cells.

A. Schneider 2 cells were transfected with P516 and then incubated in staining solution containing X-gal at 37°C to allow the colour to develop. **B**. Schneider 2 cells were stained for ß -galactosidase activity as negative control.

Table 7.5 Number of cells stained blue in each well of 24 well plate after transfection with different constructs.

	A	B	C
P171	4	1	1
P516	350	350	320
P861	8	6	1
P1206	150	120	170

Three plasmid DNA purification methods were performed for each construct.

A, B, C refer to Bulk prep, Qiagen Miniprep and Qiagen Midiprep respectively.

probably due to position effects that alter the level of expression of *lacZ*, the *t* statistics for each of the pairwise comparisons show there is no significant difference between these groups. Dramatic position effects in *mariner* expression were also observed in a *mariner* element *Ma351*. When introduced into the germline of w^{pch} flies, *Ma351* yielded various levels of w^{pch} masaicism depending on insertion site. Another *mariner* element *Ma86d* identical in sequence to *Mos1* yielded lower levels of mosaicism implying that adjacent flanking sequences have important effects on *Mos1* activity (Medhora et al., 1991).

Chapter 8

General discussion and future work

Mariner is a member of the *Tc1/mariner* superfamily of eukaryotic transposons which transpose in a "cut-and-paste" manner via a DNA intermediate. Transposition occurs in a highly organised nucleoprotein synaptic complex or transpososome. The interaction of a transposase with the ends of its cognate transposable element is the initial step of transposition. Mos1 transposase recognises the inverted repeats at the ends of *Mos1* in a sequence specific manner via a DNA binding domain located within the N-terminal 120 amino acids. This binds the right hand inverted repeat about five times more strongly than the left end (Chapter 4). The differential binding pattern of Mos1 transposase on its two ends affects the cleavage reaction. As shown in Chapter 6, cleavage at the right end is about 10 times more efficient than at the left end of *Mos1*.

The ends of the transposable element are brought together through transposase-transposase interactions. The synapsis of two transposon ends is required before catalytic steps (strand nicking or strand transfer) occur. The role of protein-protein interactions in *Mos1* transposition has been studied in this thesis. Twelve single point mutations which reduce subunit interactions have been isolated by yeast two-hybrid system. The Mos1(L124S) mutant reduces transposase activity in both excision and transposition assays, indicating that subunit interactions are involved in the transposition reaction. A change in target site selection observed with this mutant may reflect a change in the conformation of the transpososome resulting from altered protein-protein interaction. These data provide useful information for the better understanding of the organisation of *Mos1* transpososome.

Regulation of transposition is a key issue in the application of *mariner*-like elements to methods of germline transformation. The results from the transgenic flies demonstrate the first 516 bp of *Mos1* are required to regulate the expression of this element. No effect of Mos1 transposase on the activity of its own promoter was observed. This suggests that "overproduction inhibition" may not act at the level of transcription through autorepression of the *Mos1* promoter. It may act at the post-translational level if an excess of subunits promotes the formation of catalytically inactive or less active oligomers. Further study of subunit interactions may help define the molecular basis of such mechanism.

8.1 Isolation of hyperactive transposase mutants

It has been proposed that some wild-type transposases have evolved to work at less than maximal activity to ensure that transposition occurs at a low frequency *in vivo*, and is therefore less deleterious to a host organism (Engels et al., 1987).

When wild-type Tn5 transposase was utilised in an *in vitro* system, occasional, barely detectable levels of transposition could be measured by a transformation assay. Its activity can be increased by mutations and Tn5 transposase containing three mutations (EK54/MA56/LP372) is hyperactive *in vitro* (Goryshin and Reznikoff, 1997). EK54 enhances outside end (OE) binding activity of transposase (Zhou and Reznikoff, 1997). MA56 blocks the synthesis of inhibitor protein (Inh) (Wiegand and Reznikoff, 1992). LP372 enhances transposase activity possibly by facilitating the conformational change required to remove self-inhibitory interactions between the NH_2- and COOH- terminal domains, thus allowing the formation of a synaptic complex (Davies et al., 2000). These three mutations are synergistic, creating an extraordinarily active transposase.

Although TnsA and TnsB together form a heteromeric Tn7 transposase, no transposition was observed in the presence of just TnsA and TnsB. The activation of the TnsAB transposase is modulated directly by TnsC. A papillation assay was employed to screen for both TnsA and TnsB "gain-of-function" mutations. TnsA mutants ($TnsA^{E185K}$ and $TnsA^{Q261E}$) when paired with TnsB mutants ($TnsB^{M366I}$ and $TnsB^{A325T}$) were able to promote transposition in the absence of TnsC, indicating that TnsA and TnsB contain all the essential activities to complete an intermolecular transposition event (Lu and Craig, 2000).

Two hyperactive transposase mutants of Himar1 has also been isolated (Lampe et al., 1999). The A7(H267R) mutant was about 10-fold more active in *E.coli* than the wild type whereas the C9(Q131R/E137K) mutant was 50-fold more active. Purified mutant transposases retain their hyperactivity, although to a lesser degree, in an *in vitro* transposition assay. The A7 mutant was 4.8-fold more active than the wild type whereas C9 was 7-fold more active.

In Chapter 5, we used yeast two-hybrid system to select those Mos1 mutants with impaired protein-protein interactions. Using the same method, we can also select those mutants with increased subunit interactions, simply plating the cells on high concentration of 3-AT. The maximum 3-AT concentration tolerated by yeast cells containing wild-type Mos1-Mos1 interaction is 20 mM. Preliminary experiments showed several yeast colonies containing

mutant Mos1 can grow on 100 mM 3-AT (Fig 8.1), indicating an increased interaction. After retransforming the mutant plasmids into yeast cell to confirm the interaction phenotype, we will do sequencing analysis to check which amino acids have been changed. If the mutations are located in neither the DNA binding nor the catalytic domain, the corresponding transposase would be suitable candidates to be purified and their activity to be studied in *in vitro* transposition assay. Since the increased transposase-transposase interaction may favour the formation of synaptic complex, both excision and transposition efficiencies may be improved.

A specific class of Tn10 transposase mutants successfully promote a higher than normal level of excision of this element from its donor site, but are blocked for target site interactions (Haniford et al., 1989). Another class of Tn10 transposase mutants confers relaxed insertion specificity but does not affect other aspects of transposition (Bender and Kleckner, 1992). The Mos1(L124S) with impaired protein-protein interactions also alters *Mos1* target site recognition (Chapter 6). It will also be interesting to know whether or not hyperactive Mos1 mutants alter target site recognition.

Mariner elements are becoming increasingly important tools for genetic analysis in a variety of different organisms. Hyperactive transposase mutants would be well suited for both *in vivo* and *in vitro* work and may significantly improve the efficiency of *Mos1*-derived elements as genetic tools.

8.2 Characterisation of *Mos1* pre-cleavage synaptic complex

Synaptic complexes are ubiquitous through all transposition and integration reactions (see introduction). However, the *Mos1* synaptic complex has not been characterised so far.

