EDINBURGH
UNIVERSITY
LIBRARY

Shelf Mark Darwin Library
Smith PhD 1999

Mariner - Analysis of a *Drosophila* Transposable Element

Jill Smith

A thesis presented for the degree of PhD

University of Edinburgh

1997

EX BIBL. UNIV. E...

I dedicate this thesis to my family

Acknowledgements

I would like to thank my supervisor, Professor David Finnegan for giving me the opportunity to carry out this research, and for the advice and encouragement he has given me during the course of my PhD.

I would also like to thank members of the Finnegan group, both past and present who have helped me over the years. These people are Angela, Apinunt, Eve, Hong Mei, Ivan, Julie, Laura, Maki, Petra, Ruth and Susan. I would like to thank Angela in particular for her optimism when things weren't going well, for the many helpful discussions, for allowing me to mention her unpublished results, and for teaching me how to use hyphens! This PhD would not have been possible without you.

I am also grateful to Dr. David Leach for acting as my second supervisor. Other people I would like to mention are Jules, Lynn, Mary and John from the sixth floor and all the friendly folk from the Beggs lab.

My thanks also extend to Joan and Jean for supplying me with clean glass ware, to Frances for providing fly food, to photography for prompt slides and pictures and to Ian Gilbert for his computing heroics.

I would like to mention some great friends that I have made from ICMB, who shared the ups and downs of PhD life with me. These people are Angela Dawson (again), Andrew Mayes, Angela McCahill (formerly McCormack) and Kerrie Tosh. Andrew, especially, thanks for giving me your love and support (also, all those Indian takeaways, and the bad jokes that kept me sane!). I would also like to acknowledge other friends in Edinburgh and those further afield at the end of a telephone line.

Finally, I would like to thank my parents Joan and Ray, my brother Stephen, and my Grandma, Elsie for their encouragement, love and support at all times.

Abbreviations

A	Adenosine
Amp	ampicillin
ATP	adenosine 5'-triphosphate
bp	base pair
BSA	bovine serum albumin
C	cytosine
°C	degrees celsius
cDNA	complementary DNA
Ci	curie
cm	centimetre
d	deoxy
dd	dideoxy
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxyribonucleoside triphosphate
DTT	dithiothreitol
EDTA	ethylene diaminetetraacetic acid
g	gram
G	guanosine
GST	glutathione S-transferase
HEPES	N-[2-Hydroxyethyl]piperazine-N'-[2-ethane-sulphonic acid]
HTH	helix-turn-helix
IPTG	isopropyl-β-D-thiogalactopyranoside
kb	kilobase
kD	kiloDalton
l	litre
L-Broth	Luria broth
LINE	long interspersed nuclear element
LTR	long terminal repeat
M	molar
mg	milligram
ml	millilitre
mM	millimolar
mol	mole
MOPS	3-[n-morpholino]propanesulphonic acid
mRNA	messenger ribonucleic acid

MW	molecular weight
ng	nanogram
NP40	Nonidet-P 40
O.D.	optical density
ORF	open reading frame
OPI	over production inhibition
oz	ounce
p	plasmid
p	pico
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
POD	peroxidase
PVDF	polyvinylidene fluoride
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
RT	reverse transcriptase
ry	the *rosy* gene from *D. melanogaster*
SDS	sodium dodecyl sulphate
SINE	short intersperseed nuclear element
T	thymidine
Tris	tris (hydroxymethyl) aminomethane
u	unit
μg	microgram
μl	microlitre
UV	ultraviolet
V	volt
v	volume
v	the *vermillion* gene from *D. melanogaster*
v/v	volume per volume
w	the *white* gene from *D. melanogaster*
w^{pch}	*white peach* allele
w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indol-β-galactopyranoside
y	the *yellow* gene from *D. melanogaster*

Abstract

The *mariner* transposable element *mos*1, isolated from *Drosophila mauritiana* belongs to a family of inverted repeat DNA transposons, found in many phyla. Members of this family of elements transpose by a 'cut and paste' mechanism, in which the element is excised from its position in the genome and inserted elsewhere. *Mos*1 is 1286bp long, and contains a single open reading frame of 345 amino acids. This ORF encodes a transposase which is essential for the transposition mechanism.

The existence of various forms of extrachromosomal copies of transposons has been documented for a number of elements and it has been assumed that some of these are intermediates in transposition. Active *mariner* elements exist in the genome of *D. simulans*. Several forms of free copies of *mariner* have been isolated from DNA prepared from this species. Some of these are circular in conformation. Sequence analysis, performed to determine the structures across junctions of the circular copies, has revealed that all elements are deleted for some *mariner* nucleotides. The majority of elements missing three bases from one or the other end.

Recombinant *mos*1 transposase purified from *E. coli* is able to stimulate excision of *mos*1 elements from plasmids. These excised copies are linear in structure. Primer extension analysis has been employed to investigate the terminal most nucleotides of these elements. The excision of *mos*1 occurs, leaving the terminal most three base pairs from each 5' strand behind in the plasmid. No circular copies were detected in this reaction. An explanation for these findings is discussed and a model for transposition of *mos*1 and *mariner* elements from *D. simulans* is proposed.

The biological activity of purified *mos*1 transposase has been tested. Co-injection of transposase and a *white* marked *mos*1 element into *white*$^-$ *Drosophila* embryos was performed in experiments based on the *P* element paradigm of germline transformation. Several fertile G_0 flies were found to exhibit a *white*$^+$ eye colour phenotype, presumably due to transposition events. This method has implications for the germline transformation of other organisms outside of *Drosophila*. Experiments investigating the activity of the inverted repeat DNA transposable element *Tc3*, from *C. elegans* in the genome of *D. melanogaster* are also described.

Table of Contents

Title.......... i
Dedication.......... ii
Declaration.......... iii
Acknowledgements.......... iv
Abbreviations.......... v
Abstract.......... vii
Table of Contents.......... ix

Chapter 1 - Introduction.......... 1

1.1 Introduction to transposable elements.......... 2
1.1.1 Classification of eukaryotic transposable elements.......... 2
1.1.1.1 Class I elements.......... 2
(i) LTR elements.......... 4
(ii) LINEs.......... 4
(iii) SINES.......... 5
1.1.1.2 Class II elements.......... 5

1.2 Introduction to *mariner*.......... 7
1.2.1 Isolation of the *mariner* element.......... 7
1.2.1.1 Instability of the w^{pch} allele.......... 7
1.2.1.2 Identification of *mos* factors.......... 8
1.2.1.3 Structure of *mos*1.......... 8
1.2.2 Distribution of *mariner* elements.......... 9
1.2.2.1 Distribution of *mariner* elements in the melanogaster species subgroup.......... 9
1.2.2.2 Widespread distribution of *mariner* elements.......... 11
1.2.2.3 Distribution of other class II elements.......... 12
1.2.2.4 Superfamily of transposable elements.......... 14
1.2.3 Autonomous and non-autonomous elements.......... 15
1.2.3.1 Non-autonomous *mariner* elements contain nucleotide substitutions.......... 15
1.2.3.2 Non-autonomous copies of other class II transposable elements.......... 16

1.3 The transposition reaction........ 16
1.3.1 Recognition of DNA binding sites by transposase........ 16
1.3.2 DNA binding domains of transposases........ 18
1.3.3 Cleavage and strand transfer........ 20
1.3.3.1 Retroviral integration........ 21
1.3.3.2 Bacteriophage Mu replication........ 21
1.3.3.3 'Cut and paste'........ 21
1.3.3.4 Transposition of *Tc*3........ 23
1.3.3.5 V(D)J recombination: a related reaction........ 25
1.3.3.6 The catalytic mechanism........ 25
(i) One active site for two strand transfer reactions: integrase........ 27
(ii) Two active sites for three strand transfer reactions: Tn7........ 27
(iii) One active site for three strand transfer reactions: Tn10........ 28
1.3.3.7 Higher order nucleoprotein structures........ 28
(i) Mu........ 28
(ii) Tn10........ 29
(iii) Tn7........ 29
1.3.4 Insertion site specificity........ 30
1.3.5 Donor site repair after excision........ 31

1.4 Regulation of transposition........ 32
1.4.1 Tissue specificity of transposition........ 33
1.4.2 Regulation of *P* element transposition........ 34
1.4.3 Regulation of *mariner* activity........ 37

1.5 Transposable elements as experimental tools........ 38
1.5.1 Use of the *P* element........ 39
1.5.2 Harnessing the *mariner* element........ 40

1.6 Scope of Thesis........ 42

Chapter 2 - Materials and Methods........ 43

2.1 Media........ 44
2.1.1 Bacterial media........ 44

2.1.2 *Drosophila* media 44

2.2 Materials 44
2.2.1 Chemicals 44
2.2.2 Enzymes 44
2.2.3 Isotopes 45
2.2.4 Antibiotics 45
2.2.5 Competent cell preparation solutions 45
2.2.6 DNA electrophoresis solutions 45
2.2.7 DNA minipreparation solutions 46
2.2.8 DNA extraction from *Drosophila* solution 46
2.2.9 Southern blot and colony hybridisation solutions 46
2.2.10 Protein electrophoresis solutions 46
2.2.11 Protein purification solutions for Mos1 transposase 47
2.2.12 Protein purification solutions for six-his-Mos1 protein 48
2.2.13 Western blot solutions 48
2.2.14 Excision assay solution 48
2.2.15 Bacterial strains 49
2.2.16 Plasmids 49
2.2.17 *Drosophila melanogaster* stocks 51
2.2.18 *Drosophila simulans* stocks 51
2.2.19 Oligonucleotides 52

2.3 Methods 54
2.3.1 Manipulation of bacteria 54
2.3.1.1 Growth of bacterial cultures 54
2.3.1.2 Storage of bacterial cultures 54
2.3.1.3 Preparation of competent *E.coli* cells 54
2.3.1.4 Transformation of plasmid DNA into competent *E.coli* cells 54
2.3.2 Nucleic acid preparation and manipulation techniques 55
2.3.2.1 Small scale preparation of plasmid DNA 55
2.3.2.2 Small scale preparation of plasmid DNA (for use in ABI PRISM^{TM} dye terminator cycle sequencing) 55
2.3.2.3 Large scale preparation of plasmid DNA 55
2.3.2.4 Ethanol precipitation of DNA 56
2.3.2.5 Quantification of DNA concentration 56

2.3.2.6 Restriction enzyme digestion of DNA.......................... 56
2.3.2.7 Agarose gel electrophoresis.......................... 56
2.3.2.8 Recovery of DNA from agarose gels.......................... 57
2.3.2.9 Ligation of cohesive termini.......................... 57
2.3.2.10 Creation of blunt ends from 3' recessed ends.......................... 57
2.3.2.11 Ligation of blunt termini.......................... 57
2.3.2.12 Random primed labelling of DNA.......................... 57
2.3.2.13 Southern blotting.......................... 58
2.3.2.14 Colony hybridisation.......................... 58
2.3.2.15 Hybridisation of radiolabelled DNA to nucleic acids immobilised on nitrocellulose.......................... 59
2.3.2.16 Radioactive signal detection.......................... 59
2.3.2.17 Polymerase chain reaction.......................... 59
2.3.2.18 Manual sequencing of double stranded DNA.......................... 60
2.3.2.19 Automated sequencing of double stranded DNA.......................... 60
2.3.2.20 5' end labelling oligonucleotides with γ- ^{32}P-dATP.......................... 61
2.3.2.21 Primer extension analysis of excision assays.......................... 61
2.3.2.22 Sucrose density gradient ultracentrifugation.......................... 61
2.3.3 Protein expression and detection.......................... 62
2.3.3.1 Expression of recombinant proteins in *E.coli*.......................... 62
2.3.3.2 SDS polyacrylamide gel electrophoresis.......................... 62
2.3.3.3 Determination of protein concentration.......................... 62
2.3.3.4 Purification of GST-Mos1 fusion protein.......................... 62
2.3.3.5 Purification of 6 his-Mos1 protein.......................... 63
2.3.3.6 Mos1 transposase purification.......................... 63
2.3.3.7 Preparation of Mos1 antibodies.......................... 64
2.3.3.8 Western blotting.......................... 65
2.3.3.9 Affinity purification of antibodies.......................... 65
2.3.3.10 Extraction of proteins from *Drosophila*.......................... 65
2.3.4 *Drosophila* methods.......................... 66
2.3.4.1 Preparation of genomic DNA.......................... 66
2.3.4.2 Collection of eggs for microinjection.......................... 66
2.3.4.3 Establishment of transgenic lines by *P* element transformation.......................... 67
2.3.5 Activity assays using Mos1 transposase.......................... 67
2.3.5.1 *In vitro* excision assay.......................... 67

2.3.5.2 Germline transformation of *D. melanogaster* using purified Mos1 transposase 68

Chapter 3 - Overexpression and purification of the Mos1 ORF 69

3.1 Introduction 70

3.2 Results 70

3.2.1 Production of a Mos1 translational fusion protein 70

3.2.1.1 Sub-cloning the Mos1 ORF into prokaryotic expression vectors for production of a translational fusion protein 70

3.2.1.2 Expression of GST-Mos1 fusion protein 71

3.2.1.3 Purification of GST-Mos1 fusion protein 72

3.2.1.4 Preparation of GST-Mos1 antibodies 72

3.2.1.5 Expression of 6-his-Mos1 fusion protein 74

3.2.1.6 Purification of 6-his-Mos1 fusion protein 74

3.2.2 Direct expression of the Mos1 ORF 74

3.2.2.1 Sub-cloning the Mos1 ORF into prokaryotic expression vectors for direct translation 74

3.2.2.2 Expression of Mos1 protein from pET13a 75

3.2.2.3 Expression of Mos1 protein from pBCP378 75

3.2.2.4 Purification of Mos1 protein 76

3.2.2.5 Affinity purification of antibodies 76

3.2.2.6 Confirmation of the identity of the purified protein using anti-Mos1 antibodies 78

3.2.2.7 Sequencing the Mos1 ORF in plasmid pBCP378-Mos1 78

3.3 Discussion 80

Chapter 4 - *In vitro* activity of purified Mos1 protein 82

4.1 Introduction 83

4.2 Results 83

4.2.1 *In vitro* excision of *mos*1 DNA sequences from plasmid p*mos* 83

4.2.2 A mock purification of transposase is unable to support excision of *mos*1 DNA from plasmid p*mos* 86

4.2.3 Mos1 transposase requires a divalent cation for excision activity..... 86
4.2.4 Mos1 transposase requires a supercoiled DNA substrate for excision........ 88
4.2.5 Excision of *mos*1 is inversely related to the size of the element....... 88

4.3 Discussion........ 94

Chapter 5 - Analysis of excised *mariner* elements........ 97

5.1 Introduction........ 98

5.2 Results........ 99
5.2.1 Analysis of *mariner* elements excised *in vivo*........ 99
5.2.1.1 Detection of extrachromosomal *mariner* sequences........ 99
5.2.1.2 Separating extrachromosomal *mariners* from genomic DNA by centrifugation........ 101
5.2.1.3 Detection of circular *mariner* elements by PCR........ 101
5.2.1.4 Amplification is across circular *mariner* elements not tandemly inserted copies........ 104
5.2.1.5 Structures of joined *mariner* ends........ 108
5.2.2 Analysis of *mos*1 elements excised *in vitro*........ 114
5.2.2.1 Excised copies of *mos*1 obtained *in vitro* are linear........ 114
5.2.2.2 Mapping *in vitro mos*1 cleavage sites at the nucleotide level........ 116

5.3 Discussion........ 123
5.3.1 Circular forms of *mariner*........ 123
5.3.2 Linear forms of *mos*1 obtained *in vitro*........ 126

Chapter 6 - Recombinant Mos1 transposase is active *in vivo*........ 131

6.1 Introduction........ 132

6.2 Results........ 133
6.2.1 *D. melanogaster* is the ideal organism for investigating the *in vivo* activity of Mos1 transposase........ 133

6.2.2 Construction of p*mos-white*, a germline transformation vector....... 134
6.2.3 Recombinant transposase mediated transformation of *D. melanogaster*........ 134
6.2.4 Analysis of transformed lines........ 136

6.3 Discussion........ 143

Chapter 7 - Introduction of the *C. elegans* transposable element *Tc3* into *D. melanogaster*........ 146

7.1 Introduction........ 147

7.2 Results........ 147
7.2.1 Choosing suitable genetic markers........ 147
7.2.2 Plasmid construction........ 148
7.2.3 Construction of *D. melanogaster* strain y^1 v^{36f}; ry^{506}........ 148
7.2.4 The creation of transgenic lines by *P* element mediated mutagenesis........ 150
7.2.5 Localising the chromosome of insertion........ 153
7.2.6 Mobilisation of the marked Tc3 element........ 156
7.2.7 Detecting Tc3 transposase by Western blot analysis........ 161

7.3 Discussion........ 163

Chapter 8 - Summary and Future Work........ 165

Bibliography........ 168

Chapter 1

Introduction

1.1 Introduction to transposable elements

Transposable elements are segments of DNA present in both prokaryotic and eukaryotic genomes that have the ability to move to new sites and to undergo deletion or amplification without the extensive DNA sequence homology needed for classical recombination (Berg and Howe, 1989).

Transposable elements were first discovered in maize (*Zea mays*) by Barbara McClintock in the 1950s. Using genetic experiments which detected variegation in the phenotypes of maize kernels, the presence of mobile elements originally called 'controlling elements', which were able to transpose to new chromosomal locations were reported (McClintock, 1957). Bacterial insertion sequences and resistance transposons, and the mobile elements of *Drosophila melanogaster* and other eukaryotes were isolated subsequently.