Incubation of purified Tn10 transposase and IHF with a short DNA fragment containing the outside end of *IS10* resulted in the formation of a stable protein-DNA complex that formed a discrete band in a gel retardation assay (Sakai et al., 1995). In reactions containing a mixture of two fragments differing 16 bp in length (S and L), three complexes were formed and migrated with significant different mobility. So the observed transposase-DNA complex consisted of a pair of transposon end fragments held together by proteins. This pre-cleavage "paired ends complex" (PEC) is an obligatory precursor to the subsequent cleavage and strand transfer steps.

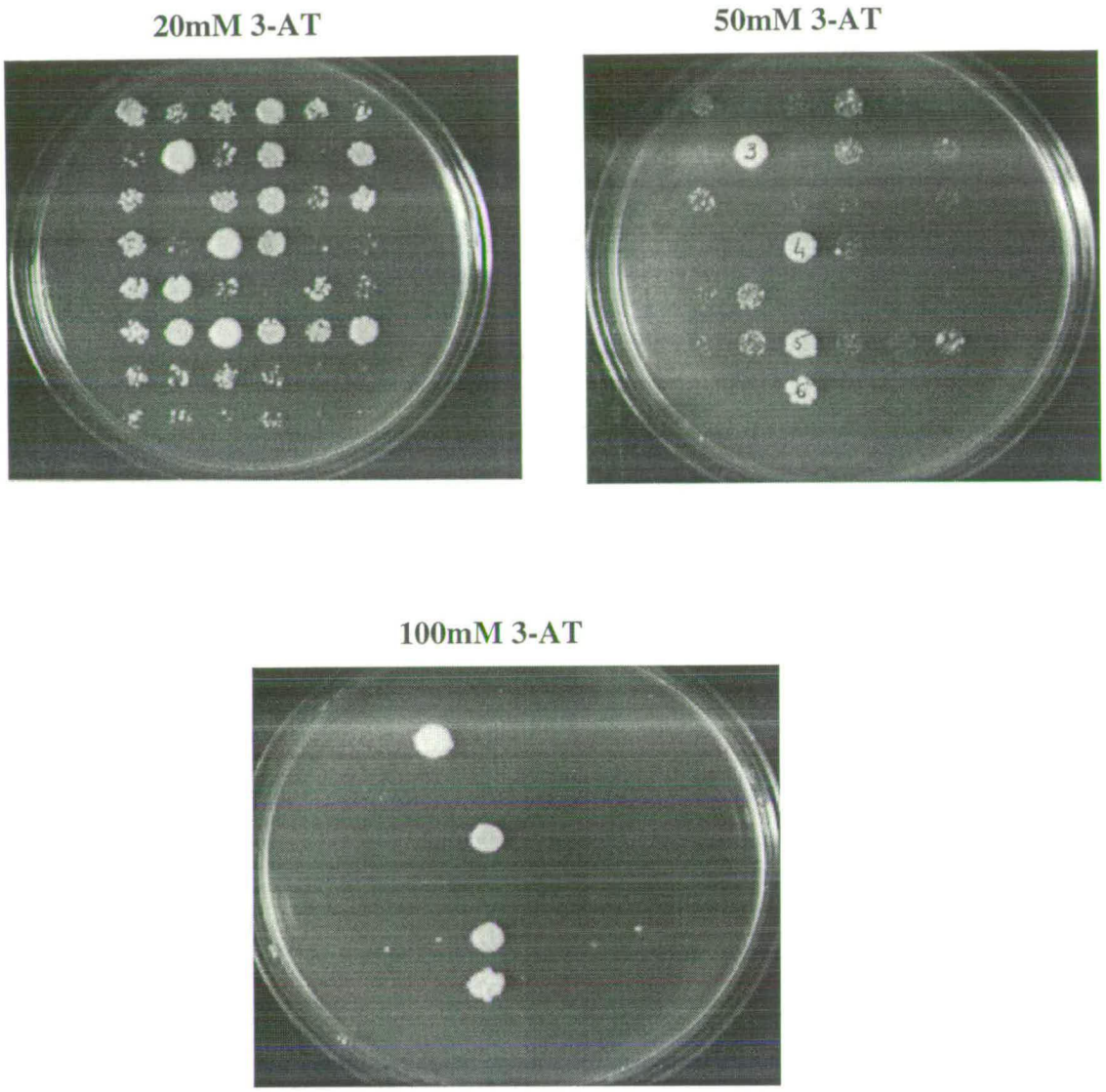


Figure 8.1 Growth phenotypes of hyperactive Mos1 mutant alleles.

The Mos1 coding region was mutagenized by the use of a PCR favouring single misincorporations. The PCR products were subsequently introduced into pACTIIst-Mos1(wt)-containing yeast cells (L40) by gap-repair transformation with a linearized pBTM116 plasmid. The transformants were streaked onto -LWH plates containing 20mM, 50 mM or 100mM 3-AT to test for the ability to form colonies. The plates were photographed after incubation at 30°C for 5 days. Wild-type Mos1-Mos1 interaction induces *HIS3* expression that allows growth at 3-AT concentration limited to 20 mM (colony at the top-left), while hyperactive Mos1 mutants allow growth at 3-AT level up to 100mM.

The *Tn5* pre-cleavage synaptic complex has been identified in a gel retardation assay recently (Bhasin et al., 2000). The outside end (OE) and the inside end (IE) of *IS50* differ by seven bases. A hyperactive, hybrid end sequence, the mosaic end (ME) was radio-labelled and incubated with EK54/LP372 Tn5 transposase. After electrophoresing the binding reactions through native polyacrylamide gels, a single retarded band representing the *Tn5* paired ends complex was seen. Using ME DNAs of different lengths or transposase of different molecular masses, the author have shown the stoichiometry of the *Tn5* PEC to be two DNA molecules and two transposase molecules.

We can perform a similar gel retardation assay in which we mix both 191bp left end and 176 bp right end probe of *Mos1* with full-length wild-type Mos1 transposase. We would expect three retarded bands on the gel: the fastest of the three containing two right ends, the slowest band containing two long DNA fragments (left end), and the middle band containing both left and right end fragments. Then we can do the same experiment with Mos1(L124S) mutant instead of wild-type transposase. Since this mutant is defective in protein-protein interactions, and therefore unfavourable for the formation of synaptic complex, we would expect PEC bands will form only very inefficiently. On the other hand, such synaptic complex may form more efficiently in the presence of the hyperactive mutants described in 8.1. These experiments will provide direct and strong evidence that subunit interactions are involved in formation of the synaptic complex.