Transposable elements have since been found in all genomes in which they have been sought. They comprise a major fraction of eukaryotic genomes, occurring as families of dispersed repeat sequences scattered throughout the genome, the number of copies depending on the transposable element and the species concerned. In *D. melanogaster*, over thirty different elements have been found constituting around 10% of the total genome size.

1.1.1 Classification of eukaryotic transposable elements

According to their structure and presumed mechanism of transposition eukaryotic transposable elements can be grouped into two classes (see figure 1.1):- Class I elements, or retrotransposons, which transpose by reverse transcription of an RNA intermediate and Class II elements which transpose via a DNA mediated excision-insertion mechanism. Prokaryotic transposons are most similar to class II elements as they also transpose by a mechanism involving a DNA intermediate.

1.1.1.1 Class I elements

During transposition of retrotransposons, an RNA copy of the element is transcribed. Reverse transcriptase encoded by the element, then converts the RNA intermediate into a double stranded DNA copy. The resulting DNA copy is then inserted at a new site in the genome by element-encoded integrase. Because the original DNA copy of the

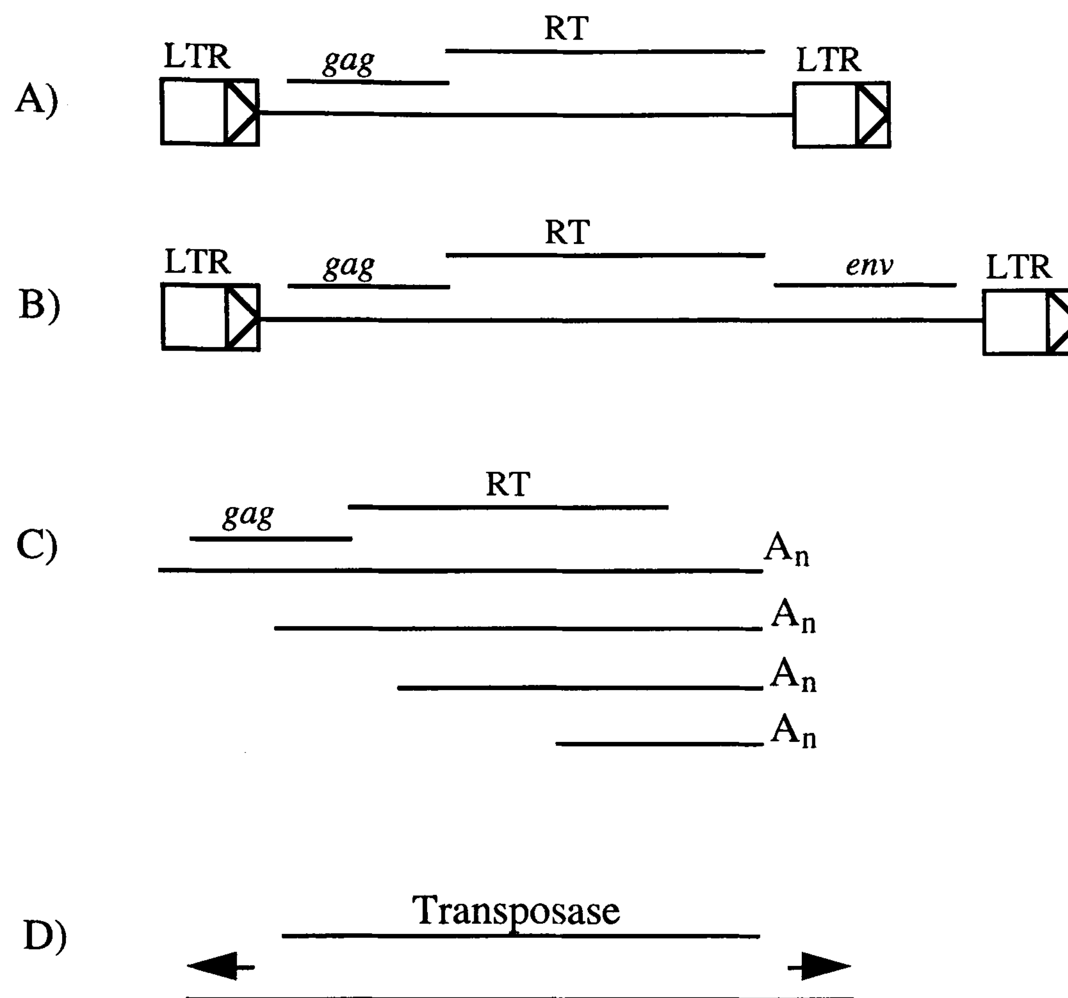


Figure 1.1 Schematic representation of the structures of eukaryotic transposable elements (adapted from Finnegan, 1992).

A) Class I elements that have long terminal repeats (LTRs). All have a *gag* like ORF and an ORF that encodes a reverse transcriptase (RT). eg. *copia*, *Ty1*.
B) LTR elements that contain a third, env-like ORF. eg. *gypsy*.
C) Class I, LINE like elements that have A rich sequences at their 3' termini. eg. *I* factor, *L1* elements.
D) Class II transposable elements containing terminal inverted repeats and a gene encoding a transposase.
See text for further details.

element remains in the genome, the mechanism of transposition of class I elements is necessarily replicative. Each round of transposition generating at least one additional copy of the element.

Retrotransposons are of two types. These are the retrovirus-like elements, which have long terminal repeat sequences (LTRs), and the LINES (long interspersed nuclear elements), also known as 'non-LTRs', which have no terminal repeats.

(i) LTR elements

The LTR elements contain open reading frames (ORFs) similar to the *gag* and *pol* genes of retroviruses. Transcription of the RNA transposition intermediate initiates in one LTR and terminates in the other. Host RNA polymerase is responsible for this. The RNA is then packaged into virus-like particles constructed of proteins encoded by the *gag* and *pol* like ORFs (Boeke, 1989; Burns *et al*, 1992). The RNA then becomes reverse transcribed yielding linear double stranded DNA and integrated into new sites in the genome of the host cell. The enzymes responsible for these reactions, reverse transcriptase and integrase, are encoded by the reverse transcriptase and integrase domains of the *pol* gene. Examples of LTR retrotransposons containing two ORFs are the *Ty*-1 elements from *Saccharomyces cerevisiae* and *copia*-like elements from *D. melanogaster.*

Some LTR elements also have a third ORF; *env*, which is similar to the *envelope* gene of retroviruses. In retroviruses, *env* codes for components of the viral envelope which are required for entry of the retrovirus into the cell. *Gypsy*, a retrotransposon from *Drosophila*, contains a third ORF and has infectious properties similar to retroviruses. High levels of *gypsy* insertion activity are observed when *Drosophila* larvae lacking *gypsy* are exposed to *gypsy* viral-like particles (Song *et al*, 1994; Kim *et al*, 1994).

(ii) LINEs

LINE-like elements also contain both *gag* and *pol*-like ORFs and are believed to transpose by a similar mechanism to the LTR elements. Transposition through an RNA intermediate has been demonstrated for the *I* factor of *D. melanogaster* and mouse *L1* elements (Pelisson *et al*, 1991; Jensen and Heidmann, 1991; Evans and Palmiter, 1991). The second ORFs of LINE-like elements do not include recognisable integrase domains. Domains similar to the apurinic family of nucleases exist, these

have been demonstrated to be important for transposition (Feng *et al*, 1996; Moran *et al*, 1996). The LINE-like elements have no terminal repeat sequences but have A-rich sequences at the 3' end of their coding strand. Within a family of elements many copies may be truncated at their 5' ends.

(iii) SINEs

SINEs are short interspersed nuclear elements, which are found in hundreds of thousands of copies in human and other mammalian genomes. Like LINEs, they have an A rich sequence at the 3' end of one strand. SINES are related in sequence to the transcripts of pol III transcribed genes and appear to have been produced by reverse transcription and integration. They have no coding capacity themselves and must rely on enzymes from elsewhere for their mobility (Deininger, 1989).

1.1.1.2 Class II elements

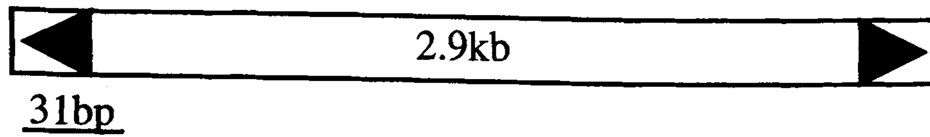
Members of this class possess inverted terminal repeat sequences of various lengths and encode proteins (transposases) which are essential for transposition. The mechanism of transposition is a DNA-DNA mediated one, also called 'cut and paste' where the element is physically removed from the chromosome and inserted elsewhere. Transposition of class II elements is intrinsically conservative in nature. However, if the sequence of the transposable element is copied back into the site from which it excised, or transposition is from replicated to unreplicated DNA then an extra copy of the element is created.

Representatives of class II elements can be grouped into four main families based on sequence comparisons. Members of these families are shown diagramatically in figure 1.2.

The *P* family, composed of the *P* element from *D. melanogaster* (Engels, 1989) and a few other related transposons,

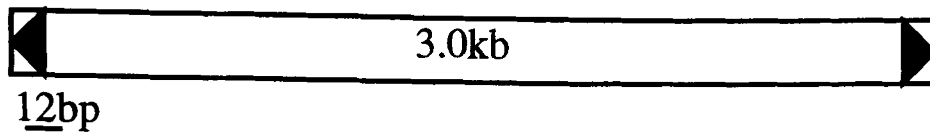
The hAT superfamily (*hobo-Ac-Tam3*) represented by *hobo* from *D. melanogaster*, *Ac* of *Z. mays*, the original 'controlling element' and Tam3 from the snap dragon *Antirrhinum majus* (Calvi *et al*, 1991),

I. The ***P*** element (*D. melanogaster*)

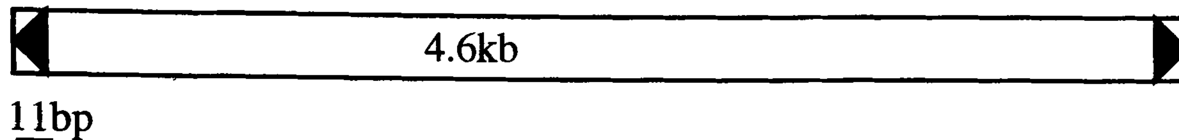


II. Representative members of the hAT family

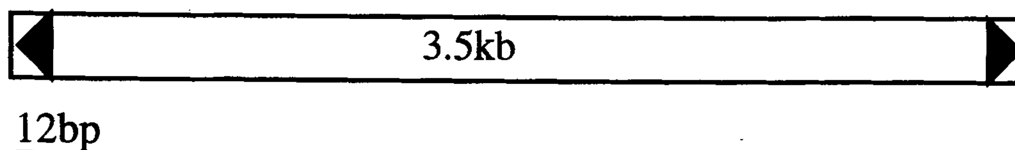
Hobo (*D. melanogaster*)



Ac (*Z. mays*)

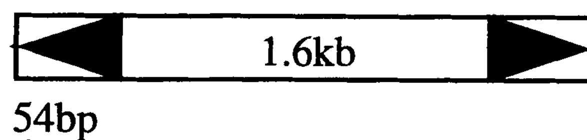


Tam3 (*A. majus*)

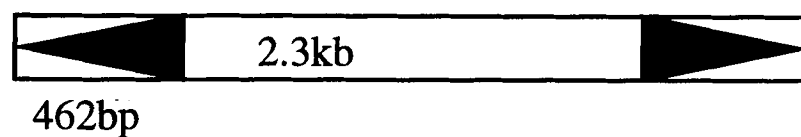


III. Representative members of the *Tc1/mariner* family

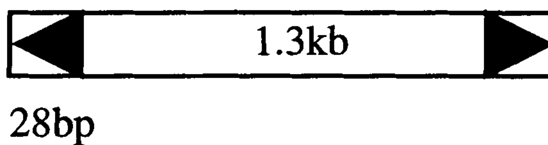
Tc1 (*C. elegans*)



Tc3 (*C. elegans*)



***mariner mos*1** (*D. mauritiana*)



IV. The *spm* element (*Z. mays*)

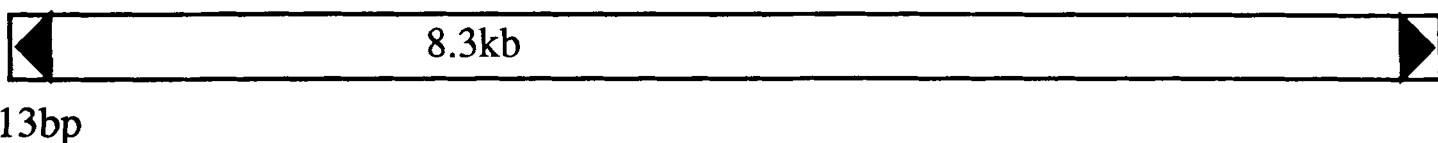


Figure 1.2. Structural features of representative members of the *P*, *hAT* and *Tc1/ mariner* families of transposable elements

Lengths of elements and their respective inverted repeats are indicated.

The *Tc1/ mariner* superfamily, so called from the transposons *Tc1* of *C. elegans* and *mariner* of *D. mauritiana* (Henikoff and Henikoff, 1992; Robertson, 1995),

The *Spm* (*En*) family, containing the *suppressor-mutator* (*spm*) element from Zea mays (Fedoroff, 1989).

1.2 Introduction to *mariner*

1.2.1 Isolation of the *mariner* element

The first *mariner* element to be isolated was discovered on analysis of an unstable mutation in the *white* gene (*w*) of *Drosophila mauritiana*, a sibling species of *Drosophila melanogaster.* The mutant eye colour observed was peach and was due to the insertion of a novel transposable element into the 5' untranslated leader sequence of *white*. Further analysis revealed the presence of approximately twenty copies of this element in the genome of *D. mauritiana.*

The transposable element was named *mariner* and the particular copy inserted into the *white* gene designated the *peach* element. The allele of the insertion mutation was titled *white peach* (w^{pch})(Jacobson and Hartl 1985, Haymer and Marsh, 1986).

1.2.1.1 Instability of the w^{pch} allele

The instability of the w^{pch} allele is due to excision of the *peach* element from the *white* gene. Somatic excision of *peach* restores pigment production in these cells and is detected phenotypically by the presence of wild-type patches of pigment on a peach coloured background, this effect is called somatic mosaicism. Germline excision is detected by reversion to wild-type eye colour in the progeny.

The peach copy of *mariner* is inactive or nonautonomous and in the absence of active *mariner* elements elsewhere in the genome can not transpose (Garza *et al*, 1991). However, in most w^{pch} bearing strains of *D. mauritiana* functional but low activity elements exist and excision events do occur, albeit infrequently (Jacobson and Hartl, 1985).

1.2.1.2 Identification of *mos* factors

Some strains of *D. mauritiana* exhibit a high rate of somatic excision of the *peach* element, resulting in mosaic eye colour in almost every fly (Bryan *et al,* 1987, Medhora *et al*, 1988). The high rate of somatic excision is accompanied by a high rate of *mariner* transposition (Bryan *et al*, 1990) and depends on the presence of one or more highly active, autonomous copies of *mariner* in the genetic background (Medhora *et al*,1988). These highly active *mariner* elements are called *mos* (mosaic) factors, and are believed to activate transposition of other *mariner* elements by supplying transposase activity *in trans*.

The mosaic factor, *mos*1 has been introduced into a strain of *D. simulans* carrying the w^{pch} allele, by repeated backcrossing from *D. mauritiana. Mos*1 was identified as a segregating *mariner* containing DNA fragment whose presence resulted in eye colour mosaicism. The element was subsequently cloned (Medhora *et al*, 1991).

1.2.1.3 Structure of *mos*1

The mariner element, *mos*1 is 1286 bp in length. It contains 28 bp inverted repeats (IR) at its termini which are flanked by duplications of a TA dinucleotide at the genomic target site. Within the inverted repeats a single uninterrupted open reading frame (ORF) of 1035 nucleotides is present which encodes a predicted polypeptide of 345 amino acids, with a molecular weight of 40.8 kD.