8.3 Structure study of Mos1 transposase

Over the last 20 years, remarkable progress has been made in the molecular analysis of transposition through the use of genetics and biochemistry. Structural studies have focused on isolated domains rather than intact proteins. The three-dimensional structure of the catalytic core domain of Mu transposase was first characterised (Rice and Mizuuchi, 1995) and a co-crystal structure of the DNA-binding domain of Tc3 transposase in complex with double-stranded DNA that corresponds to the termini of *Tc3* transposon has been solved (van Pouderoyen et al., 1997). This showed an HTH fold, forming a dimer that brings together the two DNA ends. The DNA-binding domain of Mos1 transposase has been identified within the first 120 amino acids of the protein and contains a HTH motif (Chapter 4). Sufficient soluble Mos1-N120 protein can be obtained from pET His-Tag system (Chapter 3) for structural analysis by NMR (nuclear magnetic resonance). Preliminary NMR experiments indicate that Mos1-N120 forms a monomer rather than a

dimer in the absence of DNA (Richardson, personal communication). This is consistent with our yeast two-hybrid results (Chapter 5). The HTH motif is necessary but not sufficient for sequence-specific DNA binding and the structural studies should provide further insight into the overall folding of the intact DNA binding domain revealing other regions required for DNA binding. We are especially interested to know why the first 30 amino acids are not dispensable. Do they help the correct folding of the HTH motif or are they directly involved in DNA binding ?

The first structure of a transposase synaptic complex has been determined (Davies et al., 2000). This structure consists of intact Tn5 transposase bound to *Tn5* transposon ends. The molecular assembly of *Tn5* synaptic complex is dimeric, where each transposon ends is bound by both transposase subunits, orienting the transposon cleavage sites into the transposase catalytic sites. To fully understand the molecular basis of *Mos1* transposition, the structure of the *Mos1* synaptic complex is eagerly awaited. With a detailed architecture of a transposase-DNA complex available, we should be able to answer the following questions: 1) How does an intact transposase bind to the ends of its cognate DNA? 2) What is the interaction between the transposon cleavage site and the transposase active site? 3) How are divalent metal ion cofactors bound to the active site? The structure may also indicate whether or not the catalytic domains of Mos1 also cut *in trans*, thus ensuring that co-ordinated cleavage of the two transposon ends takes place. Those mutants affecting Mos1-Mos1 interactions (Chapter 5) can also be understood in terms of structural basis.

The crystal structure of neither Mos1 nor Himar1 transposase is currently available because of the poor solubility of the full length protein. Mos1(L124S) mutant is more soluble than wild-type Mos1 transposase, indicating that mutants may be uncovered which are amenable to crystallisation and structure study.

References

Andrews, J. D., and Gloor, G. B. (1995). A role for the KP leucine zipper in regulating P element transposition in Drosophila melanogaster. Genetics *141*, 587-94.

Arciszewska, L. K., McKown, R. L., and Craig, N. L. (1991). Purification of TnsB, a transposition protein that binds to the ends of Tn7. J Biol Chem *266*, 21736-44.

Ariyoshi, M., Vassylyev, D. G., Iwasaki, H., Nakamura, H., Shinagawa, H., and Morikawa, K. (1994). Atomic structure of the RuvC resolvase: a holliday junction-specific endonuclease from E. coli. Cell *78*, 1063-72.

Atkinson, P. W., Warren, W. D., and O'Brochta, D. A. (1993). The hobo transposable element of Drosophila can be cross-mobilized in houseflies and excises like the Ac element of maize. Proc Natl Acad Sci U S A *90*, 9693-7.

Auge-Gouillou, C., Bigot, Y., Pollet, N., Hamelin, M. H., Meunier-Rotival, M., and Periquet, G. (1995). Human and other mammalian genomes contain transposons of the mariner family. FEBS Lett *368*, 541-6.

Bainton, R., Gamas, P., and Craig, N. L. (1991). Tn7 transposition in vitro proceeds through an excised transposon intermediate generated by staggered breaks in DNA. Cell *65*, 805-16.

Bainton, R. J., Kubo, K. M., Feng, J. N., and Craig, N. L. (1993). Tn7 transposition: target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified in vitro system. Cell *72*, 931-43.

Baker, T. A., Kremenstova, E., and Luo, L. (1994). Complete transposition requires four active monomers in the mu transposase tetramer. Genes Dev *8*, 2416-28.

Baker, T. A., and Luo, L. (1994). Identification of residues in the Mu transposase essential for catalysis. Proc Natl Acad Sci U S A *91*, 6654-8.

Baker, T. A., Mizuuchi, M., and Mizuuchi, K. (1991). MuB protein allosterically activates strand transfer by the transposase of phage Mu. Cell *65*, 1003-13.

Barth, P. T., Datta, N., Hedges, R. W., and Grinter, N. J. (1976). Transposition of a deoxyribonucleic acid sequence encoding trimethoprim and streptomycin resistances from R483 to other replicons. J Bacteriol *125*, 800-10.

Beall, E. L., Admon, A., and Rio, D. C. (1994). A Drosophila protein homologous to the human p70 Ku autoimmune antigen interacts with the P transposable element inverted repeats. Proc Natl Acad Sci U S A *91*, 12681-5.

Bender, J., and Kleckner, N. (1992). IS10 transposase mutations that specifically alter target site recognition. Embo J *11*, 741-50.

Berg, D. E., and Howe, M. M. (1989). Mobile DNA (Washington D.C.: American Society for Microbiology).

Bhasin, A., Goryshin, I. Y., and Reznikoff, W. S. (1999). Hairpin formation in Tn5 transposition. J Biol Chem *274*, 37021-9.

Bhasin, A., Goryshin, I. Y., Steiniger-White, M., York, D., and Reznikoff, W. S. (2000). Characterization of a Tn5 pre-cleavage synaptic complex. J Mol Biol *302*, 49-63.

Black, D. M., Jackson, M. S., Kidwell, M. G., and Dover, G. A. (1987). KP elements repress P-induced hybrid dysgenesis in Drosophila melanogaster. Embo J *6*, 4125-35.

Boeke, D. J. (1989). Transposable elements in S.cerevisiae. In Mobile DNA, D. E. Berg and M. M. Howe, eds. (Washington, D.C.: American Society for Microbiology), pp. 335-374.

Bolland, S., and Kleckner, N. (1996). The three chemical steps of Tn10/IS10 transposition involve repeated utilization of a single active site. Cell *84*, 223-33.

Bolland, S., and Kleckner, N. (1995). The two single-strand cleavages at each end of Tn10 occur in a specific order during transposition. Proc Natl Acad Sci U S A *92*, 7814-8.

Braam, L. A., Goryshin, I. Y., and Reznikoff, W. S. (1999). A mechanism for Tn5 inhibition. carboxyl-terminal dimerization. J Biol Chem *274*, 86-92.

Braam, L. A., and Reznikoff, W. S. (1998). Functional characterization of the Tn5 transposase by limited proteolysis. J Biol Chem *273*, 10908-13.

Breathnach, R., and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Annu Rev Biochem *50*, 349-83.

Brennan, R. G., and Matthews, B. W. (1989). The helix-turn-helix DNA binding motif. J Biol Chem *264*, 1903-6.

Brunet, F., Godin, F., Bazin, C., and Capy, P. (1999). Phylogenetic analysis of Mos1-like transposable elements in the Drosophilidae. J Mol Evol *49*, 760-8.

Brunet, F., Godin, F., Bazin, C., David, J. R., and Capy, P. (1996). The mariner transposable element in natural populations of Drosophila teissieri. J Mol Evol *42*, 669-75.

Brunet, F., Godin, F., David, J. R., and Capy, P. (1994). The mariner transposable element in the Drosophilidae family. Heredity *73*, 377-85.

Bryan, G., Garza, D., and Hartl, D. (1990). Insertion and excision of the transposable element mariner in Drosophila. Genetics *125*, 103-14.

Bryan, G. J., and Hartl, D. L. (1988). Maternally inherited transposon excision in Drosophila simulans. Science *240*, 215-7.

Bryan, G. J., Jacobson, J. W., and Hartl, D. L. (1987). Heritable somatic excision of a Drosophila transposon. Science *235*, 1636-8.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A., and Skalka, A. M. (1996). The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations. Structure *4*, 89-96.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A., and Skalka, A. M. (1995). High-resolution structure of the catalytic domain of avian sarcoma virus integrase. J Mol Biol *253*, 333-46.

Burns, N. R., Saibil, H. R., White, N. S., Pardon, J. F., Timmins, P. A., Richardson, S. M., Richards, B. M., Adams, S. E., Kingsman, S. M., and Kingsman, A. J. (1992). Symmetry, flexibility and permeability in the structure of yeast retrotransposon virus-like particles. Embo J *11*, 1155-64.

Capy, P., Koga, A., David, J. R., and Hartl, D. L. (1992). Sequence analysis of active mariner elements in natural populations of Drosophila simulans. Genetics *130*, 499-506.

Capy, P., Langin, T., Bigot, Y., Brunet, F., Daboussi, M. J., Periquet, G., David, J. R., and Hartl, D. L. (1994). Horizontal transmission versus ancient origin: mariner in the witness box. Genetica *93*, 161-70.

Capy, P., Maruyama, K., David, J. R., and Hartl, D. L. (1991). Insertion sites of the transposable element mariner are fixed in the genome of Drosophila sechellia. J Mol Evol *33*, 450-6.

Capy, P., Vitalis, R., Langin, T., Higuet, D., and Bazin, C. (1996). Relationships between transposable elements based upon the integrase-transposase domains: is there a common ancestor? J Mol Evol *42*, 359-68.

Chaconas, G., Lavoie, B. D., and Watson, M. A. (1996). DNA transposition: jumping gene machine, some assembly required. Curr Biol *6*, 817-20.

Clubb, R. T., Schumacher, S., Mizuuchi, K., Gronenborn, A. M., and Clore, G. M. (1997). Solution structure of the I gamma subdomain of the Mu end DNA-binding domain of phage Mu transposase. J Mol Biol *273*, 19-25.

Coates, C. J., Jasinskiene, N., Miyashiro, L., and James, A. A. (1998). Mariner transposition and transformation of the yellow fever mosquito, Aedes aegypti. Proc Natl Acad Sci U S A *95*, 3748-51.

Coates, C. J., Turney, C. L., Frommer, M., O'Brochta, D. A., and Atkinson, P. W. (1997). Interplasmid transposition of the mariner transposable element in non-drosophilid insects. Mol Gen Genet *253*, 728-33.

Coates, C. J., Turney, C. L., Frommer, M., O'Brochta, D. A., Warren, W. D., and Atkinson, P. W. (1995). The transposable element mariner can excise in non-drosophilid insects. Mol Gen Genet *249*, 246-52.

Colloms, S. D., van Luenen, H. G., and Plasterk, R. H. (1994). DNA binding activities of the Caenorhabditis elegans Tc3 transposase. Nucleic Acids Res *22*, 5548-54.

Craig, N. L. (1997). Target site selection in transposition. Annu Rev Biochem *66*, 437-74.

Craig, N. L. (2000). Tn7: a smarter transposon. In Keystone symposia: Transposition and other genome rearrangements, pp. 26.

Craig, N. L. (1996). Transposon Tn7. Curr Top Microbiol Immunol *204*, 27-48.

Craig, N. L. (1995). Unity in transposition reactions. Science *270*, 253-4.

Craigie, R., and 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.

Curcio, M. J., and Belfort, M. (1996). Retrohoming: cDNA-mediated mobility of group II introns requires a catalytic RNA. Cell *84*, 9-12.

Davies, D. R., Braam, L. M., Reznikoff, W. S., and Rayment, I. (1999). The three-dimensional structure of a Tn5 transposase-related protein determined to 2.9-A resolution. J Biol Chem *274*, 11904-13.

Davies, D. R., Goryshin, I. Y., Reznikoff, W. S., and Rayment, I. (2000). Three-dimensional structure of the Tn5 synaptic complex transposition intermediate [see comments]. Science *289*, 77-85.

de la Cruz, N. B., Weinreich, M. D., Wiegand, T. W., Krebs, M. P., and Reznikoff, W. S. (1993). Characterization of the Tn5 transposase and inhibitor proteins: a model for the inhibition of transposition. J Bacteriol *175*, 6932-8.

Derbyshire, K. M., Hwang, L., and Grindley, N. D. (1987). Genetic analysis of the interaction of the insertion sequence IS903 transposase with its terminal inverted repeats. Proc Natl Acad Sci U S A *84*, 8049-53.

Doak, T. G., Doerder, F. P., Jahn, C. L., and Herrick, G. (1994). A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. Proc Natl Acad Sci U S A *91*, 942-6.

Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R., and Davies, D. R. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases [see comments]. Science *266*, 1981-6.

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

Engels, W. R. (1989). P elements in *Drosophila melanogaster*. In Mobile DNA, D. E. Berg and M. M. Howe, eds. (Washington D.C.: American Society for Microbiology), pp. 437-484.

Engels, W. R., Benz, W. K., Preston, C. R., Graham, P. L., Phillis, R. W., and Robertson, H. M. (1987). Somatic effects of P element activity in Drosophila melanogaster: pupal lethality. Genetics *117*, 745-57.

Evans, J. P., and Palmiter, R. D. (1991). Retrotransposition of a mouse L1 element. Proc Natl Acad Sci U S A *88*, 8792-5.

Fadool, J. M., Hartl, D. L., and Dowling, J. E. (1998). Transposition of the mariner element from Drosophila mauritiana in zebrafish. Proc Natl Acad Sci U S A *95*, 5182-6.

Fawcett, D. H., Lister, C. K., Kellett, E., and Finnegan, D. J. (1986). Transposable elements controlling I-R hybrid dysgenesis in D. melanogaster are similar to mammalian LINEs. Cell *47*, 1007-15.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245-6.

Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. Trends Genet *5*, 103-7.

Finnegan, D. J. (1992). Transposable elements. Curr Opin Genet Dev *2*, 861-7.

Finnegan, D. J. (1997). Transposable elements: how non-LTR retrotransposons do it. Curr Biol *7*, R245-8.

Fujiwara, T., and Craigie, R. (1989). Integration of mini-retroviral DNA: a cell-free reaction for biochemical analysis of retroviral integration. Proc Natl Acad Sci U S A *86*, 3065-9.

Fujiwara, T., and Mizuuchi, K. (1988). Retroviral DNA integration: structure of an integration intermediate. Cell *54*, 497-504.