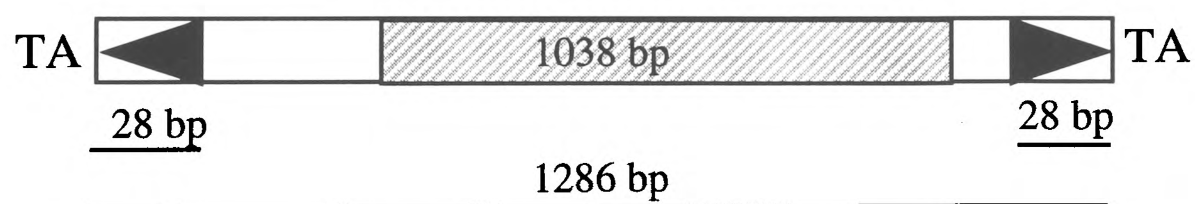


Figure 1.3. Structure of the *mariner* transposable element, *mos*1

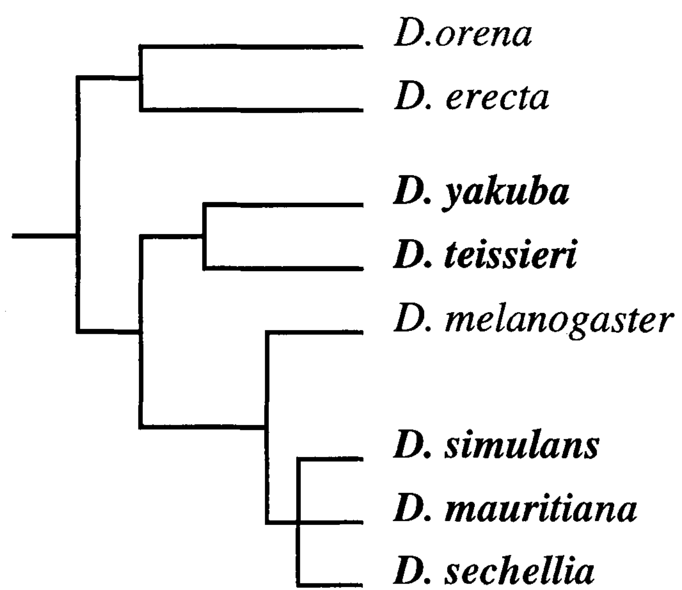
1.2.2 Distribution of *mariner* elements

1.2.2.1 Distribution of *mariner* elements in the melanogaster species subgroup

The abundance of *mariner* differs dramatically in the genomes of the closely related species *D. mauritiana, D. simulans, D. sechellia* and *D. melanogaster.*
The first three species cross-hybridise (Lachaise *et al*, 1988). Test crosses between *D. simulans* females homozygous for w^{pch}, carrying *peach* as the only *mariner* element and males from *D. mauritiana, D. simulans and D. sechellia* have been performed to assess the excision activity of *mariner* elements in these different species. W^{pch} was initially transferred from *D. mauritiana* to *D. simulans* by a series of backcrosses. Female hybrids from the interspecific cross were then backcrossed to *D. simulans* males in each generation (Bryan and Hartl, 1988). In *D. mauritiana,* each individual contains 20-30 copies of *mariner* distributed widely at many polymorphic sites throughout the genome (Jacobson *et al*, 1986), some of which are active in transposition and excision (Maruyama *et al*, 1991). In *D. simulans*, the number of *mariner* elements per genome varies from 1-10, and the overall activity of the elements varies from one population to another (Capy *et al*, 1992). *D. sechellia* contains two *mariner* elements present at fixed sites in the genome, one is full length and has low biological activity, the other has a 3' end truncation (Capy *et al*, 1991). Elements in the *mariner* family have not been detected in natural isolates of *D. melanogaster* (Jacobson *et al*, 1986).

Mariner occurs in five of the eight species of the *melanogaster* species subgroup (figure 1.4.). The discontinuous distribution of *mariner* within this subgroup may be due to stochastic loss of the element. According to this theory, *mariner* would have been present in the ancestral species before the *melanogaster* subgroup diverged and been lost in the lineage leading to *D. melanogaster* and the lineage leading to *D. erecta* and *D. orena* .

An alternative mechanism for the distribution is horizontal transfer. This is the way in which *P* elements have spread to *D. melanogaster* in the past few decades (Daniels *et al*, 1990). In this case, *mariner* would have been transferred from an outside source into some species of the *melanogaster* species sub-group, subsequent to its divergence.



Species in bold contain *mariner* elements

Figure 1.4. The *melanogaster* species subgroup

Several observations have been made which suggest that the discontinuous distribution of *mariner* is due to stochastic loss. Firstly, the *mariner* gene phylogeny is congruent with the species subgroup phylogeny based on nucleotide sequences of the *alcohol dehydrogenase* gene. The pattern of divergence of elements between species is not compatible with that expected due to interspecific transfer (Maruyama and Hartl, 1991a). Secondly, some laboratory strains of *D. simulans* have lost *mariner* elements. The loss occurring during laboratory maintenance and inbreeding (Capy *et al*, 1990, Capy *et al*, 1991). Although not conclusive, these findings are more simply explained by vertical transmission than horizontal transfer.

1.2.2.2 Widespread distribution of *mariner* elements

Mariner sequences that have been found in a distantly related drosophilid, *Zaprionus tuberculatis* may have resulted from horizontal transfer from the *melanogaster* species sub group. This species, estimated to have diverged at least 50 million years ago, carries sequences which contain 97% identity to *mariner* nucleotide sequences in the *melanogaster* species subgroup (Maruyama and Hartl, 1991b). The serendipitous finding of *mariner*-like sequences in the lepidopteran species *Hyalophora cecropia* , with an overall 49% identity at the nucleotide level to *mos1* of *D. mauritiana* (Lidholm *et al*, 1991) also suggests the possibility of horizontal transfer between even more distantly related species.

Robertson (1993), using primers designed to represent regions of contiguous, conserved amino acids between the putative transposase genes of the *D. mauritiana* and *H. cecropia* elements, was able to amplify intervening segments of the transposase genes from putative *mariner* transposases from various arthropods. The initial screen revealed presumed *mariner* elements from ten other insects representing six additional orders including those as diverse as bees, mosquitoes, silverfish, cat fleas and earwigs. Sequences of multiple clones from each species revealed a diverse array of *mariner* elements, some with multiple subfamilies in their genomes, indicating both vertical inheritance and horizontal transfer.

An intact ORF in each species was found suggesting each may carry functional transposable elements. Thirteen distinct *mariner* subfamilies have been identified in arthropods. In general the amino acid sequences within a subgroup have more than 40% identity, while members of different subfamilies show 25-40% identity.

Mariner like elements from organisms outside of arthropods have also been reported. The lower invertebrate planarian, *Dugesia tigrina*, a flatworm, contains ~8000 copies of a *mariner* like element (Garcia-Ferandez *et al*, 1993). The high copy number, and presence of an uninterrupted ORF point to an active *mariner* element in planarians. The high degree of similarity shared with *mariner* elements from some arthropods and the fact that this element is undetectable in other planaria species, strongly suggests a case of horizontal transfer between the two distant phyla (Garcia-Femandez *et al*, 1993, Garcia-Femandez *et al*, 1995).

Two types of *mariner* elements that define subfamilies not found in arthropods have been discovered in the nematode *C. elegans* (Garcia-Femandez *et al*, 1993, Robertson *et al*, 1992, Sedensky *et al*, 1994). *Mariner* like elements (MLEs) have also been discovered in vertebrates. Internal fragments of the putative transposase gene of MLEs were amplified from human, mouse, rat, Chinese hamster, sheep and bovine genomic DNAs by PCR (Auge-Gouillou *et al*, 1995). More recently, two full length MLEs, *hum*1 and *hum*7, from humans have been cloned, both contain numerous stop codons and frameshifts (Oosumi *et al*, 1996). The *hum*1 element is a member of the irritans subfamily and *hum*7, the cecropia subfamily of MLEs.

1.2.2.3 Distribution of other Class II elements

Similarities between the predicted proteins of *D. mauritiana mariner* and the *Tc1* element from *C. elegans* have been noted (Henikoff and Henikoff, 1992). The *Tc1* family of elements like the *mariner* family is diverse, with members in nematodes, flies, fungi and fish (Robertson, 1995; Robertson and Asplund, 1996). The two families are sister families sharing 18-25% amino acid identity in their transposases. A family tree depicting the relationships between different members of the *Tc1* and *mariner* families is shown in figure 1.5.

Unlike the *Tc1/mariner* family of elements, distribution of the *P* element is virtually restricted to the genus *Drosophila.* It is thought that *P* elements are a relatively recent acquisition to the genome of *D. melanogaster*. Laboratory stocks of *D. melanogaster* established in the first half of this century do not contain *P* elements, but all the recently caught strains do. The discovery that a *P* element from *D. willistoni* contains just one nucleotide different to the *P* elements of *D. melanogaster* suggests that *D. willistoni* was the source of a horizontal transfer into *D. melanogaster* (Daniels *et al*, 1990). This recent invasion hypothesis is now generally accepted.

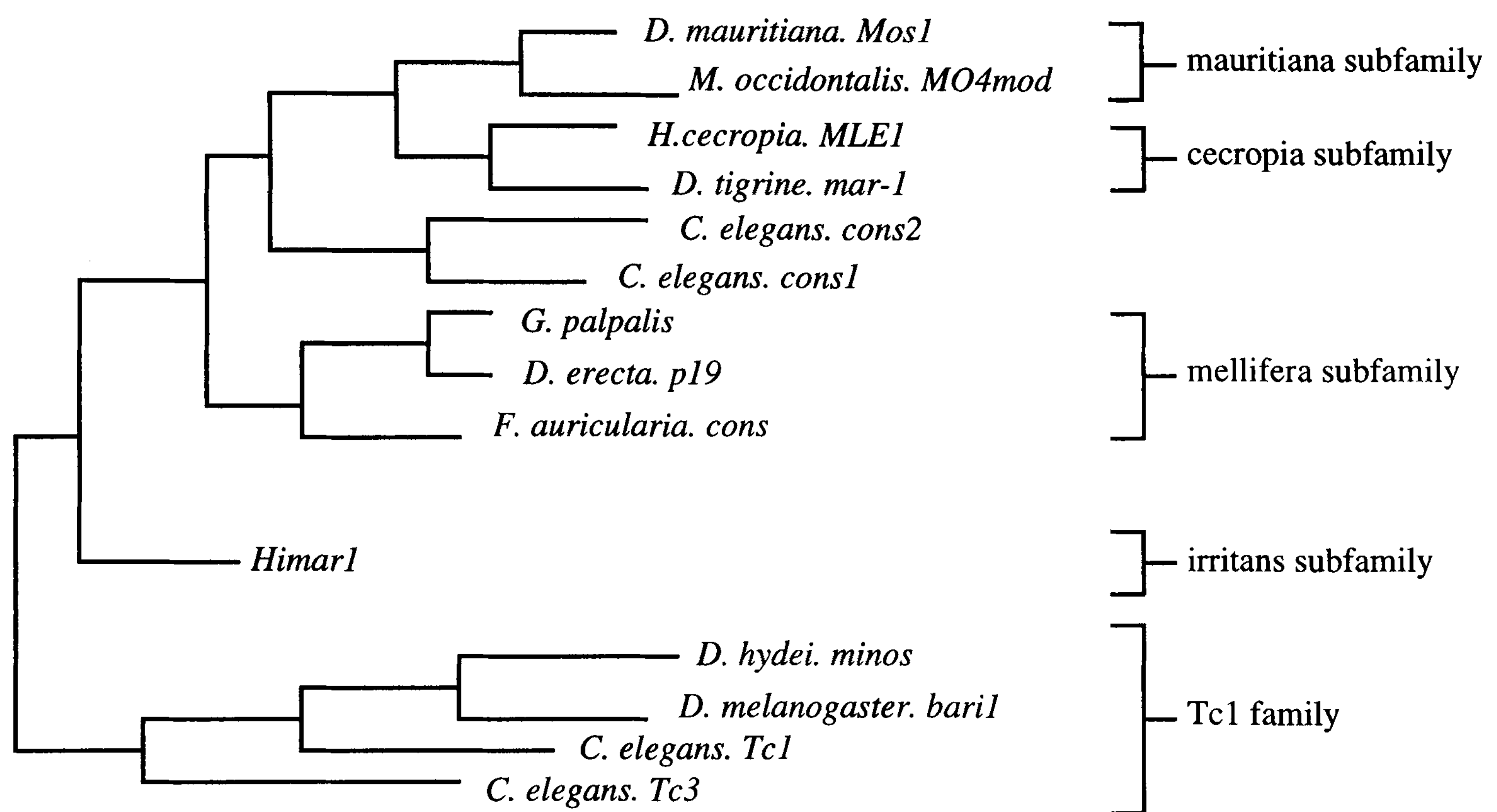


Figure 1.5. The *Tc1/mariner* family. Relationship of the various *mariner* subfamilies to the *Tc1* family of transposable elements (from Lampe *et al*, 1996).

Recently *P* elements have also been detected in the *Drosophila* species: *D. nebulosa*; *D. bifasciata; D. subobscura* and *Scaptomyza pallida*. To account for the current distribution of *P* elements analysis of their evolution indicates that in addition to the extremely recent transfer of the canonical *P* element from *D. willistoni* to *D. melanogaster*, related elements have undergone at least two more horizontal transfers within the *Drosophilidae* (Clark *et al*, 1994).

The *hobo* element is restricted to the genus *Drosophila*, and like *P* appears to have recently invaded the genome of *D. melanogaster* (Simmons, 1992). The transposase sequences of *hobo*, *Activator* (*Ac*) from maize and *Tam3* from *Antirrhinum majus* exhibit low levels of coding similarity spanning several hundred codons. It has been suggested that horizontal transfer rather than vertical inheritance from a common ancestor that existed prior to the divergence of animals and plants may be responsible for the high sequence similarity (Calvi *et al*, 1991).

1.2.2.4 Superfamily of transposable elements

A superfamily of transposases has been recognised which links together a number of different transposable element families. These distant relationships are based on a shared, presumed catalytic D,D35E domain in their encoded transposases and relate the *Tc1/mariner* families of transposases to the *Tec* and *TBE* transposons of ciliates and the *IS630* and *IS3* families in bacteria (Doak *et al*, 1994). This domain is also present in transposases of the prokaryotic elements Tn7, Tn10 and phage Mu transposase as well as retroviral and retrotransposon integrases (see Craig, 1995; Grindley and Leschziner, 1995).

The 'D,D35E' motif consists of two aspartic acid residues (D), typically separated by more than ninety amino acids, followed by a glutamic acid residue (E), typically 34 or 35 amino acids further towards the carboxy terminal. The spacing between the latter D and E is essentially invariant at 34 amino acids in eukaryotic elements, but 35 in prokaryotic ones. *Mariner* and *mariner*-like elements are unique in that their acidic signature is D,D34D (Robertson and Macleod, 1993). The acidic residues are thought to be part of the catalytic site of the transposase and involved in co-ordinating a metal ion necessary for catalysis (Kulkosky *et al*, 1992; Mizuuchi, 1992, Sarnovsky *et al*, 1997).

1.2.3 Autonomous and non-autonomous elements

Autonomous transposable elements are elements which produce active transposase and can be mobilised within their resident genome. Non-autonomous elements are derivatives of autonomous copies that are unable to produce active transposase themselves, but nevertheless can transpose if active transposase is produced *in trans* from an autonomous copy of the same element.

1.2.3.1 Non-autonomous *mariner* elements contain nucleotide substitutions

Sequencing and analysis of the mosaic factor *mos1*, *peach* and other copies of the *mariner* element from the genomes of *D. mauritiana* and *D. simulans*, has revealed that all are 1286bp in length and differ from each other generally by single nucleotide changes. *Mos*1 and *peach* differ in eleven nucleotide positions distributed throughout their length. These are three substitutions upstream of the open reading frame, two in the putative promoter region (defined in Jacobson *et al,* 1986), two single base changes 3' of the ORF and five in the ORF, four of which result in amino acid replacements (Maruyama *et al*, 1991).

Analysis of the functional effects of nucleotide substitutions by molecular (site-directed mutagenesis) and evolutionary (cladistic) techniques has been performed, revealing that the nucleotide sequence of the element is the primary determinant of function. Missense point mutations within the putative transposase coding region appear to be responsible for inactivation (Maruyama *et al*, 1991). Inactivating mutations may result in frameshift alterations or in premature translation termination. Transposase activity supplied *in trans* is thus also a prerequisite for transposition of these non autonomous elements.

In general, the *Tc1* and *mariner* classes of elements are homogeneous in length. Deleted elements have been found though in these classes of elements. A 3' end truncated *mariner* in *D. tiesseiri* and internally deleted copies of Tan1, a transposon from *Aspergillus niger*, a member of the Fot1 family of the *Tc1*/ *mariner* superfamily have been detected but these are exceptional cases (Nyyssonen *et al*, 1996).

1.2.3.2 Non-autonomous copies of other class II transposable elements

In contrast to the *mariner/Tc*1 family, many non-autonomous transposable elements are inactive because they contain internal deletions and thus cannot code for a functional transposase.

P element containing strains of *D. melanogaster* typically possess 40 to 50 *P* elements. These elements can be subdivided into two classes: full length elements of 2.9kb and heterogeneous elements with internal deletions (Engels, 1989). The smaller *P* elements are the major type of *P* sequence in the genome. They do not encode transposase and are non-autonomous or inactive, although they code for regulatory proteins (see 1.4.2). However, they can be mobilised when transposase is supplied *in trans* as they contain all the cis-acting sequences required for transposition.

The *hobo* element of *D. melanogaster*, appears to be analogous to *P*. Full length, functionally complete *hobo* elements are 3.0kb, but defective, internally deleted *hobo*s which rely on complete *hobos* for their mobilisation also exist (Streck *et al*, 1986).

Non-autonomous dissociation, (Ds) elements from maize are internally deleted copies of the autonomous element, activator (Ac). The *Ac* element is 4.6kb in length. *Ds* elements are deletion derivatives, which range from *Ac* elements with deletions to sequences that are virtually unrelated to that of the *Ac* element, except for the terminal inverted repeats (Federoff, 1989). *Ds* sequences although inactive themselves, can be promoted to transpose in the presence of *Ac*.

1.3 The transposition reaction

1.3.1 Recognition of DNA binding sites by transposase

The initial step in the mobilisation of a transposable element is the recognition and binding of specific DNA sites near the ends of the element by an element encoded transposase. The formation of a stable DNA-protein complex is a prerequisite to catalysis. Different transposable elements form different protein-DNA complexes, varying in the number of monomers of transposase bound and the number of binding sites. Some transposable elements also require host proteins. Sequence-specific DNA binding has been demonstrated for at least three members of the *Tc1/mariner* family .

DNase1 footprinting and gel retardation assays have been used to demonstrate that *Tc1A*, the *Tc1* transposase, is a site-specific DNA binding protein which recognises the inverted repeats of the *Tc1* element. The primary recognition sequence is between base pairs 5 and 26 of each inverted repeat (Vos *et al*, 1993).

Similarly, *Himar*1 and *Tc3* transposases have been found to bind to the inverted terminal repeat sequences of their cognate transposon (Lampe *et al*, 1996; van Leunan *et al*, 1993). *Himar*1 transposase protects its inverted terminal repeats from DNaseI cleavage, covering 28 nucleotides on both strands. The top strand is protected between nucleotides 4 and 32, and the bottom between 2 and 30.

Tc3A binds specifically to a radiolabelled fragment containing the terminal 40 nucleotides of the *Tc3* inverted repeat as detected by gel retardation experiments (van Leunan *et al*, 1993).

Recently transposase binding sites have been delineated in the *pogo* element. *Pogo* contains two binding sites at each end of the element, it also contains a number of sites approximating to the binding site consensus, which are clustered just upstream of the translation start site (Wang, 1997).

Transposases encoded by prokaryotic elements such as IS30 and IS903, also recognise sequences within the inverted repeats of their elements (Stalder *et al*, 1990: Derbyshire and Grindley, 1992). Mu contains a transposase binding site at each end of the transposon, but also has additional binding sites that are asymmetrically spaced at the left and right arm of the transposon (Craigie *et al*, 1984).

The *P* transposase differs from those of the *Tc1/mariner* superfamily so far characterised, in that it does not bind to sequences within the inverted repeats. The transposase instead mediates binding to a 10bp internal sequence (Kaufman *et al*, 1989; Misra and Rio, 1990). The transposase binding site is located 16bp and 4bp from the 5' and 3' inverted repeats, respectively. However, the 31bp inverted repeats of the *P* element are absolutely required for *P* element transposition *in vivo* (Engels, 1989). Inverted repeat binding protein (IRBP), a *Drosophila*, cellular, non *P* element encoded DNA binding protein is able to specifically interact with the 31bp inverted terminal repeat sequence of the *P* element. Sequence analysis has shown that the IRBP is related to the 70kD subunit of the human Ku autoimmune antigen which is implicated in double strand break repair (Beall *et al*, 1994).

Transposases encoded by the plant transposons *En*-1 and *Ac*, also do not bind to the inverted repeat sequences of their respective elements, but rather recognise subterminal repeat motifs that are asymmetrically located within the left and right arms of the transposon (Gierl *et al*, 1988; Kunze and Starlinger, 1989).

The sequence of the terminal few bp is relatively conserved among class II elements (figure 1.6) (Collins *et al*, 1989). However, these terminal most nucleotides are not required for recognition by their respective transposases. The terminal few bp may instead be important for steps in the transposition reaction that occur after the initial recognition of sequences by the transposase.

Element	**Species**	**5'-inverted repeat-3'**
Mariner	*D. mauritiana*	C**CAGGTG**TAC
Tc1	*C. elegans*	**CAG-TG**CTG
Tc3	*C. elegans*	**CAG-TG**TGG
P	*D.melanogaster*	**CA--TG**ATG
Hobo	*D. melanogaster*	**CAG**AGAACT
Spm/En	*Z. mays*	**CACTACAA**G
Tam1	*A. majus*	**CACTACAA**C
Tam2	*A. majus*	**CACTACAA**C
Tgm1	*G. max*	**CACTA**TTAG

Figure 1.6. A comparison of inverted repeat termini of several class II elements.

1.3.2 DNA binding domains of transposases

DNA binding domains contained within the transposase protein are responsible for localisation of a particular transposase to DNA binding sites within the element, such that a synaptic DNA-protein complex may be formed prior to mobilisation. Functional dissection has been performed for a number of different transposases to elucidate the domains involved in DNA binding.

Both *Tc1* and *Tc3* transposases have two distinct DNA binding activities, a sequence specific DNA binding activity which is found in the N-terminal part of the transposase and a non specific DNA binding activity (Vos *et al*, 1993; Colloms *et al*, 1994).

The sequence specific DNA binding domain of *Tc1A* is bipartite, as revealed by mutational analysis of the binding site and interference studies (Vos and Plasterk, 1994). Protein sequence alignments have revealed a weak similarity between the amino terminus of the *Tc*1 like *minos* transposase and the paired box sequence, a highly conservative DNA-binding protein domain found at the amino terminal end of the *Drosophila* paired gene product and several other Drosophila and mammalian genes involved in embryonic development. (Franz *et al*, 1994; Czerny *et al*, 1993).

Recent database searches have revealed the presence of a helix turn helix (HTH) motif in transposases from *Tc1*, *mariner* and *pogo* subgroups (Pietrovski and Henikoff, 1997). HTH motifs are DNA binding motifs found in many diverse proteins that bind site-specific DNA sequences (Pabo and Sauer, 1992). The position of the predicted HTH motifs are in the N-terminal regions of the transposases. This coincides with the N-terminal specific DNA binding activity observed for *Tc1, Tc3* and *pogo* transposases. Site-directed mutagenesis of single residues in the HTH domain of pogo transposase, which are predicted to be particularly important in DNA binding completely abolish DNA binding. Single substitutions of amino acids in the HTH motif to proline, which alters local protein secondary structure, also eliminate DNA binding activity (Wang, 1997).

Analysis of transposase sequences, as performed by Pietrovski (1997), demonstrates that the central and catalytic, carboxy terminal regions between *Tc1/mariner* family members appear more conserved than the N terminal domains. Despite the helix-turn-helix motif there is little conservation between N-terminal regions of these transposases. There is also no apparent conservation of the inverted repeat sequences of these elements apart from the terminal most nucleotides. The site-specific DNA binding regions of transposases of the *Tc1/mariner* family characterised to date recognise inverted repeat sequences of their own transposon. It is possible that the inverted repeat sequences are divergent between different transposons, so that the transposase protein of a given element will only act at its own end.

1.3.3 Cleavage and strand transfer

Well-characterised bacterial transposons appear to transpose by similar mechanisms (see Mizuuchi, 1992). DNA transposition mediated by the bacterial transposons Tn10 and Tn7 and the transposing bacteriophage Mu are particularly well characterised mechanistically. This type of DNA transposition is shared by many plant and animal

transposons. Integration of the DNA copies of retroviral genomes and LTR-retrotransposons into host chromosomes also occurs by a similar mechanism.

The recombination stage of transposition takes place in two distinct chemical steps: endonucleolytic cleavages at the ends of the transposon, followed by a pair of DNA strand transfer events that covalently join the two strands of a target DNA to the 3' ends of the transposon DNA. The target DNA cleavage and joining to the donor DNA ends are energetically coupled and external energy sources such as ATP are not required. The reactions also proceed without covalent DNA-protein intermediates (Mizuuchi, 1992). The subsequent gap filling processes are carried out by cellular DNA gap-repair or by replication machinery.

The first chemical step is nicking at the 3' ends of the transposon sequence by transposase protein which binds to specific sequences located at or near the two ends of the transposon (1.3.1). In this reaction, the 5'-phosphoryl is transferred to a water molecule. In the subsequent strand transfer step, a target DNA is cut on both strands and the 5'-phosphoryl ends are joined to the free 3' ends of the transposon liberated in the first step.

Sometimes, DNA cleavage reactions occur at the 5' ends of the element also. DNA cleavage reactions at the 5' end profoundly influence the structure of the transposition product (see figure 1.7). Replicative transposons such as Mu nick only at the 3' ends of the transposon (Pato, 1989;Varmus and Brown, 1989; Mizuuchi, 1992). Non-replicative elements such as the prokaryotic transposons Tn7 and Tn10 and eukaryotic elements *Tc*1 and *Tc*3 cut both strands at or near the transposon ends, excising the element from its original surroundings (Bainton *et al*, 1991; Benjamin and Kleckner, 1992; Vos *et al*, 1996; van Leunan *et al*, 1994). Thus, although all transposition reactions involve DNA breaking and joining, several different types of recombination products can emerge, depending on which DNA strands are broken and joined.

1.3.3.1 Retroviral integration

In retroviral integration, the viral RNA genome is converted to double strand DNA by reverse transcription. This DNA product is then cleaved by the retrovirally encoded integrase to expose the 3' ends of the actual retroviral DNA. Strand transfer reactions then join these exposed 3' ends to staggered positions on the target DNA, one transposon end joining to one target strand and the other to a displaced position on the

other target strand. As a result, the transposon is covalently joined to the target DNA but is flanked by short gaps that reflect the staggered positions of target joining. Host DNA repair functions then repair these flanking gaps. The resulting product in which the retroviral DNA is covalently linked to the target DNA is called a simple insertion (see figure 1.7. A).

1.3.3.2 Bacteriophage Mu replication

Bacteriophage Mu DNA is embedded in host chromosomal DNA. During replication of Mu, DNA cleavage reactions are performed, executed by MuA protein which is the Mu encoded transposase. MuA introduces single strand nicks at both ends of the element. These cleavages expose the 3'-OH ends of the transposon, separating them from flanking bacterial DNA but leaving the transposon covalently linked at its uncleaved 5' ends to flanking DNA. The exposed 3' ends of the transposon are then joined by strand transfer reactions to staggered positions on the target DNA. This transposition product is then replicated by host DNA replication to generate a product called a co-integrate, in which the donor backbone, target and two transposon copies are linked (see fig 1.7 B).

Thus, although the products of retroviral integration and Mu transposition are distinct, their differences arise from the state of the DNA substrate at the 5' ends rather than the actual transposition reaction. Mu remains linked to the donor site because no cleavage occurs at the 5' ends. In contrast, in the retrovirus, the substrate DNA is the transposable element only.

1.3.3.3 'Cut and paste'

During translocation of mobile elements by a 'cut and paste' mechanism, double strand breaks are introduced at the ends of the element to generate an excised transposon that is then inserted into a target site. This paradigm was first established for the bacterial transposons Tn10 and Tn7 (Kleckner *et al*, 1995; Chalmers and Kleckner, 1994; Craig, 1991; Craig, 1995). It is now clear that numerous other elements including the *P* element of *Drosophila*, the *Tc*1 and *Tc*3 elements of *C. elegans* and *mariner* elements use this mechanism.

The chemical reactions at the centre of cut and paste transposition are the same as that in Mu and retrovirus reactions: ie, exposure of the 3'-OH transposon ends, which are

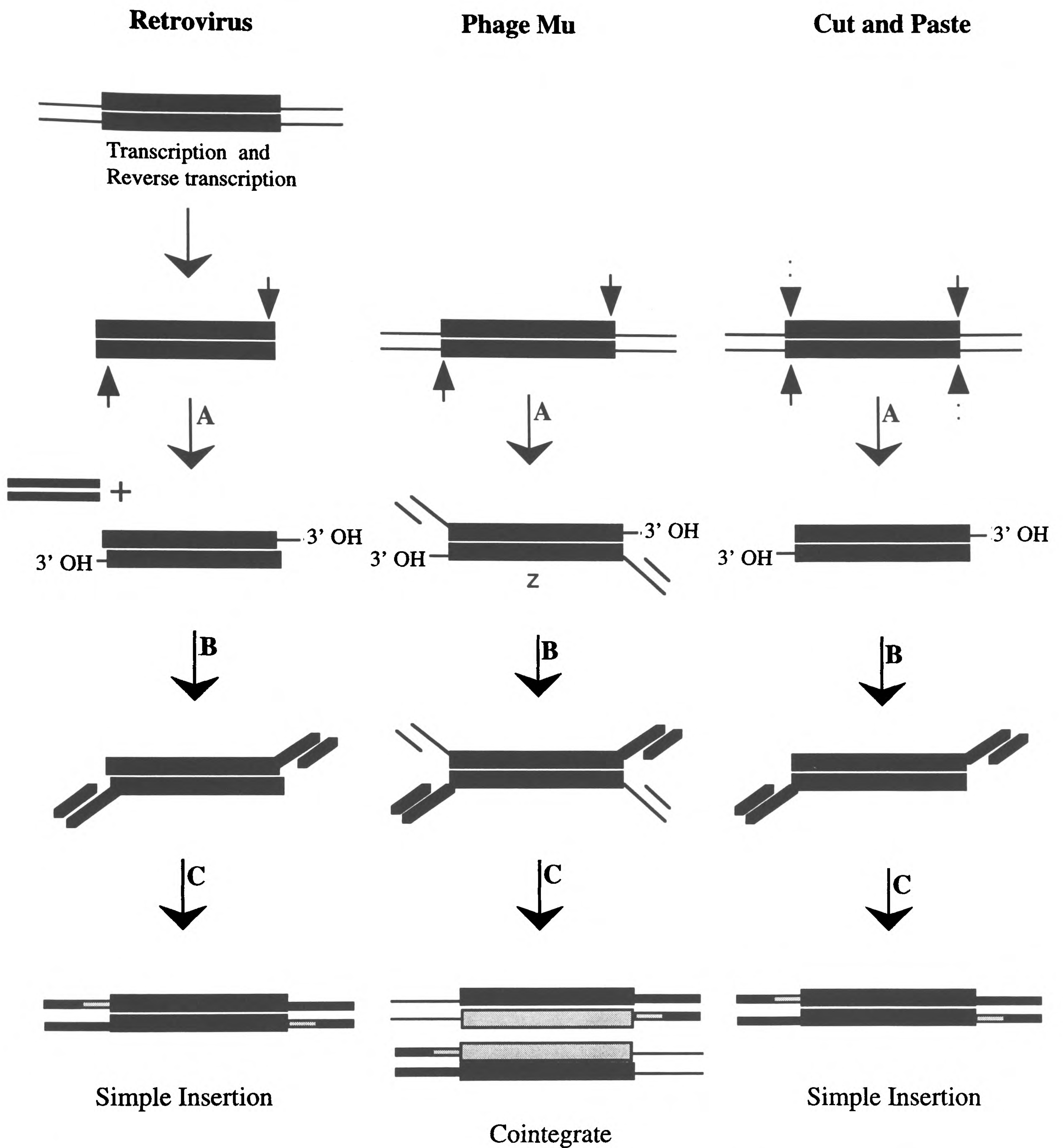


Figure 1.7. The DNA processing reactions that underlie the transposition of different types of mobile elements (adapted from Craig, 1996). Black boxes, mobile elements; thin lines, flanking donor DNA; medium lines, target DNA; small arrows, cleavage at 3' ends; dashed arrows, cleavage at 5' ends; grey shading, DNA replication. **A**. Donor cleavage, **B**. Target joining, **C**. Repair/Replication. See text for detailed description.

then joined to target DNA. However, with cut and paste elements nicking also takes place at the 5' ends of the element. These double strand DNA breaks result in an excised transposon. Transposition is completed by strand transfer reactions that join the 3' transposon ends to the target DNA (fig 1.7 C).

1.3.3.4 Transposition of *Tc*3

The *Tc*3 element from *C. elegans* transposes via a 'cut and paste' mechanism. Excised, linear *Tc*3 molecules have been detected upon forced expression of the *Tc*3 transposase gene under control of a heat shock promoter in *C. elegans*. The appearance of these molecules is coincident with Tc3 transposition and they have been presumed to be intermediates in the transposition reaction (van Leunan *et al*, 1993). The 3' ends of the excised elements are located between the last nucleotide of the *Tc3* sequence and the TA target site duplication and contain a 3' hydroxyl group. The 5' most ends of the transposon which contain 5' phosphate groups are 2 nucleotides shorter than the 3' ends. The last two nucleotides of the *Tc3* sequence at the 5' end appear not to be excised from the chromosome. Based on these findings a model for *Tc*3 transposition has been proposed (van Leunan *et al*, 1994) (figure 1.8).

Transposition of *Tc3* is initiated by the binding of the *Tc3* transposase (and maybe other cellular factors) to the ends of the transposon (van Leunan *et al*, 1993). Double strand breaks with a 2bp stagger are then generated, resulting in an excised element (see fig 1.8). The 3' end which contains a 3'-OH group is then believed to perform a nucleophilic attack during integration into target DNA. The 3'-OH groups are predicted to attack the phosphate bonds on both strands of the target DNA 5' of each T residue of the TA dinucleotide, in a one-step phosphoryl transfer reaction. As the excised element contains the complete *Tc*3 sequence at the 3' ends but lacks the terminal two nucleotides at each 5' end, use of the 3' end for integration prevents loss of transposon sequences. After integration of *Tc*3, the four nucleotide gap at each end of the transposon is repaired by the cellular machinery to produce a complete *Tc*3 element flanked by a duplicated TA sequence (van Leunan *et al*, 1994).