Garcia-Fernandez, J., Bayascas-Ramirez, J. R., Marfany, G., Munoz-Marmol, A. M., Casali, A., Baguna, J., and Salo, E. (1995). High copy number of highly similar mariner-like transposons in planarian (Platyhelminthe): evidence for a trans-phyla horizontal transfer. Mol Biol Evol *12*, 421-31.

Garcia-Fernandez, J., Marfany, G., Baguna, J., and Salo, E. (1993). Infiltration of mariner elements [letter; comment]. Nature *364*, 109-10.

Goldgur, Y., Dyda, F., Hickman, A. B., Jenkins, T. M., Craigie, R., and Davies, D. R. (1998). Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. Proc Natl Acad Sci U S A *95*, 9150-4.

Goryshin, I., Kil, Y. V., and Reznikoff, W. S. (1994). DNA length, bending, and twisting constraints on IS50 transposition. Proc Natl Acad Sci U S A *91*, 10834-8.

Goryshin, I. Y., and Reznikoff, W. S. (1998). Tn5 in vitro transposition. J Biol Chem *273*, 7367-74.

Grenier, E., Abadon, M., Brunet, F., Capy, P., and Abad, P. (1999). A mariner-like transposable element in the insect parasite nematode Heterorhabditis bacteriophora. J Mol Evol *48*, 328-36.

Grindley, N. D., and Leschziner, A. E. (1995). DNA transposition: from a black box to a color monitor. Cell *83*, 1063-6.

Gueiros-Filho, F. J., and Beverley, S. M. (1997). Trans-kingdom transposition of the Drosophila element mariner within the protozoan Leishmania [see comments] [published erratum appears in Science 1997 Aug 8;277(5327):753]. Science *276*, 1716-9.

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. J Mol Biol *166*, 557-80.

Haniford, D., and Kleckner, N. (1994). Tn 10 transposition in vivo: temporal separation of cleavages at the two transposon ends and roles of terminal basepairs subsequent to interaction of ends. Embo J *13*, 3401-11.

Haniford, D. B., Chelouche, A. R., and Kleckner, N. (1989). A specific class of IS10 transposase mutants are blocked for target site interactions and promote formation of an excised transposon fragment. Cell *59*, 385-94.

Haren, L., Ton-Hoang, B., and Chandler, M. (1999). Integrating DNA: transposases and retroviral integrases. Annu Rev Microbiol *53*, 245-81.

Harrison, S. C. (1991). A structural taxonomy of DNA-binding domains. Nature *353*, 715-9.

Hartl, D. L. (1989). Transposable element mariner in Drosophila species. In Mobile DNA, D. E. Berg and M. M. Howe, eds. (Washington D.C.: American Society for Microbiology), pp. 531-536.

Haymer, D. S., and Marsh, J. L. (1986). Germ line and somatic instability of a white mutation in Drosophila mauritiana due to a transposable genetic element. Dev Genet *6*, 281-91.

Hazelrigg, T., Levis, R., and Rubin, G. M. (1984). Transformation of white locus DNA in drosophila: dosage compensation, zeste interaction, and position effects. Cell *36*, 469-81.

Heinlein, M., Brattig, T., and Kunze, R. (1994). In vivo aggregation of maize Activator (Ac) transposase in nuclei of maize endosperm and Petunia protoplasts. Plant J *5*, 705-14.

Hengen, P. N. (1997). False positives from the yeast two-hybrid system. Trends Biochem Sci *22*, 33-4.

Hickman, A. B., Li, Y., Mathew, S. V., May, E. W., Craig, N. L., and Dyda, F. (2000). Unexpected structural diversity in DNA recombination: the restriction endonuclease connection. Mol Cell *5*, 1025-34.

Hollenberg, S. M., Sternglanz, R., Cheng, P. F., and Weintraub, H. (1995). Identification of a new family of tissue-specific basic helix-loop-helix proteins with a two-hybrid system. Mol Cell Biol *15*, 3813-22.

Holzinger, A., Philips, K. S., and Weaver, T. E. (1996). Single-step purification/solubilization of 6XHis-tagged proteins on Ni-NTA agarose. QIAGEN NEWS *4*, 14-15.

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

Jacobson, J. W., and 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., and Hartl, D. L. (1986). Molecular structure of a somatically unstable transposable element in Drosophila. Proc Natl Acad Sci U S A *83*, 8684-8.

Janknecht, R., de Martynoff, G., Lou, J., Hipskind, R. A., Nordheim, A., and Stunnenberg, H. G. (1991). Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus. Proc Natl Acad Sci U S A *88*, 8972-6.

Johnson, R. C., Yin, J. C., and Reznikoff, W. S. (1982). Control of Tn5 transposition in Escherichia coli is mediated by protein from the right repeat. Cell *30*, 873-82.

Junop, M. S., and Haniford, D. B. (1997). Factors responsible for target site selection in Tn10 transposition: a role for the DDE motif in target DNA capture. Embo J *16*, 2646-55.

Junop, M. S., and Haniford, D. B. (1996). Multiple roles for divalent metal ions in DNA transposition: distinct stages of Tn10 transposition have different Mg^{2+} requirements. Embo J *15*, 2547-55.

Kachroo, P., Leong, S. A., and Chattoo, B. B. (1994). Pot2, an inverted repeat transposon from the rice blast fungus Magnaporthe grisea. Mol Gen Genet *245*, 339-48.

Karess, R. E. (1985). P element mediated germline transformation of Drosophila. In DNA cloning Volume II, a practical approach, D. M. Glover, ed. (Oxford : IRS Press).

Karess, R. E., and Rubin, G. M. (1984). Analysis of P transposable element functions in Drosophila. Cell *38*, 135-46.

Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T., and Morikawa, K. (1990). Three-dimensional structure of ribonuclease H from E. coli. Nature *347*, 306-9.

Katzman, M., Katz, R. A., Skalka, A. M., and Leis, J. (1989). The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration. J Virol *63*, 5319-27.

Kaufman, P. D., Doll, R. F., and Rio, D. C. (1989). Drosophila P element transposase recognizes internal P element DNA sequences. Cell *59*, 359-71.

Kaufman, P. D., and Rio, D. C. (1991). Drosophila P-element transposase is a transcriptional repressor in vitro. Proc Natl Acad Sci U S A *88*, 2613-7.

Kaufman, P. D., and Rio, D. C. (1992). P element transposition in vitro proceeds by a cut-and-paste mechanism and uses.GTP as a cofactor. Cell *69*, 27-39.

Kenna, M. A., Brachmann, C. B., Devine, S. E., and Boeke, J. D. (1998). Invading the yeast nucleus: a nuclear localization signal at the C terminus of Ty1 integrase is required for transposition in vivo. Mol Cell Biol *18*, 1115-24.

Kennedy, A. K., Guhathakurta, A., Kleckner, N., and Haniford, D. B. (1998). Tn10 transposition via a DNA hairpin intermediate. Cell *95*, 125-34.