In vitro studies have demonstrated that *Tc1* and *Himar* transposases are the only gene products required for transposition of these elements. Mechanisms for transposition of *Tc1* and *Himar* have recently been proposed based on the model for transposition of *Tc*3. These models also predict a staggered intermediate, in which bases from the 5'

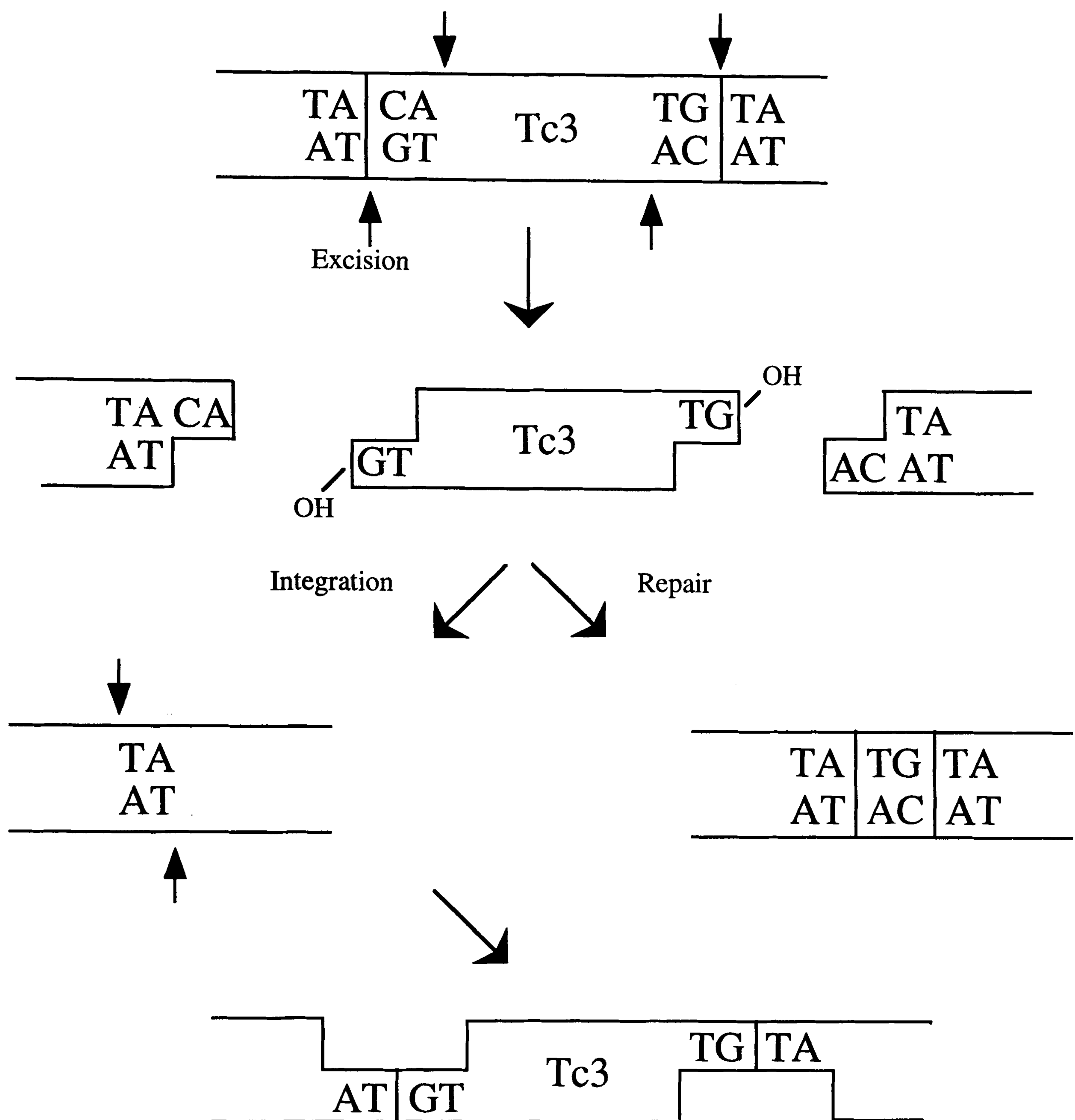


Figure 1.8. Model for *Tc3* transposition (adapted from van Luenen *et al*, 1994).

A schematic model for *Tc3* transposition explaining the excision and integration of element, the target duplication, and the generation of footprints. The model is described in the text. Short arrows indicate the positions of the cleavage sites at the ends of the transposon and also indicate where strand transfer reactions take place during integration.

strand remain behind in the chromosome after excision (Vos *et al*, 1996; Lampe *et al*, 1996).

1.3.3.5 V(D)J recombination: a related reaction

Recent studies of the V(D)J recombination reaction of mammalian immunoglobulin genes have revealed its striking similarity to transposition reactions (Gellert, 1996). The initial part of V(D)J recombination, like transposition proceeds through two endonucleolytic steps (McBlane, 1995). These steps are catalysed by the recombination activating proteins Rag-1 (Schatz *et al*, 1989) and Rag-2 (Oettinger *et al*, 1990). The first step is a nicking event, which takes place at the end of a recombination signal sequence, generating a 3'-OH group. This 3'-OH group is then used in a nucleophilic attack on the phosphodiester bond of the other DNA strand. The difference between V(D)J recombination and transposition is in the second strand transfer reaction, which joins the 3'-OH group to the complementary DNA strand rather than to a new target site. The reactions of V(D)J recombination, like those of transposition proceed by isoenergetic shuffling of phosphodiester bonds without formation of a protein-linked intermediate (van Gent *et al*, 1996).

1.3.3.6 The catalytic mechanism

The similarity of the chemical reactions underlying Mu replication, retroviral insertion and transposition of 'cut and paste' elements suggests that the same or closely related active sites execute these reactions. Indeed, a domain has been recognised which exists in a wide variety of transposons including Mu, Tn10 and Tn7, retroviral and retrotransposon integrases and a number of eukaryotic elements such as *mariner*, Tc1 and *Tc*3. This domain consists of a conserved triad of acidic residues; DDE (see 1.2.2.4).

For functional strand transfer, a Lewis acid is required. This is an electron deficient entity such as Mg^{2+}, which polarises the bonds between non-exchanging oxygen atoms and the phosphorus atom, making it a better target for the nucleophile. The three acidic residues of the DDE motif are believed to be involved in co-ordinating a divalent metal ion and to play a central role in catalysis. Biochemical, genetic and structural studies of members of the DDE superfamily provide strong evidence that the DDE residues play a critical role in both the endonucleolytic cleavage step of transposition,

which exposes the 3' ends of the element, and in the subsequent strand transfer of these ends to target DNA:

Tc3 transposases containing site-directed mutations at amino acids in the DDE motif and at other D and E residues have been assessed for their ability to catalyse excision and transposition reactions. No excision or integration products were observed with transposases containing mutations affecting residues of the DDE triad. Alterations outwith this domain though gave positive results for excision and transposition supporting the notion that the DDE motif contributes to the active site of the *Tc1/mariner* transposases (van Leunan *et al*, 1994).

Site directed mutants of HIV-1 integrase and MuA transposase at the conserved DDE residues are defective in both donor strand cleavage and strand transfer, suggesting that these amino acids are part of the active sites of these proteins also (Engelman and Craigie, 1992; Kulkosky *et al*, 1992; van Gent *et al*, 1992; Baker and Luo, 1994; Kim *et al*, 1995).

EMS induced alleles of *mos*1 transposase which reduce excision of the *peach* element *in vivo*, have been found to map to a small region of the ORF that includes part of the D,D34D motif (Lohe *et al*, 1997; Lohe *et al*, 1996). It is interesting to note that the site-directed change D284E in the final D of the D,D34D motif, which creates the consensus D,D34E as found in *Tc1* like elements is unable to support excision of the *peach* element (Lohe *et al*, 1997)

The structures of the catalytic domains of MuA transposase (Rice and Mizuuchi, 1995) and two retroviral integrases: HIV-1 and ASV (Dyda *et al* 1994; Bujacz *et al*, 1995) have been elucidated at the atomic level by X-ray crystallography. The overall topology of the domains bear structural similarity despite little extended primary sequence homology between them. In both structures the D,D35E amino acids are close together forming a plausible binding site for a metal ion cofactor (Dyda *et al* 1994; Rice and Mizuuchi, 1995).

(i) One active site for two strand transfer reactions: integrase

The two strand transfer events central to retroviral integration differ in the choice of nucleophile. A water molecule is used for the first step in which 3' end cleavage, or end processing, takes place and the 3'-OH of a deoxyribose is used in the second step

for the strand transfer to target DNA. Mutations of the phylogenetically invariant aspartate and glutamate residues of the DDE motif abolish both of these activities. The parallel effects of mutations at these residues on both strand transfer activities, combined with their close, physical proximity (Dyda *et al* 1994; Bujacz *et al,* 1995) point to the existence of a single catalytic site for both of these reactions within each integrase protomer.

Local deformation of DNA structure provides a means by which the active site could catalyse two reactions. Substrates in the two reactions could be distorted such that similarities between the two are maximised. Integrase has recently been demonstrated to be capable of DNA distortion at the ends of the retroviral DNA (Scottoline *et al,* 1997). MuA protein also appears to cause DNA deformation at the termini of Mu DNA (Lavoie *et al,* 1991).

(ii) Two active sites for three strand transfer reactions: Tn7

Tn7 transposon translocates by a 'cut and paste' mechanism. The transposase that catalyses this reaction is a heteromeric complex consisting of two Tn7 encoded proteins: TnsA and TnsB. Site directed mutagenesis of these proteins has demonstrated that TnsA mediates DNA cleavage reactions at the 5' ends of Tn7, and that TnsB mediates DNA breakage and joining at the 3' ends of Tn7. The double strand breaks that underlie excision of Tn7 therefore result from a collaboration between two active sites, one in TnsA and one in TnsB. The same (or a closely related) active site in TnsB also mediates the subsequent joining of the 3' ends to the target (Sarnovsky *et al,* 1997).

(iii) One active site for three strand transfer reactions: Tn10

Tn10 is a composite transposon. It comprises a pair of IS10 insertion sequences that flank a tetracycline resistance determinant (Kleckner, 1989). IS10-right is structurally and functionally intact and encodes a single transposase protein which mediates transposition by interacting with a specific sequence at the Tn10 termini. Tn10 transposition, like that of Tn7, involves three chemical steps at each transposon end. However, with IS10 transposase all three steps utilise a common set of critical amino acid residues. Mutations in any of these residues individually abolishes each of the three chemical steps. Furthermore, the same IS10 monomer molecule is responsible for catalysing strand transfer, as for catalysing DNA cleavage at each end. Thus, a

single active site unit within a single monomer of IS10 transposase catalyses the three chemical steps of Tn10/IS10 transposition in ordered succession at a single transposon end (Bolland and Kleckner, 1996).

1.3.3.7 Higher order nucleoprotein structures

Higher order nucleoprotein-DNA complexes are generally formed as a prerequisite to initiation of the transposition reaction. This requirement prior to transposition prevents partial reactions from taking place that could result in abortive events. Assembly of a synaptic protein-DNA complex is thus a way of achieving a highly accurate mechanism of transposition.

(i) Mu

The mechanism of transposition of phage Mu has been well-studied and the Mu strand transfer reaction duplicated *in vitro* (Craigie *et al*, 1985). For the chemical steps of transposition to take place, MuA and the Mu DNA ends must be assembled into a higher order protein-DNA complex called a transpososome (Craigie and Mizuuchi, 1987; Surette *et al*, 1987). Mu contains six consensus binding sites for the MuA protein, R1, R2 and R3 at the right end and L1, L2 and L3 at the left end. In the transpososome, only three of these sites are occupied by MuA. Mu also contains a transpositional enhancer which is recognised by a site-specific DNA binding domain in MuA distinct from that which recognises the Mu ends (Leung *et al*, 1989; Mizuuchi and Mizuuchi, 1989). Catalytically inactive MuA monomers initially bind to the MuA binding sites. In the presence of *E. coli* HU and IHF and divalent metal ions, conversion of the monomer to a catalyticly active tetrameric form can then take place. This is probably due to a complex set of interactions between the two Mu ends and the transpositional enhancer (Mizuuchi *et al*, 1992; Baker and Mizuuchi, 1992). Mu end synapsis coupled with tetramer formation is believed to induce a structural transition that changes the MuA active site from the inactive to the active configuration (Rice and Mizuuchi, 1995).

(ii) Tn10

Tn10 transposition is non-replicative. The transposon is excised from its donor site by a pair of double-strand cleavages at the ends of the element and subsequently inserts into a new target DNA site. A double strand gap is left behind at the donor site

(Morisato and Kleckner, 1984; Roberts and Kleckner, 1988; Benjamin and Kleckner, 1989; Haniford *et al*, 1991; Chalmers and Kleckner, 1994).

The chemical steps of Tn10 transposition occur within the context of a stable synaptic complex mediated by the transposase and the transposon ends. *In vivo*, interaction of the transposon ends is required prior to and prerequisite to cleavage at either end (Haniford and Kleckner, 1994). A corresponding, stable, precleavage synaptic complex has also been identified *in vitro* (Sakai *et al*, 1995). Within a synaptic Tn10 complex, two monomers carry out the chemical reactions of transposition. Additional monomers may also be present, but if so, play purely structural roles (Bolland and Kleckner, 1996).

(iii) Tn7

Tn7 transposition involves the Tn7 encoded proteins: TnsA, TnsB, TnsC, TnsD and TnsE. All Tn7 recombination reactions require TnsABC gene products. In the presence of TnsABC+D, transposition to attTn7, a specific site in the chromosomes of many bacteria, takes place at high frequency. In the presence of TnsABC+E transposition to many non attTn7 sites is observed.

The Tn7 transposase is a heteromeric complex of two Tn7 encoded proteins TnsA and TnsB (Sarnovsky *et al*, 1997). The two proteins execute different processing reactions. TnsB is a sequence-specific DNA-binding protein which binds to multiple positions at both ends of the Tn7 transposon. TnsB mediates DNA breakage and joining reactions at the 3' ends of Tn7. TnsA mediates DNA cleavage at the 5' ends of Tn7. For transposition to take place into attTn7, all the components of the reaction: TnsABC+D, 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 requirement suggests that the formation of a nucleoprotein complex involving synapsis of the transposon ends both with each other and with the target DNA is a prerequisite for transposition.

1.3.4 Insertion site specificity

Genomic clones from a variety of *mariner* transposons have demonstrated that *mariner* elements have little insertion specificity except for an absolute requirement for the palindromic dinucleotide TA, which is duplicated on insertion. TA is the target site for

a number of other transposable elements including the bacterial insertion sequences IS630 (Tenzen *et al*, 1990) and IS895 (Alam *et al*, 1991), *C. elegans* elements *Tc1* (Moermann and Waterstone, 1989) and *Tc3* (van Leunan *et al*, 1994) and *pogo* from *D. melanogaster* (Tudor *et al*, 1992).

Substitution of Mg^{2+} for Mn^{2+} in *in vitro* transposition reactions, however has been demonstrated to circumvent the absolute requirement of *mariner* transposable elements *himar1* and *mos1* for a TA dinucleotide for insertion. In the presence of Mg^{2+} in the transposition reaction, insertion of *mariner* elements occurs exclusively into TA residues. With Mn^{2+}, the target site specificity is altered (Lampe *et al*, 1996; Dawson, personal communication). The D,D35E motif as discussed above is thought to be involved in co-ordinating a metal ion required for DNA cleavage and strand transfer. It appears that this domain may also be involved in or closely linked to target site selection, which can be perturbed when Mn^{2+} rather then Mg^{2+} is bound (Lampe *et al*, 1996).

Elements with a specific target sequence do not integrate into each possible site with equal frequency. *Tc1* and *Tc3* for example do not integrate into every TA dinucleotide, but have a strong and element specific preference for certain TA dinucleotides (van Luenen and Plasterk, 1994). Alignment of the *Tc1* or *Tc3* integration sites has not revealed a strong consensus sequence for either transposon.

Hot spots for integration have also been observed for elements which do not integrate into a specific sequence. The consensus insertion sequence for *P* elements is GGCCAGAC (O'Hare and Rubin, 1983). However, some hot spots for *P* integration have no homology to this consensus sequence whereas insertions into some potential integration sites, which perfectly fit the consensus sequence have never been isolated (Engels, 1989). Also, the 6-9bp flanking the Tn10 consensus insertion sequence (Halling and Kleckner, 1982) can confer variations of at least one thousand fold in insertion frequency (Bender and Kleckner, 1992). However, no discernible consensus for the Tn10 flanking sequence has been found. These observations suggest that target site specificity relies on more than the primary sequence of the DNA.

DNA secondary structure may be an important consideration in target site selection. DNA secondary structure plots for nucleotides either side of the central TA of a number of different *mariner* insertions *in vivo* has revealed the presence of hairpin structures in which the TA residues are exposed (Bigot, unpublished observations). It

may be that certain DNA secondary structures render some TA dinucleotides more accessible to the recognition machinery than others, thus resulting in hot spots for insertion.

The functional state of the DNA (eg. transcription, replication) can affect the distribution of integration sites. *P* element insertions are preferentially found at or near transcription start sites (Tsubota *et al*, 1985; Kelley *et al*, 1987). The Tn5 transposon integrates more frequently into regions with increased negative supercoiling (Isberg and Syvenan, 1982; Datta and Rosner, 1987), and in the vicinity of promoter regions of transcribed genes (Mckinnon *et al*, 1985; Berg *et al*, 1983). Insertions of the yeast elements Ty1 and Ty3 are preferentially into regions containing tRNA genes, long terminal repeats or pre-existing transposable elements (Sandmeyer *et al*, 1990; Ji *et al*, 1993), and Tn10 and Mu preferentially integrate into non transcribed regions (Casadesus and Roth, 1989; Pato, 1989).

1.3.5 Donor site repair after excision

The excision of class II transposable elements from chromosomes results in the generation of double strand DNA breaks which must be repaired to prevent chromosome loss. The breaks can be repaired by two alternative mechanisms. First, by repair of the broken chromosome through 'gap repair' or gene conversion using the homologous chromosome, sister chromatid or ectopic copies on other chromosomes as a template, and secondly by ligation of the broken ends followed by mismatch repair.

Repair of double strand breaks following *Tc1* excision in *C. elegans* can occur by gene conversion. The frequency of loss of *Tc1* elements from the *unc22* gene depends on the sequence of the homologous chromosome. Heterozygous *Tc1* mutations revert at a high frequency, with precise loss of the element. Homozygous *Tc1* mutants however, revert 100 times less frequently and a leave a 'footprint' of *Tc1* sequence behind. In homozygotes the transposon is usually copied back into the excised site. Loss of the transposon is only seen in rare cases in which repair is interrupted and a *Tc1* footprint is left behind (Plasterk, 1991; Plasterk and Groenen, 1992). Excision sites remaining after infrequent mobilisation of marked *mos*1 sequences in *D. melanogaster* have been observed to repair via a gene conversion mechanism (Lohe and Hartl, 1996).

P element induced breaks in *D. melanogaster* are mainly repaired by gene conversion (Engels *et al*, 1990). Failure of the repair process to go to completion may account for the frequent occurrences of internally deleted *P* elements.

In most cases, somatic excision of *Tc1* leaves footprints which consist of 2bp from the 5' or 3' end of the transposon flanked by the TA duplication (Ruan and Emmons, 1987; Eide and Anderson, 1988). Similarly, *Tc3* leaves 2bp footprints behind (van Leunan *et al*, 1994). The most common footprint after *mariner* excision in *D. mauritiana* consists of 3bp from one or other end of the transposon flanked by the TA target site duplication (Bryan *et al*, 1990). Staggered cuts within the transposon ends have been proposed to explain the origin of these footprints (Eide and Anderson 1988; van Luenen *et al*, 1994; Bryan *et al*, 1990) and the structure of linear, excised *Tc3* elements *in vivo* is compatible with this (van Luenen *et al*, 1994).

Excision of the *hobo* element from *D. melanogaster* results in repaired donor sites which more closely resemble those left by the excision of *Ac* from maize and *Tam3* from snapdragon chromosomes, than those reported from other transposable elements in insects. *Hobo* excision in *D. melanogaster* results in the complete removal of all *hobo* sequence and the addition of nucleotides at the resulting breakpoints. These additional nucleotides appear to be duplications of the flanking genomic sequence (Atkinson *et al*, 1993).

1.4 Regulation of transposition

High frequencies of transposition result in increased frequencies of mutation and reduced viability and fertility in the host genome. Regulation of the activity of transposable elements is thus necessary for the long term persistence of the element and the genome within which it resides.

1.4.1 Tissue specificity of transposition

P element transposition is tissue-specific. Together the 4 exons found in autonomous *P* elements produce an 87kD protein required for *P* element mobility (Karess and Rubin, 1984; Rio *et al*, 1986). Tissue-specific splicing of the transposase pre-mRNA ensures that the elements are mobile only in the germline (Laski *et al*, 1986). In somatic tissues only two of the three *P* element introns are spliced out creating mRNA which, when translated yields a protein of 66kD. This protein is a negative regulator of

P element transposition (Kaufman, 1989). In germline tissues an additional splicing event removes the third intron to produce full length mRNA which expresses the 87kD transposase protein. This tissue specificity can be eliminated by the artificial removal of the 2-3 intron to create a delta 2-3 element (Laski *et al*, 1986). The putative splicing inhibitor, *P* element somatic inhibitor protein (PSI) of the 2-3 intron in somatic cells has recently been identified (Adams *et al*, 1997). Somatic inhibition of the third intron involves the assembly of a multiprotein complex on a regulatory sequence in the exon 5' to the intron. The PSI is a component of this ribonucleoprotein complex.

Activity of the *hobo* element is restricted to the germline. This is due to regulation of transposase production at the level of transcription as evidenced by the ability of the *hsp70* promoter to direct *hobo* activity in the soma, and for *hobo* sequences to direct germline specificity of the *P* element delta 2-3 gene (Calvi and Gelbart, 1994). Thus, although *hobo* is similar to the *P* element at the level of organisation and tissue specificity, it differs in the underlying mechanism governing the germline specificity.

In the nematode *C. elegans*, somatic and germline activity of *Tc1* and *Tc3* has been observed. *Tc1* and *Tc3* excision is detectable in somatic cells of the common laboratory strain Bergerac TR679 (Emmons and Yesner, 1984; Collins *et al*, 1989), and *Tc1* excision also in strain Bristol N2. In the Bergerac strain, *Tc1* excisions in somatic cells occur about 1000x more frequently per cell than in the germline (Emmons and Yesner, 1984; Eide and Anderson, 1988). Excision of *Tc1* from its site of insertion generates 'empty sites' which, because of the very high frequency of somatic excision, can be detected on genomic Southern blots as restriction fragments of wild type mobility (Emmons and Yesner, 1984). Somatic excision of *Tc3* does not occur at frequencies comparable to *Tc1* (Collins et al, 1989). Germline transposition of these elements is restricted to the Bergerac strain, in which *Tc1* is responsible for most spontaneous mutations generated.

Excision of a *Tc1* element from the *unc54* gene in the germline of an EMS mutagenised isolate of the Bristol strain has recently been reported (Plasterk, Keystone symposium). The EMS induced mutation, *Mut-7*, is a mendelian segregating gene, which alleviates germline repression of *Tc1* excision in the Bristol strain. *Mut-7* is believed to be a general regulator of germline excision and transposition, since germline activity of the non-complementary elements *Tc3*, *Tc4* and *Tc5* is also derepressed.

Mariner activity has been observed in both somatic and germline cells. *In vivo* activity has been assayed using excision of the *peach* element from a *white peach* transgene on the X chromosome. Excision restores the ability of the gene to produce eye pigmentation. In the presence of active transposase, excision of the *peach* element restores wild type expression of the w^{pch} transgene. In the soma, excision results in eye- colour mosaicism; in the germline, it results in a reverse mutation of w^{pch} to wild type. In general the level of somatic excision observed in a strain is positively correlated with the level of germline excision observed in the same strain (Jacobson and Hartl, 1985).

Transposition of *mariner* in the germline has also been detected. A *white* marked *mariner* element present on chromosome 2 has been observed to jump to chromosome 3, on overexpression of the *mos* transposase behind an *hsp26-sgs3* promoter enhancer cassette. Transposition and excision events of this element though are rare (Lohe and Hartl, 1996).

1.4.2 Regulation of *P* element transposition

P element transposition is cytotype dependent. The progeny of certain crosses exhibit a phenomenon called hybrid dysgenesis. This is a syndrome of abnormal traits including gonadal dystrophy, high mutability, segregation and distortion sterility exhibited when males of a *P* (paternally contributing) strain are mated with females of an M (maternally contributing) strain. P strains (P cytotype) differ from M strains by their possession of multiple *P* elements which produce dysgenesis when placed in the maternally derived background of an M strain (M cytotype).

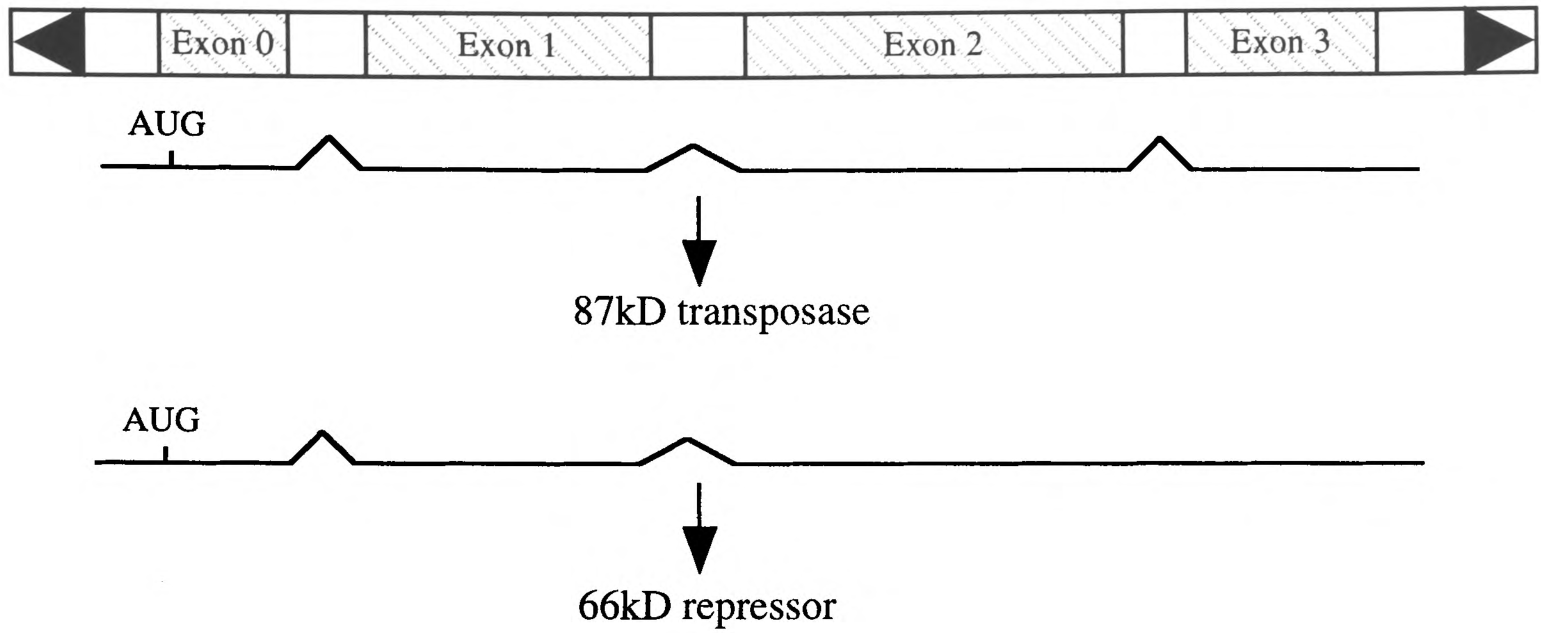
Alternative splicing of full length *P* elements results in either production of 87kD transposase or the 66kD repressor, as described earlier. The 66kD protein is colinear with the amino terminal of the 87kD transposase protein and contains all the first three *P* element exons. *P* element derivatives with internal deletions can also encode proteins that repress transposition. One of the most commonly found deletion derivatives encodes a repressor protein known as *KP*. This element fuses the second and fourth exons out of frame. The resulting ORF encodes a 207 amino acid protein, the first 199 of which are identical to the N terminus of the transposase (see figure 1.9).

P element mobility is regulated on at least two levels by *P* element encoded polypeptides. The *P* element encoded proteins *KP* and 66kD repressor have been

shown to repress transcription from a *P* element promoter *in vivo* (Lemaitre and Coen, 1991; Lemaitre and Coen, 1993), presumably by preventing assembly of an RNA polymerase II complex at the *P* element promoter. This transcriptional regulation can be inherited maternally, accounting for some of the reciprocal cross differences seen in the matings between P and M strain flies (Lemaitre *et al*, 1993; Ronsseray *et al*, 1993). Alternative splicing of the *P* transcript also explains the germ line specific phenotypes associated with hybrid dysgenesis.

A second level of regulation of *P* element mobility been demonstrated. Here, regulation acts to repress transposase activity, even when transposase is directed by a heterologous promoter (Steller and Pirotta, 1986). Purified *KP* protein is able to dimerise *in vitro*, and can bind to multiple sites on the ends of *P* element DNA with high affinity. These sites include the high affinity transposase binding site, an 11bp transpositional enhancer and terminal 31bp inverted repeats (Lee *et al*, 1996). Binding of *KP* to these sites may compete with transposase, thus blocking assembly of an active transposase multimer and inhibiting transposition. This mechanism of inhibition is used by phage Mu. Mu transposase monomers bind to multiple recognition sites on Mu DNA, leading to the formation of a higher order active tetrameric nucleoprotein complex. Mu repressor inhibits transposition by binding to one of these sites, the internal activation sequence, and blocks assembly of the catalytically active transposase tetramer (Leung *et al*, 1989; Mizuuchi and Mizuuchi, 1989). The bacterial insertion sequence IS1 also produces a repressor protein InsA, that inhibits transposition by binding to specific sites on the transposon DNA (Zerbib *et al*, 1990a; Zerbib *et al*, 1990b).

Full length *P* element



***KP* element**

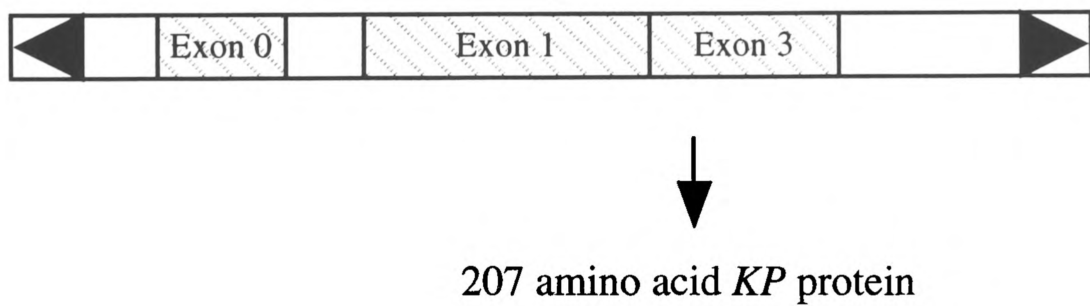


Figure 1.9. *P* element encoded proteins

Genetic experiments have suggested that the *KP* protein may inhibit transposition through heteromultimer formation with *P* transposase (Gloor and Andrews, 1995). Transgenic fly strains containing *KP* proteins with mutant leucine zipper domains were found no longer to repress *P* element mobility *in vivo* (Lee *et al*, 1996). A similar mechanism of repression has been proposed for Tn5 which produces an inhibitory protein that forms heterodimers with transposase (de la Cruz et al, 1993; Reznikoff, 1993).

1.4.3 Regulation of *mariner* activity

Genetic studies of the *mos*1 *mariner* element in *Drosophila* have revealed the presence of at least two putative regulatory mechanisms controlling *mariner* activity.

Overproduction of the wild type *mos*1 transposase by introducing more than one copy of an *hsp70:mos*1P(promoter):*mos*1ORF construct or by heat shocking reduces the overall level of transposase activity as assayed by the excision of the non autonomous *peach* element from the w^{pch} allele (Lohe and Hartl, 1996). This mechanism is termed overproduction inhibition (OPI). The mechanism of overproduction inhibition has not been determined, but studies of another *hsp70:mos*1P construct suggest that it is not mediated by direct, mutually inhibitory effects of transposase subunits. A biological assay system using the maternal effect of a transgene, *hsp70:mos*1P:*Cre*, in which the Cre site specific recombinase is driven by the *hsp70:mos*1P chimeric promoter has been employed. In the presence of maternally transmitted Cre, a *mini-white* reporter gene flanked by *loxP* recombination sites is deleted, resulting in white eyes (Siegal and Hartl, 1996). The extent of maternal Cre transmission can be regulated by the presence of the *hsp70:mos*1P:*mos*1ORF. In the presence of two copies of *hsp70:mos*1P:*mos*1ORF, compared to one or no copies, maternal transmission of Cre results in a 22% reduction of the number of white eyed offspring (Hartl *et al*, 1997). The *hsp70:mos*1P:*Cre* construct includes both the 5' end of *mos*1 (nucleotides 58-172) and the 3' end (nucleotides 1185-1286), hence overproduction inhibition could be mediated by either the 5' or 3' ends of *mos*1 or both.

The implication of OPI for *mariner* regulation is that as the number of active copies within an organism passes an optimum at which the rate of transposition per element is maximised, then the overall transposase activity will decrease and the rate of increase in copy number will decrease over time. High levels of expression of an element will need less copies for OPI. Different levels of expression in different genomes may help

explain why some species have only a few copies of an element whereas others have thousands of copies.

Another mechanism of regulation has been observed in a class of hypomorphic missense mutations in the *mos*1 transposase. In the presence of wild type *mos*1 transposase, these mutations exhibit dominant negative complementation. Possibilities for this heteroallelic effect include competition of the mutant proteins with wild type transposase for DNA binding sites on the transposon and the formation of oligomers of reduced activity that contain both wild type and defective subunits. Results form the yeast two-hybrid system imply that *mos*1 subunits can interact, suggesting the possibility of subunit oligomerisation (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).

The majority of *mariner* like elements do not produce an active transposase even though some, such as the *peach* element have an intact ORF. These elements are candidates for regulation through dominant negative complementation with wild type transposase in nature. Defective *mariner* elements may also play a role in regulation through titration of transposase if they have retained their transposase binding sites. A particular deletion is common in natural populations of *D. teissieri*, which eliminate about one half of the 3' end of the ORF. This deleted element has been proposed as a candidate for a titrator element, the part of the ORF remaining may also play a role in *mariner* regulation similar to the *KP* element in *P* element regulation (Brunet *et al*, 1996).

A system of regulation analogous to cytotype determination for *P*, has not been reported thus far for *mariner* or for any other members of the *Tc1/mariner* superfamily. Likewise, no syndrome of hybrid dysgenesis mobilising a *mariner* element has been reported.

1.5 Transposable Elements as Experimental Tools

Transposable elements are valuable experimental tools for molecular genetics in many organisms They serve as agents of chromosomal insertion, deletion and rearrangement, and provide the basis for transformation of somatic or germline cells.

In *Drosophila* exogenous DNA has not been shown to replace endogenous sequences by recombination. A different system for introducing modified sequences at ectopic sites in the genome, employing transposable elements has been developed. Several class II transposable elements in insect genomes have been characterised in detail and evaluated as potential germline transformation vectors, including *P* (Engels, 1989), *hobo* (Blackmann and Gelbart, 1989; O'Brochta *et al*, 1994), *minos* (Loukeris *et al*, 1995a; Loukeris *et al*, 1995b) and *mariner* (Lidholm *et al*, 1993). The importance of transforming species outside of drosophilids has been highlighted as a technical development which is essential for a genetic programme of insect control (Ashburner, 1995). Germline transformation with transposable elements would enable the genetic manipulation of organisms that are disease vectors or agricultural pests. The potential ability of transposable elements to sweep through populations may enable the genetic manipulation of natural populations.

1.5.1 Use of the *P* element

The paradigm for germline transformation is the use of the *P* element to transform the germline of *D. melanogaster* (Rubin and Spradling, 1982). Here, the transformation vector consists of a *P* element with a large internal deletion within which is inserted the DNA sequence of interest. The defective *P* element containing the sequence of interest is introduced into the genome by transposition mediated by an active, 'helper' *P* element co-injected into the embryonic germline. This transformation system has been widely used for identifying functional genetic units in *D. melenogaster.* Another type of engineered *P* element, the enhancer trap, has been widely used to identify genomic insertions near genes expressed only at particular times or in particular tissues (O'Kane and Gehring, 1987, Bellen *et al*, 1989, Wilson *et al*, 1989). The enhancer trap contains the *lac Z* gene from *E. coli* under the control of the weak *P* element promoter, β-galactosidase is expressed only when the element transposes to a site in the genome near a strong enhancer. An analogous system of germline transformation has been developed based on the *hobo* element (Blackman and Gelbart, 1989).