Ketting, R. F., Fischer, S. E. J., and Plasterk, R. H. (1997). Target choice determinants of the Tc1 transposon of Caenorhabditis elegans. Nucleic Acids Res *25*, 4041-7.

Kidwell, M. G. (1993). Evolutionary biology. Voyage of an ancient mariner [news; comment]. Nature *362*, 202.

Kim, A., Terzian, C., Santamaria, P., Pelisson, A., Purd'homme, N., and Bucheton, A. (1994). Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of Drosophila melanogaster. Proc Natl Acad Sci U S A *91*, 1285-9.

Kimura, M., and Yamaguchi, I. (1998). Convergent transcription units and their promoters at both ends of pot2, an inverted repeat transposon from the rice blast fungus. J Biochem (Tokyo) *124*, 268-73.

Kleckner, N., Chalmers, R. M., Kwon, D., Sakai, J., and Bolland, S. (1996). Tn10 and IS10 transposition and chromosome rearrangements: mechanism and regulation in vivo and in vitro. Curr Top Microbiol Immunol *204*, 49-82.

Kruklitis, R., Welty, D. J., and Nakai, H. (1996). ClpX protein of Escherichia coli activates bacteriophage Mu transposase in the strand transfer complex for initiation of Mu DNA synthesis. Embo J *15*, 935-44.

Kubo, K. M., and Craig, N. L. (1990). Bacterial transposon Tn7 utilizes two different classes of target sites. J Bacteriol *172*, 2774-8.

Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P., and Skalka, A. M. (1992). Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol Cell Biol *12*, 2331-8.

Kunze, R., Behrens, U., Courage-Franzkowiak, U., Feldmar, S., Kuhn, S., and Lutticke, R. (1993). Dominant transposition-deficient mutants of maize Activator (Ac) transposase. Proc Natl Acad Sci U S A *90*, 7094-8.

Lampe, D. J., Akerley, B. J., Rubin, E. J., Mekalanos, J. J., and Robertson, H. M. (1999). Hyperactive transposase mutants of the Himar1 mariner transposon. Proc Natl Acad Sci U S A *96*, 11428-33.

Lampe, D. J., Churchill, M. E., and Robertson, H. M. (1996). A purified mariner transposase is sufficient to mediate transposition in vitro [published erratum appears in EMBO J 1997 Jul 1;16(13):4153]. Embo J *15*, 5470-9.

Lampe, D. J., Grant, T. E., and Robertson, H. M. (1998). Factors affecting transposition of the Himar1 mariner transposon in vitro. Genetics *149*, 179-87.

Lavoie, B. D., and Chaconas, G. (1996). Transposition of phage Mu DNA. Curr Top Microbiol Immunol *204*, 83-102.

Lee, C. C., Beall, E. L., and Rio, D. C. (1998). DNA binding by the KP repressor protein inhibits P-element transposase activity in vitro. Embo J *17*, 4166-74.

Lee, C. C., Mul, Y. M., and Rio, D. C. (1996). The Drosophila P-element KP repressor protein dimerizes and interacts with multiple sites on P-element DNA. Mol Cell Biol *16*, 5616-22.

Leung, D. W., Chen, E., and Goeddel, D. V. (1989). A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction technique. J Methods Cell Mol Biol *1*, 11-15.

Levis, R., Hazelrigg, T., and Rubin, G. M. (1985). Effects of genomic position on the expression of transduced copies of the white gene of Drosophila. Science *229*, 558-61.

Li, B., and Fields, S. (1993). Identification of mutations in p53 that affect its binding to SV40 large T antigen by using the yeast two-hybrid system. Faseb J *7*, 957-63.

Lidholm, D. A., Lohe, A. R., and Hartl, D. L. (1993). The transposable element mariner mediates germline transformation in Drosophila melanogaster. Genetics *134*, 859-68.

Lohe, A. R., De Aguiar, D., and Hartl, D. L. (1997). Mutations in the mariner transposase: the D,D(35)E consensus sequence is nonfunctional. Proc Natl Acad Sci U S A *94*, 1293-7.

Lohe, A. R., and Hartl, D. L. (1996a). Autoregulation of mariner transposase activity by overproduction and dominant-negative complementation. Mol Biol Evol *13*, 549-55.

Lohe, A. R., and Hartl, D. L. (1996b). Reduced germline mobility of a mariner vector containing exogenous DNA: effect of size or site? Genetics *143*, 1299-306.

Lohe, A. R., Lidholm, D. A., and Hartl, D. L. (1995). Genotypic effects, maternal effects and grand-maternal effects of immobilized derivatives of the transposable element mariner. Genetics *140*, 183-92.

Lohe, A. R., Sullivan, D. T., and Hartl, D. L. (1996). Subunit interactions in the mariner transposase. Genetics *144*, 1087-95.

Lu, F., and Craig, N. L. (2000). Isolation and characterization of Tn7 transposase gain-of-function mutants: a model for transposase activation. Embo J *19*, 3446-57.

Luan, D. D., Korman, M. H., Jakubczak, J. L., and Eickbush, T. H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell *72*, 595-605.

Luo, G., Ivics, Z., Izsvak, Z., and Bradley, A. (1998). Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. Proc Natl Acad Sci U S A *95*, 10769-73.

Manning, J. E., Schmid, C. W., and Davidson, N. (1975). Interspersion of repetitive and nonrepetitive DNA sequences in the Drosophila melanogaster genome. Cell *4*, 141-55.

Maruyama, K., and Hartl, D. L. (1991a). Evolution of the transposable element mariner in Drosophila species. Genetics *128*, 319-29.

Maruyama, K., and Hartl, D. L. (1991b). Evidence for interspecific transfer of the transposable element mariner between Drosophila and Zaprionus. J Mol Evol *33*, 514-24.

May, E. W., and Craig, N. L. (1996). Switching from cut-and-paste to replicative Tn7 transposition. Science *272*, 401-4.

McClintock, B. (1957). Controling elements and the gene. Cold Spring Harbour Symp. Quant. Biol. *21*, 197-216.

Medhora, M., Maruyama, K., and Hartl, D. L. (1991). Molecular and functional analysis of the mariner mutator element Mos1 in Drosophila. Genetics *128*, 311-8.

Mizuuchi, K. (1997). Polynucleotidyl transfer reactions in site-specific DNA recombination. Genes Cells *2*, 1-12.

Mizuuchi, K. (1992). Transpositional recombination: mechanistic insights from studies of mu and other elements. Annu Rev Biochem *61*, 1011-51.

Mizuuchi, M., Baker, T. A., and Mizuuchi, K. (1992). Assembly of the active form of the transposase-Mu DNA complex: a critical control point in Mu transposition. Cell *70*, 303-11.

Mizuuchi, M., Baker, T. A., and Mizuuchi, K. (1991). DNase protection analysis of the stable synaptic complexes involved in Mu transposition. Proc Natl Acad Sci U S A *88*, 9031-5.

Moreira, L. A., Edwards, M. J., Adhami, F., Jasinskiene, N., James, A. A., and Jacobs-Lorena, M. (2000). Robust gut-specific gene expression in transgenic aedes aegypti mosquitoes. Proc Natl Acad Sci U S A *97*, 10895-8.