The host range of *P* and *hobo* elements appears to be limited to *Drosophila* and numerous attempts to transform species outside the genus *Drosophila* using the *P* element system have been unsuccessful (Engels, 1989). In *Drosophila* a host encoded inverted repeat binding protein (IRBP) binds specifically to the outer halves of the *P* element inverted repeats. Because the inverted repeats are absolutely required for transposition and are not bound by the *P* element encoded transposase, IRBP is a

candidate for a host factor involved in transposition. This may be why *P* element activity is limited in its host range to the *Drosophilidae* (Beall *et al*, 1994).

1.5.2 Harnessing the *mariner* element

The *mariner* element however, looks increasingly promising as a tool for molecular biology due to its extraordinarily broad host range. Extremely similar *mariner* elements can occupy the genomes of species even in different phyla, indicating that these elements were recently horizontally transferred into their genomes (Robertson, 1993; Robertson and Macleod, 1993; Garcia-Fernandez *et al*, 1995; Lohe *et al*, 1995). An example of this phenomenon occurs in the irritans subfamily of the *mariner* elements, where two flies in different suborders (>200 million years diverged) and a green lacewing (>265 million years diverged from the flies) each contain *mariner* elements whose consensus sequences encode transposases that differ from each other by only six amino acids out of 348 (Robertson and Lampe, 1995). It appears, from this type of data that *mariner* transposition relies solely on the *mariner* transposase and not species specific host factors beyond those necessary to transcribe and translate the transposase and host repair enzymes necessary to repair single stranded gaps at the sites of transposon insertions. Indeed this suggestion is borne out experimentally, by the demonstration that *Himar*1 transposase is the only hornfly encoded protein required for transposition of the *Himar* element in an *in vitro* environment (Lampe *et al*, 1996), and *mos*1 transposase, the only *Drosophila* encoded protein for *in vitro* mobilisation of *mos*1 (Dawson, personal communication).

The ability of *mariner* elements to undergo horizontal transfer across a wide range of insect species, and the apparent independence of host factors in the transposition reaction bear favourably in the development of *mariner* as a versatile system for germline transformation.

Mariner elements are capable of integrating and being maintained in genomes in which they are not normally present. Initially, the w^{pch} allele and the *mos*1 element of *D. mauritiana* were introduced into the germline of *D. melanogaster* using the *P* element mediated system of transformation. Transformed lines containing the w^{pch} allele are stable as the *peach* element is inactive. In the presence of *mos*1, excision of the *peach* element occurs resulting in eye colour mosaics and germline revertants (Garza *et al*, 1991).

*Mos*1 is able to insert autonomously into the genome of *D. melanogaster* on injection into the pole plasm of *Drosophila* embryos. Over 30% of fertile G_O flies crossed individually to *w^pch^* *D. melanogaster* yielded mosaic offspring. Southern blot analysis of subsequent generations of these lines revealed multiple *mos*1 containing bands indicating that the integrated *mos*1 element is capable of active transposition (Garza *et al*, 1991).

*Mos*1 is also able mediate the transposition of *mos*1 based transformation vectors (Lidholm *et al*, 1993). Vector pM1wB, which contains the *mos*1 element disrupted by insertions containing the wild type *white* gene from *D. melanogaster*, the *LacZ* gene from *E. coli,* and sequences that enable plasmid replication and selection in *E. coli* was introduced into the germline of *D. melanogaster* by co-injection with an active *mos1* helper plasmid. The M1wB element was found to insert in the genome in a *mos1* dependent manner. The termini of the inverted repeats inserted at a TA dinucleotide.

In the presence of active transposase, mobilisation of M1wB and other *mos*1 based constructs occurs at frequencies lower than 0.1% per chromosome per generation (Lohe and Hartl, 1996). The stability of the *mos*1 derived insertions are not ideal for genetic manipulation. However, they may be of considerable significance in assessing the practicability of *mos*1 as a general transformation vector in the implementation of schemes in which stable transformants are required such as insect population regulation.

The potential of *mariner* to serve as a general transformation vector in insects has been demonstrated by the transformation of *mos*1 into *D. virilis*, a species that last shared a common ancestor with *D. melanogaster* >40million years ago (Lohe and Hartl, 1996). Integration of *mos*1 into *D. virilis* is accompanied by duplication of the TA target sequence site and transfer of DNA sequences bound only by the inverted repeats. The ability of *mariner* elements to excise and transpose in non drosophilids has also been reported. Plasmid based assays in embryonic soma of the Australian sheep blowfly *Lucilia cuprina* and the Queensland fruitfly *Bactrocera tryoni* have been performed. Excision products show a 3bp footprint at the excision site as is observed at empty sites in *D. mauritiana* chromosomes (Coates *et al*, 1995), and transposition products recovered are identical in structure to *mariner* transposition events observed in the chromosomes of *D. melanogaster* and *D. mauritiana* (Coates *et al*, 1996). These results suggest that a gene vector system constructed form the *mos*1 element, should

be capable of producing germline transformants of these and possibly other non-drosophilid insects.
The *minos* element from *D. hydei* is a member of the *Tc1/mariner* superfamily. *Minos* has likewise been used as a gene vector to genetically transform *D. melanogaster* (Loukeris *et al*, 1995a). The genome of the mediterranean fruit fly *Ceratitis capitata* has also been successfully transformed by the *minos* element (Loukeris *et al*, 1995b).

It appears that members of the *Tc1/mariner* superfamily in general share an ability to function in heterologous cellular environments and are thus promising for the future development of germline transformation vectors functional over a large range of species.

1.6 Scope of thesis

As outlined above, the majority of information about *mariner* elements has come from studies of their structure and distribution among different insect species. At the outset of this project, details concerning the 'cut and paste' mechanism by which these elements are believed to transpose, were largely inferred from the way in which other short inverted repeat type elements transpose. The aim of this PhD thesis was to look at the mechanism of *mariner* transposition, by analysing possible intermediates in the transposition reaction. Experiments were designed such that *mariner* elements mobilised by transposase both *in vitro* and *in vivo* could be examined.

Chapter 2

Materials and methods

2.1 Media

2.1.1 Bacterial Media

Luria Broth (L-broth):

Difco Bacto tryptone, 10g; Difco Bacto yeast extract, 5g; NaCl, 5g; per litre adjusted to pH 7.2.

Luria Agar (L-agar):

Difco Bacto tryptone, 10g; Difco Bacto yeast extract, 5g; NaCl, 5g; Difco agar, 15g; per litre.

2.1.2 *Drosophila* Media

French fly food:

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

Egg collection media:

Difco Bacto Agar, 1.5g; pure apple juice, 100ml.

2.2 Materials

2.2.1 Chemicals

All chemicals used were products for molecular biology or analytical grade chemicals. They were supplied from Sigma chemical company, Fisons, BDH, Boehringer Mannheim.

2.2.2 Enzymes

All restriction enzymes, Klenow enzyme, T4 ploynucleotide kinase and T4 DNA ligase and buffers were purchased from Boehringer Mannheim. Taq Polymerase was supplied from Gibco and DNaseI from Sigma.

2.2.3 Isotopes

α-^{32}P-dCTP (3000 Ci/mM).
α-^{35}S-dATP (400 Ci/mM).
γ- ^{32}P-dATP (3000 Ci/mM).

All isotopes were supplied by Amersham International.

2.2.4 Antibiotics

Ampicillin to a final concentration of 100μg/ml, chloramphenicol to 34μg/ml, kanamycin to 25 μg/ml and tetracycline to 12 μg/ml were added to media prior to use as required.

2.2.5 Competent cell preparation solutions:

Competent cells buffer 1:
33mM CH_3COOK, pH 7.5; 100mM RbCl; 50mM $MnCl_2.4H_2O$; 10mM $CaCl_2.2H_2O$; glycerol, 15% (v/v); pH to 5.8 with 200mM acetic acid; made up to 1l with dH_2O.

Competent cells buffer 2:
25mM MOPS, pH 6.8; 10mM RbCl; 75mM $CaCl_2.2H_2O$; glycerol 15% (v/v); pH to 6.8 with KOH; made up to 1l with dH_2O.

2.2.6 DNA electrophoresis solutions:

DNA loading buffer:
30% glycerol; 0.1% bromophenol blue; made up to 10ml with dH_2O.

20x TAE:
0.8M Tris ; Glacial acetic acid, 22.8ml; 20mM EDTA pH8.0; made up to 1l with dH_2O.

10x TBE:

0.9M Tris; 550mM Boric Acid; 50mM EDTA; made up to 1l with dH_2O.

TE.
10mM Tris; 50mM EDTA.

2.2.7 DNA minipreparation solutions:

Resuspension buffer:
25mM Tris, pH 8.0; 10mM EDTA; 50mM glucose.

Lysis solution:
200mM NaOH; 1% (w/v) SDS

2.2.8 DNA extraction from *Drosophila* solution:

Solution A:
100mM Tris.HCl, pH 9.0; 100mM EDTA, pH 8.0; 1% (w/v) SDS.

2.2.9 Southern blot and colony hybridisation solutions:

Denaturation buffer:
0.5M NaOH; 1.5M NaCl.

Hybridisation solution:
0.5M NaP (1M Na_2HPO_4, 72ml; 1M NaH_2PO_4, 28ml); 7% (w/v) SDS; 1mM EDTA pH 8.0.

Neutralisation buffer:
3M NaCl; 1M Tris pH 7.5.

20x SSC:
3M NaCl; 300mM tri-Na Citrate.

Wash solution:
40mM NaP (as for hybridisation solution); 1% (w/v) SDS; 1mM EDTA; made up to 1l

2.2.10 Protein electrophoresis solutions:

Boiling Mix:
13mM Tris, pH 8.9; 6.6% (w/v) SDS; 31% (v/v) glycerol; 0.4M 2-mercaptoethanol,; 0.05% (w/v) bromophenol blue.

Destain:
10% (v/v) methanol; 10% (v/v) acetic acid.

Resolving gel buffer:
150mM Tris; 0.4% (w/v) SDS; pH to 8.9 with HCl.

10xPBS:
1.36M NaCl; 100mM Na_2HPO_4; 20mM KCl; 17mM KH_2PO_4.

Stacking gel buffer:
42mM Tris; 0.4% (w/v) SDS; pH to 6.7 with HCl.

Stain:
50% (v/v) methanol; 10% (v/v) acetic acid; 2% (w/v) Coomassie brilliant blue R-250.

10X TGS:
250mM Tris; 1.9M Glycine; 1% (w/v) SDS.

2.2.11 Protein purification solutions for Mos1 transposase:

Column wash buffer:
25mM Tris pH 7.5; 8M urea; 5mM DTT; 10% (v/v) glycerol; 50mM NaCl.

Detergent buffer:
20mM Tris pH 7.5; 4mM EDTA; 0.2M NaCl; 1% (v/v) deoxycholate; 1% (v/v) NP40; 1mM DTT.

Diluent buffer:
25mM Tris pH 7.5; 8M Urea; 5mM DTT; 10% (v/v) glycerol.

Guanidine buffer:
6M Guanidine Hydrochloride; 20mM Tris pH7.5; 5mM DTT.

Resuspension buffer:
20mM Tris pH7.5; 10% (v/v) Glycerol; 2mM $MgCl_2$; 1mM DTT.

Urea gradient buffer, 8M:
25mM Tris pH 7.5; 8M urea; 100mM NaCl; 5% (v/v) glycerol; 5mM DTT.

Urea gradient buffer, 0M:
25mM Tris pH 7.5; 100mM NaCl; 5% glycerol; 5mM DTT.

Wash buffer:
0.5% (v/v) NP40; 1mM EDTA.

2.2.12 Protein purification solutions for 6-his-Mos1 protein:

Buffer 1:

50mM Tris pH 7.4; 8M urea; 0.5M NaCl; 20% (v/v) glycerol; 100mM $MgCl_2$; 2mM 2-mercaptoethanol; 1mM Imidizole.

Buffer 2:

50mM Tris pH 7.4; 0.5M NaCl; 20% (v/v) glycerol; 100mM $MgCl_2$; 2mM 2-mercaptoethanol

Imidazole buffer:

50mM Tris pH 7.4; 0.5M NaCl; 20% (v/v) glycerol; 100mM $MgCl_2$; 2mM 2-mercaptoethanol; 250mM Imidizole.

Urea gradient buffer, 6M:

50mM Tris pH7.4; 6M urea, 0.5M NaCl; 20% (v/v) glycerol; 100mM $MgCl_2$; 2mM 2-mercaptoethanol.

Urea gradient buffer, 0M:

50mM Tris pH7.4; 0.5M NaCl; 20% (v/v) glycerol; 100mM $MgCl_2$; 2mM 2-mercaptoethanol.

2.2.13 Western blot solutions:

10x TBS:

0.5M Tris; 1.5M NaCl, pH to 7.5 with HCl.

1xTBS-T:

10xTBS, 100ml; Tween 20, 1ml; made up to 1l with dH_2O.

Western blot transfer buffer:

0.63M Glycine; 80mM Tris.

2.2.14 Excision assay solution:

25mM HEPES pH7.9; 100mM NaCl; 10% (v/v) glycerol; 2mM DTT.

2.2.15 Bacterial Strains:

Name	Genotype	Use	Reference
BL21 (DE3)	*F⁻ompT[lon]* *hsdS_B*($r_B^- m_B^-$), with DE3, a λ prophage carrying the T7 RNA polymerase gene	Strain used for over-expression of recombinant proteins.	Studier *et al*, 1990.
BL21 (DE3) pLysS	As above, plus plasmid pLysS carrying the T7 lysozyme gene.	Used for over-expression of recombinant proteins.	Studier *et al*, 1990.
NM522	F'*lacI^{q}Δ(lacZ)M15 proA+B+/supE thiΔ((lac-proAB)Δ(hsdMS-mcrB)5*($r_k^- m_k^-$ McrBC⁻)	Used for making competent cells for general cloning applica tions.	Gough and Murray (1983).

2.2.16 Plasmids:

Name	Description and Use	Reference
CaSpeR	Contains a functional copy of the 'mini-*white*' gene from *D.melanogaster.*	Klemenz *et al*, 1987.
pBCP378	For direct, over-expression of recombinant proteins from a *trc* promoter.	Velterop *et al*, 1995.
pBluescribe	General cloning applications.	Stratagene catalogue.
pBY	pBluescribe containing a functional copy of the *yellow* gene from *D. melanogaster.*	Gift from I. Busseau. Geyer and Corces, 1987.

pet13a	For direct, over-expression of ORFs from a T7 promoter.	Gift from D.Lampe.
pGEM-T	TA cloning vector, for direct cloning of PCR products.	Promega catalogue.
pGEX-2T	For over-expression of recombinant proteins as GST-fusions. Contains a thrombin cleavage site.	Pharmacia catalogue.
pGEX-3X	For over-expression of recombinant proteins as GST-fusions. Contains a Factor X_a cleavage site.	Pharmacia catalogue.
p*mos*	pBluescript containing the *mos*1 element present in a 5kb genomic clone.	Medhora *et al*, 1991.
pπ25.7wc	Helper plasmid used in *P*-element transformation.	Karess and Rubin, 1984.
pRP1406	*P* element transformation vector containing the *Drosophila vermillion* gene, and a *Tc3* element with a *yellow* gene insertion into the ORF.	Gift from R. Plasterk.
pRP1386	*P* element transformation vector with a *rosy* marker gene, containing the *Tc3* ORF downstream of the *D. Melanogaster hsp26 sgs3* promoter enhancer cassette.	Gift from R. Plasterk.
pRSETC	For over-expression of 6-histidine tagged fusion proteins.	Invitrogen catalogue.

2.2.17 *Drosophila melanogaster* Stocks:

Name	Genotype	Reference
Oregon R	wild-type	Lindsley and Grell, 1968.
v^{36f}; ry^{506}	v^{36f}; ry^{506}	Gift from I. Busseau.
y^1 w^-; *CyO/Bl*; ry^{506}	y^1 w^-; *CyO/Bl*; ry^{506}	Gift from I. Busseau.
y^1 w^-; *D/Sb* ry^{506}	y^1 w^-; *D/Sb* ry^{506}	Gift from I. Busseau.
w^{1118}	*white*$^-$	Hazelrigg *et al*, 1984.
y^1; ry^{506}	y^1; ry^{506}	Gift from I. Busseau.

2.2.18 *Drosophila simulans* Stocks:

Name	Genotype	Reference
238	wild-type, contains active *mariner* elements	Gift from P. Capy
244	Wild type, contains a single inactive *mariner* element inserted into the 5' untranslated region of the *white* gene	Gift from P. Capy
yw	*yellow*$^-$ *white*$^-$	Gift from P. Capy

For explanation of gene symbols see Lindsley and Zimm, 1992.