Morisato, D., and Kleckner, N. (1984). Transposase promotes double strand breaks and single strand joints at Tn10 termini in vivo. Cell *39*, 181-90.

Mullins, M. C., Rio, D. C., and Rubin, G. M. (1989). cis-acting DNA sequence requirements for P-element transposition. Genes Dev *3*, 729-38.

O'Hare, K., and Rubin, G. M. (1983). Structures of P transposable elements and their sites of insertion and excision in the Drosophila melanogaster genome. Cell *34*, 25-35.

O'Neill, M., Dryden, D. T., and Murray, N. E. (1998). Localization of a protein-DNA interface by random mutagenesis. Embo J *17*, 7118-27.

Oosumi, T., Belknap, W. R., and Garlick, B. (1995). Mariner transposons in humans [letter]. Nature *378*, 672.

Patience, C., Wilkinson, D. A., and Weiss, R. A. (1997). Our retroviral heritage. Trends Genet *13*, 116-20.

Phadnis, S. H., and Berg, D. E. (1987). Identification of base pairs in the outside end of insertion sequence IS50 that are needed for IS50 and Tn5 transposition. Proc Natl Acad Sci U S A *84*, 9118-22.

Pietrokovski, S., and Henikoff, S. (1997). A helix-turn-helix DNA-binding motif predicted for transposases of DNA transposons. Mol Gen Genet *254*, 689-95.

Plasterk, R. H. (1996). The Tc1/mariner transposon family. Curr Top Microbiol Immunol *204*, 125-43.

Plasterk, R. H., Izsvak, Z., and Ivics, Z. (1999). Resident aliens: the Tc1/mariner superfamily of transposable elements. Trends Genet *15*, 326-32.

Polard, P., and Chandler, M. (1995). Bacterial transposases and retroviral integrases. Mol Micro *15*, 13-23.

Reiter, L. T., Liehr, T., Rautenstrauss, B., Robertson, H. M., and Lupski, J. R. (1999). Localization of mariner DNA transposons in the human genome by PRINS. Genome Res *9*, 839-43.

Reiter, L. T., Murakami, T., Koeuth, T., Pentao, L., Muzny, D. M., Gibbs, R. A., and Lupski, J. R. (1996). A recombination hotspot responsible for two inherited peripheral neuropathies is located near a mariner transposon-like element [published erratum appears in Nat Genet 1998 Jul;19(3):303]. Nat Genet *12*, 288-97.

Reznikoff, W. S. (1993). The Tn5 transposon. Annu Rev Microbiol *47*, 945-63.

Rice, P., Craigie, R., and Davies, D. R. (1996). Retroviral integrases and their cousins. Curr Opin Struct Biol *6*, 76-83.

Rice, P., and Mizuuchi, K. (1995). Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. Cell *82*, 209-20.

Rio, D. C., Barnes, G., Laski, F. A., Rine, J., and Rubin, G. M. (1988). Evidence for Drosophila P element transposase activity in mammalian cells and yeast. J Mol Biol *200*, 411-5.

Robertson, H. M. (1993). The mariner transposable element is widespread in insects [see comments]. Nature *362*, 241-5.

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

Robertson, H. M., and Asplund, M. L. (1996). Bmmar1: a basal lineage of the mariner family of transposable elements in the silkworm moth, Bombyx mori. Insect Biochem Mol Biol *26*, 945-54.

Robertson, H. M., and Lampe, D. J. (1995). Distribution of transposable elements in arthropods. Annu Rev Entomol *40*, 333-57.

Robertson, H. M., and Lampe, D. J. (1995). Recent horizontal transfer of a mariner transposable element among and between Diptera and Neuroptera. Mol Biol Evol *12*, 850-62.

Robertson, H. M., and MacLeod, E. G. (1993). Five major subfamilies of mariner transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. Insect Mol Biol *2*, 125-39.

Rogers, M., Ekaterinaki, N., Nimmo, E., and Sherratt, D. (1986). Analysis of Tn7 transposition. Mol Gen Genet *205*, 550-6.

Rowland, S. J., Sherratt, D. J., Stark, W. M., and Boocock, M. R. (1995). Tn552 transposase purification and in vitro activities. Embo J *14*, 196-205.

Rubin, G. M., and Spradling, A. C. (1982). Genetic transformation of Drosophila with transposable element vectors. Science *218*, 348-53.

Sakai, J., Chalmers, R. M., and Kleckner, N. (1995). Identification and characterization of a pre-cleavage synaptic complex that is an early intermediate in Tn10 transposition. Embo J *14*, 4374-83.

Sakai, J., and Kleckner, N. (1997). The Tn10 synaptic complex can capture a target DNA only after transposon excision. Cell *89*, 205-14.

Sakai, J. S., Kleckner, N., Yang, X., and Guhathakurta, A. (2000). Tn10 transpososome assembly involves a folded intermediate that must be unfolded for target capture and strand transfer. Embo J *19*, 776-85.

Sarnovsky, R. J., May, E. W., and Craig, N. L. (1996). The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. Embo J *15*, 6348-61.

Schlappi, M., Raina, R., and Fedoroff, N. (1994). Epigenetic regulation of the maize Spm transposable element: novel activation of a methylated promoter by TnpA. Cell *77*, 427-37.

Schouten, G. J., van Luenen, H. G., Verra, N. C., Valerio, D., and Plasterk, R. H. (1998). Transposon Tc1 of the nematode Caenorhabditis elegans jumps in human cells. Nucleic Acids Res *26*, 3013-7.

Schumacher, S., Clubb, R. T., Cai, M., Mizuuchi, K., Clore, G. M., and Gronenborn, A. M. (1997). Solution structure of the Mu end DNA-binding ibeta subdomain of phage Mu transposase: modular DNA recognition by two tethered domains. Embo J *16*, 7532-41.

Scofield, S. R., English, J. J., and Jones, J. D. (1993). High level expression of the Activator transposase gene inhibits the excision of Dissociation in tobacco cotyledons. Cell *75*, 507-17.

Sedensky, M. M., Hudson, S. J., Everson, B., and Morgan, P. G. (1994). Identification of a mariner-like repetitive sequence in C. elegans. Nucleic Acids Res *22*, 1719-23.

Sharratt, D. J. (1995). Mobile Genetic Elements (Oxford: Oxford University Press).

Sherman, A., Dawson, A., Mather, C., Gilhooley, H., Li, Y., Mitchell, R., Finnegan, D., and Sang, H. (1998). Transposition of the Drosophila element mariner into the chicken germ line [see comments] [published erratum appears in Nat Biotechnol 1999 Jan;17(1):81]. Nat Biotechnol *16*, 1050-3.

Simmons, M. J., and Bucholz, L. M. (1985). Transposase titration in Drosophila melanogaster: a model of cytotype in the P-M system of hybrid dysgenesis. Proc Natl Acad Sci U S A *82*, 8119-23.

Smit, A. F., and Riggs, A. D. (1996). Tiggers and DNA transposon fossils in the human genome. Proc Natl Acad Sci U S A *93*, 1443-8.

Smith, J. (1997). Mariner--Analysis of a Drosophila transposable element. (PhD thesis, University of Edinburgh).

Song, S. U., Gerasimova, T., Kurkulos, M., Boeke, J. D., and Corces, V. G. (1994). An env-like protein encoded by a Drosophila retroelement: evidence that gypsy is an infectious retrovirus. Genes Dev *8*, 2046-57.

Spee, J. H., de Vos, W. M., and Kuipers, O. P. (1993). Efficient random mutagenesis method with adjustable mutation frequency by use of PCR and dITP. Nucleic Acids Res *21*, 777-8.

Spradling, A. C., and Rubin, G. M. (1983). The effect of chromosomal position on the expression of the Drosophila xanthine dehydrogenase gene. Cell *34*, 47-57.

Steiniger-White, M., and Reznikoff, W. S. (2000). The C-terminal alpha helix of Tn5 transposase is required for synaptic complex formation. J Biol Chem *275*, 23127-33.

Surette, M. G., Buch, S. J., and Chaconas, G. (1987). Transpososomes: stable protein-DNA complexes involved in the in vitro transposition of bacteriophage Mu DNA. Cell *49*, 253-62.

Suzuki, M., and Brenner, S. E. (1995). Classification of multi-helical DNA-binding domains and application to predict the DBD structures of sigma factor, LysR, OmpR/PhoB, CENP-B, Rapl, and Xy1S/Ada/AraC [published erratum appears in FEBS Lett 1996 Mar 4;381(3):270]. FEBS Lett *372*, 215-21.

Thummel, C. S., Boulet, A. M., and Lipshitz, H. D. (1988). Vectors for Drosophila P-element-mediated transformation and tissue culture transfection. Gene *74*, 445-56.

Tosi, L. R., and Beverley, S. M. (2000). cis and trans factors affecting Mos1 mariner evolution and transposition in vitro, and its potential for functional genomics. Nucleic Acids Res *28*, 784-90.

Trentmann, S. M., Saedler, H., and Gierl, A. (1993). The transposable element En/Spm-encoded TNPA protein contains a DNA binding and a dimerization domain. Mol Gen Genet *238*, 201-8.

Tudor, M., Lobocka, M., Goodell, M., Pettitt, J., and O'Hare, K. (1992). The pogo transposable element family of Drosophila melanogaster. Mol Gen Genet *232*, 126-34.

van Gent, D. C., Ramsden, D. A., and Gellert, M. (1996). The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. Cell *85*, 107-13.

van Luenen, H. G., Colloms, S. D., and Plasterk, R. H. (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., and 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*, 6044-54.

Velterop, J. S., Dijkhuizen, M. A., van 't Hof, R., and Postma, P. W. (1995). A versatile vector for controlled expression of genes in Escherichia coli and Salmonella typhimurium. Gene *153*, 63-5.

Vidal, M., and Legrain, P. (1999). Yeast forward and reverse 'n'-hybrid system. Nucleic Acids Res *27*, 919-29.

Vos, J. C., and Plasterk, R. H. (1994). Tc1 transposase of Caenorhabditis elegans is an endonuclease with a bipartite DNA binding domain. Embo J *13*, 6125-32.

Waddell, C. S., and Craig, N. L. (1988). Tn7 transposition: two transposition pathways directed by five Tn7-encoded genes. Genes Dev *2*, 137-49.

Wang, H., Hartswood, E., and Finnegan, D. J. (1999). Pogo transposase contains a putative helix-turn-helix DNA binding domain that recognises a 12 bp sequence within the terminal inverted repeats. Nucleic Acids Res *27*, 455-61.

Warren, A. M., and Crampton, J. M. (1994). Mariner: Its prospects as a DNA vector for the genetic manipulation of medically important insects. Parasitology Today *10*, 58-63.

Watson, M. A., and Chaconas, G. (1996). Three-site synapsis during Mu DNA transposition: a critical intermediate preceding engagement of the active site. Cell *85*, 435-45.

Weinreich, M. D., Mahnke-Braam, L., and Reznikoff, W. S. (1994). A functional analysis of the Tn5 transposase. Identification of domains required for DNA binding and multimerization. J Mol Biol *241*, 166-77.

White, M. A. (1996). The yeast two-hybrid system: forward and reverse [comment]. Proc Natl Acad Sci U S A *93*, 10001-3.

Wiegand, T. W., and Reznikoff, W. S. (1992). Characterization of two hypertransposing Tn5 mutants. J Bacteriol *174*, 1229-39.

Wiley, L. J., Riley, L. G., Sangster, N. C., and Weiss, A. S. (1997). mle-1, a mariner-like transposable element in the nematode Trichostrongylus colubriformis. Gene *188*, 235-7.

Williams, T. L., and Baker, T. A. (2000). Molecular biology. Transposase team puts a headlock on DNA [see comments]. Science *289*, 73-4.

Williams, T. L., Jackson, E. L., Carritte, A., and Baker, T. A. (1999). Organization and dynamics of the Mu transpososome: recombination by communication between two active sites. Genes Dev *13*, 2725-37.

Wintjens, R., and Rooman, M. (1996). Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. J Mol Biol *262*, 294-313.

Yant, S. R., Meuse, L., Chiu, W., Ivics, Z., Izsvak, Z., and Kay, M. A. (2000). Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. Nat Genet *25*, 35-41.

Yip, T. T., and Hutchens, T. W. (1994). Immobilized metal ion affinity chromatography. Mol Biotechnol *1*, 151-64.

York, D., and Reznikoff, W. S. (1996). Purification and biochemical analyses of a monomeric form of Tn5 transposase. Nucleic Acids Res *24*, 3790-6.

Yoshioka, K., Kanda, H., Akiba, H., Enoki, M., and Shiba, T. (1991). Identification of an unusual structure in the Drosophila melanogaster transposable element copia: evidence for copia transposition through an RNA intermediate. Gene *103*, 179-84.

Young, M. W. (1979). Middle repetitive DNA: a fluid component of the Drosophila genome. Proc Natl Acad Sci U S A *76*, 6274-8.

Zang, X., Yazdanbakhsh, M., Jiang, H., Kanost, M. R., and Maizels, R. M. (1999). A novel serpin expressed by blood-borne microfilariae of the parasitic nematode Brugia malayi inhibits human neutrophil serine proteinases. Blood *94*, 1418-28.

Zhang, L., Sankar, U., Lampe, D. J., Robertson, H. M., and Graham, F. L. (1998). The Himar1 mariner transposase cloned in a recombinant adenovirus vector is functional in mammalian cells. Nucleic Acids Res *26*, 3687-93.

Zhou, M., and Reznikoff, W. S. (1997). Tn5 transposase mutants that alter DNA binding specificity. J Mol Biol *271*, 362-73.

Zhou, Y. H., Zhang, X. P., and Ebright, R. H. (1991). Random mutagenesis of gene-sized DNA molecules by use of PCR with Taq DNA polymerase. Nucleic Acids Res *19*, 6052.