2.2.19 Oligonucleotides

Name	Sequence	Comments
337	**cgggatcc**tgtcgagtttcgtgcc	"+" strand oligonucleotide containing the first 16 bases of the Mos1 ORF.
338	**gcgaattc**ttattcaaagtatttgcc	"-" strand oligonucleotide containing 18 terminal most bases of the Mos1 ORF.
AJ1	cgcggtaccgtatcgtttcatcca	primer containing the first bases of the atonal ORF.
AJ2	ataggtaccgggaattcagcgcagc	primer containing the terminal bases of the atonal ORF.
M4269	ataggtggtacacggatggggcc	"-" strand primer, from 1033 to 1010 in *mos*1
M4270	ggccccatccgtgtaccacctat	'+' strand primer, from 1010 to 1033 in *mos*1.
N6799	**gccat**atgtcgagtttcgtgcc	"+" strand oligonucleotide containing first 17 bp of Mos1 ORF.
P3855	cccatccgattaccacctattc	"+" strand primer corresponding to nucleotides 1014 to 1035 of *mos*1.
P3856	cacagttggtacttgttcgcc	"-"strand primer corresponding to nucleotides 300 to 280 of *mos*1.
P4795	ggttgacacttcacaaggtc	"-" strand primer from nucleotides 102 to 83 within *mos*1.
SH1	ttcaatgatgtccagtgcag	primer containing the terminal sequences of the *Drosophila* mini-*white* gene.

V5054	gggaaaaatgtgtagctagcgacggc	"+" strand oligonucleotide from 1169 to 1194 in *mos*1.
V6842	ctatggtggttcgacagtcaagg	"-" strand oligonucleotide from 123 to 101 in *mos*1.
V8523	**ggggtacc**ttaaatgccaatcgaagttcc	"+" strand oligonucleotide directed towards *mos*1 element in p*mos* from 5' flanking DNA.
V8524	**ggggtacc**ggctttttcttaaagctaccc	"-" strand oligonucleotide directed towards *mos*1 element in p*mos* from 3' flanking DNA.

All oligonucleotides are shown in 5' - 3' orientation. Non-homologous additional nucleotides are shown in bold; restriction enzyme recognition sites are shown underlined - *Bam*HI = 5' ggatcc 3'; *Eco*RI = 5' gaattc 3'; *Nde*I = 5' catatg 3'; *Kpn*I = 5' ggtacc 3'

2.3 Methods

2.3.1 Manipulation of Bacteria

2.3.1.1 Growth of bacterial cultures:

Liquid culture of *E.coli* was performed in L-broth, supplemented with antibiotics as required, by innoculating with a single colony using a sterile metal loop. Cultures of 5mls were grown in 2oz glass bottles, larger volumes were grown in conical flasks with a total capacity 10x that of the culture volume. Cultures were shaken at 37°C for 14-16 hours.

2.3.1.2 Storage of bacterial cultures:

Cultures were stored for a long period by mixing 1ml of a fresh overnight culture of *E.coli* with 1 ml of glycerol and storing in a sterile vial at -70°C. For recovery the culture was thawed and streaked out onto L-agar plates, with antibiotic if required.

Bacteria were stored as 5ml overnight cultures or on agar plates for short term storage at 4°C (4-6 weeks).

2.3.1.3 Preparation of competent *E.coli* cells:

One colony of *E.coli* was used to innoculate 5ml of L-broth and incubated at 37°C overnight with shaking. A subculture of 2.5ml of cells in 500ml of L-broth was incubated until an $O.D._{550}$ of 0.500-0.550 was reached. The cells were cooled on ice for 30 minutes, then pelleted for 15 minutes at 2500rpm in a pre-chilled Sorvall centrifuge. The cells were resuspended in 20ml of ice-cold competent cell buffer 1, and kept on ice for 15 minutes before centrifugation at 2500rpm for 9 minutes. The cell pellet was resuspended in 3.5ml of chilled competent cell buffer 2, and kept on ice for 15 minutes, before dividing into 200μl aliqouts and freezing in a dry ice/ ethanol bath. Frozen cells were stored at -70°C.

2.3.1.4 Transformation of plasmid DNA into competent cells

Five μl of a ligation reaction, or 1-100ng of plasmid DNA were added to 200μl of competent cells thawed slowly on ice. After 30 minutes on ice, cells were heat shocked at 42°C for 2 minutes. 200μl of L-broth was then added and incubated at 37°C for 1

hour. Cells were then spread onto L-agar with antibiotic as required and incubated overnight at 37°C.

For blue/white selection of cells, 100μl of 100mM IPTG and 20μl of 50 mg/ml X-gal were added to cultures immeadiately prior to spreading.

2.3.2 Nucleic acid preparation and manipulation techniques

2.3.2.1 Small scale preparation of plasmid DNA (miniprep) (Birnboim and Doly, 1979).

Single bacterial colonies harbouring a plasmid DNA of interest were innoculated into 5ml of L-broth containing antibiotic. After 14-16 hours shaking at 37°C, 1.5ml of the culture was pelleted in an eppendorf. The pellet was resuspended in 100μl of resuspension buffer. 200μl of lysis solution was then added before neutralisation with 150μl of 3M NaAcetate pH 5.5. Eppendorf tubes were left standing on ice for 5 minutes. Genomic DNA and cell debris were precipitated by centrifugation in a microfuge for 10 minutes at 12,000 rpm. The supernatant was then transferred to a new tube and an equal volume of phenol: chloroform, 1:1 added. After thorough mixing, then centrifugation for 2 minutes, the upper aqueoues phase was removed to a clean tube. An equal volume of chloroform was then added, mixed and centrifuged, before tranfer of the supernatant to a clean tube. Plasmid DNA in the supernatant was ethanol precipitated (2.3.2.4) and resuspended in 30-100μl of TE.

2.3.2.2 Small scale preparation of plasmid DNA (for use in ABI PRISM^{TM} Dye terminator cycle sequencing):

DNA was prepared as in section 2.3.2.1. After drying the DNA under vacuum, the pellet was dissolved in 32μl of dH_2O and precipitated by addition of 8μl of 4M NaCl and 40μl of 13% PEG8000. After mixing, the sample was incubated on ice for 20 minutes, before pelleting at 12,000rpm for 15 minutes at 4°C. The DNA was then rinsed with 70% ethanol and dried under vacuum before resuspension in 20μl of dH_2O.

2.3.2.3 Large scale preparation of plasmid DNA (midi/maxi prep):

Large scale preparation of plasmid DNA was carried out by using QIAGEN^{TM} plasmid midi/maxi kits (QIAGEN GmbH and QIAGEN Inc.) according to the

manufacturer's handbook instructions. This method is based on an alkaline lysis procedure, followed by binding to an anion-exchange resin under low salt and pH. Impurities such as RNA and protein are removed by a medium salt wash and plasmid DNA is eluted by high salt and concentrated by isopropanol precipitation.

2.3.2.4 Ethanol precipitation of DNA:

DNA was precipitated from solution by addition of 0.1 volumes of 3M Na Acetate pH 5 and 2-2.5 volumes of absolute ethanol at room temperature. The solution was mixed thoroughly and left at -20°C overnight or -70°C for 15 minutes. The DNA was then pelleted by centrifugation for 10 minutes at 12,000rpm, the supernatant discarded and the pellet washed with 70% ethanol. After removal of the supernatant, the pellet was dried under vacuum for 4 minutes and the DNA dissolved in TE.

2.3.2.5 Quantification of DNA concentration:

DNA concentrations were measured by spectrophotometry. Gene Quant (Pharmacia) and a lambda 15 UV/VIS spectrophotometer (Perkin Elmer) were used. O.D. readings were translated to DNA conentrations using the conversion factor of 50μg/ml of double stranded DNA giving an $O.D._{260}$ reading of 1.

2.3.2.6 Restriction enzyme digestion of DNA :

DNA was digested by restriction endonuclease in 30μl of appropriate 1x restriction buffer, in a 37°C water bath for 1-12 hours.

2.3.2.7 Agarose gel electrophoresis:

DNA was fractionated in 0.8-1.0% (w/v) Flowgen SEA KEM LE agarose made up in 1xTAE buffer containing 0.5mg/ml ethidium bromide. Prior to loading, DNA samples were mixed with 0.1 volumes of DNA loading buffer. Bacteriophage lcI857 DNA, cut with *Hind*III was used as size markers. Electrophoresis was performed horizontally across a potential difference of 1-10V/cm. DNA was visualised by UV trans-illumination and photographed.

2.3.2.8 Recovery of DNA from agarose gels (gene clean):

After agarose gel electrophoresis, the desired DNA fragment was cut from the gel and extracted using Geneclean^{TM} (Bio101). This involves binding of the DNA to a silica matrix formed by "glassmilk" and sodium iodide. Supplier's instructions were followed. Fragments less than 200bp were extracted using Mermaid^{TM} (Bio101) in which DNA binds to "glassfog". Again manufacturer's directions were followed.

2.3.2.9 Ligation of cohesive termini:

Plasmid and insert DNA to be ligated, were first digested with appropriate restriction endonucleases to yield complementary overhanging ends. Equimolar quantities of vector and foreign DNA were then mixed and 1µl 10x T4 ligase buffer, 1 Weiss unit T4 ligase and dH_2O to 10µl added. Reactions were incubated at 16°C overnight.

2.3.2.10 Creation of blunt ends from 3' recessed ends:

Recessed 3' ends were filled in to create blunt ends using the Klenow fragment of DNA polymerase I. DNA to be end filled was added to 2µl of 10x Klenow buffer and 1.5µl of 2mM dNTPs. dH2O to 19µl and 1 unit of Klenow enzyme were then added and incubated at 23°C for 1 hour. The DNA was then phenol extracted, ethanol precipitated (2.3.2.4) and resuspended in T.E. before use in a ligation reaction (2.3.2.11).

2.3.2.11 Ligation of blunt termini:

Blunt ended insert and plasmid DNA to be ligated were mixed in a 3:1 molar ratio at as high concentration as possible for a 10µl final reaction volume. 1µl of 10x T4 DNA ligase buffer, 1µl of 10µM hexammine cobalt chloride solution and 1 Weiss unit of T4 DNA ligase were then added and the samples incubated at 16°C for 20 hours.

2.3.2.12 Random-primed labelling of DNA:

Approximately 40ng of linear purified DNA was heat denatured by boiling and snap cooled on ice. ^{32}P-dCTP was employed to label the DNA, using a T7 oligolabelling kit (Pharmacia) following the manufacturer's instructions. Labelling reactions were incubated at 37°C for 2 hours.

G25 sephadex spun columns, made in 1ml syringes plugged with glass wool, were then used to separate radiolabelled DNA from unincorporated nucleotides.

2.3.2.13 Southern blotting (Southern, 1975):

Southern blotting was used to transfer DNA fragments separated on an agarose gel to nitrocellulose membrane. For efficient transfer of fragments, gels were initially immersed in 0.2M HCl for 20 minutes with gentle shaking. Gels were then soaked in denaturation buffer for 30 minutes followed by neutralisation buffer for 1hr again with gentle agitation. Soaking the gel in 0.2M HCl partially depurinates tha DNA. Addition of denaturation buffer, which contains a strong base, cleaves the DNA at the sites of depurination creating shorter fragments which transfer much more efficiently then larger ones.

Gels were placed on a glass support containing a blotting paper wick with both ends resting in a tray of 20xSSC. A nitrocellulose membrane (Hybond N, Amersham Intl.) soaked in 2xSSC was then placed carefully on top of the gel taking care to remove any air bubbles trapped. Three sheets of 2xSSC-soaked blotting paper were then placed over the membrane followed by some dry pieces of blotting paper, a stack of paper towels and a weight, usually a Sigma catalogue.

The transfer was allowed to proceed for 20 hours. The membrane was then rinsed in 2x SSC and UV cross-linked in a Stratalinker (Stratagene) to bind the DNA to the membrane.

2.3.2.14 Colony hybridisation (Benton and Davis, 1977):

Colony hybridisation was employed to identify bacteria containing recombinant plasmids with an insert of interest. Colonies of bacteria were transferred from L-agar plates to 9cm diameter Hybond-N nitrocellulose membrane filters (Amersham Intl.) by placing filters directly onto the plates for 2 minutes. Cells were lysed and DNA was denatured by soaking the discs in denaturation buffer for 2 minutes. Filters were then placed in neutralisation buffer for 2 minutes, washed in 2x SSC + 0.1% SDS for 2 minutes, and washed twice, for 2 minutes, in 2x SSC. The membranes were then UV cross linked as in 2.3.2.13.

2.3.2.15 Hybridisation of radiolabelled DNA to nucleic acids immobilised on nitrocellulose:

Nitrocellulose membranes, with bound nucleic acid, were placed in Techne hybridisation cylinders and 15ml of hybridisation solution, prewarmed to 65°C was added. The cylinders were put in Techne hybridisation ovens at 65°C and prehybridised for at least 60 minutes. ^{32}P-labelled probe was denatured by boiling for 2 minutes and added to the cylinders. Hybridisation was carried out overnight at 65°C. Excess probe and hybridisation solution were decanted and the filters washed 3 times in wash solution for 30 minutes at 65°C. The membranes were then air dryed before being wrapped in Saran wrap^{TM} and placed in autoradiographic or phosphorimager cassettes.

Probes were kept in hybridisation solution at room temperature and then reused, by boiling for 3 minutes and replacing prehybridisation solution. Nitrocellulose membranes were stripped of probe, when necceasry, by incubating for 30 minutes in 0.4M NaOH at 45°C, followed by 30 minutes in 0.1xSSc, 0.1% SDS, 0.2M Tris pH7.5 at 45°C. Filters were then reprobed.

2.3.2.16. Radioactive signal detection:

Radioactive ^{32}P and ^{35}S signals were detected by exposing membranes to a pre-flashed X-ray film (HA West) in an autoradiographic cassette. Cassettes were stored at -70°C, for ^{32}P membranes and room temperature for ^{35}S, for an appropriate length of time, before developing the film in an automatic X-ray film developer (X-OGRAPH COMPACT X-2).

Phosphorimage cassettes were used for quantative analysis of ^{32}P containing membranes. Scans from phosphorimager cassettes were analysed using Molecular Dynamics Image Quant^{TM} software.

2.3.2.17 Polymerase chain reaction (PCR):

For amplification of specific DNA fragments, primers were designed which shared at least 18 nucleotides of homology to the template. Some were designed with restriction sites at their 5' ends. All primers were provided by the Oswell DNA service.

All reactions were carried out using a Hybaid OmniGene programmable PCR machine. For a 50μl reaction volume, 100ng template DNA was mixed with 10nmoles dNTPs, 50pmoles each primer, 1.0U *TaqI* polymerase, 5μl 10x PCR buffer, supplied by the manufacturers, (200mM Tris.HCl pH8.4; 500mM KCl), and 5μl 15mM $MgCl_2$. Reactions were overlaid with 75ml mineral oil (Sigma).

A usual PCR run consisted of thirty cycles of denaturation at 94°C for 1min, annealing at 50°C for 1min and polymerisation at 72°C for 1min. Conditions were changed depending on the T_m of specific primers. Products from PCR amplifications were checked by agarose gel electophoresis (2.3.2.7). They were cloned, if necessary, by extraction from the gel, and, if primers had restriction endonuclease sites at their ends, by digesting and ligating as described in 2.3.2.6-2.3.2.11. Alternatively, use was made of the fact that *TaqI* polymerase adds a 3' terminal adenine to both strands. In this case pGEM-T Vector Systems (Promega) were used, following manufacturer's directions, to "TA clone" the products.

2.3.2.18 Manuel sequencing of double-stranded DNA (Sanger *et al*, 1977):

DNA sequencing was carried out using Sequenase^{TM} version 2.0 USB (United States Biochemicals) and α-^{35}S-dATP. Template and primer were annealed by boiling 9μl DNA solution (containing 3-5ng DNA) and 1μl of primer for 4 minutes before snap freezing on dry ice. Standard labelling and termination reactions, as described by USB were performed. Labelling reactions were carried out at room temperature for 4 minutes, and termination reactions were left for 3 minutes at 37°C.

Samples were boiled for 2 minutes before loading on a 6% polyacrylamide denaturing gel. Electrophoresis was carried out in 1xTBE buffer at 40W, to maintain a gel temperature of 50-55°C using a Sequi-Gen^{TM} sequencing gel (BIORAD). The gels were fixed in 10% methanol (v/v), 10% acetic acid (v/v) for 15 minutes, then dried under vacuum in gel dryer model 583 (BIORAD) for 1hr at 80°C and exposed to X-ray film overnight.

2.3.2.19 Automated sequencing of double-stranded DNA:

2.3pmol of primer and 500ng of template DNA (2.3.2.2) were added to 8.0μl of ABI PRISM^{TM} dye terminator cycle sequencing ready reaction mix (Perkin Elmer) and dH_2O added to 20μl. Reactions were overlaid with 40 μl of mineral oil. A cycle sequencing programme of 25 cycles of denaturation at 96°C for 10 seconds, annealing

at 50°C for 5 seconds and extension at 60°C for 4 minutes was then performed. Ethanol precipitation (2.3.2.4) to remove unincorporated terminater nucleotides was then carried out and samples were dried under vacuum and stored at -20°C. Samples were resuspended in loading buffer (5:1 deionized formamide:25mM EDTA containing 50mg/ml blue dextran) before loading on to an ABI PRISM^{TM} 377 DNA sequencer. DNA sequences were analysed using the gene jockey programme.

2.3.2.20 5' end labelling oligonucleotides with γ- ^{32}P-dATP:

Oligonucleotides were 5' ^{32}P end labelled for use in primer extension assays (2.3.2.21). 2µl of 10x polynucleotide kinase buffer, 1µl γ- ^{32}P-dATP, 50pmoles oligonucleotide, 10 units of polynucleotide kinase and dH_2O to 20µl were incubated at 37°C for 45 minutes, before heat denaturing the enzyme at 68°C for 10 minutes. G25 sephadex spun columns, made in 1ml syringes plugged with glass wool, were then used to separate radiolabelled DNA from unincorporated nucleotides.

2.3.2.21 Primer extension analysis of excision assays:

PCR based primer extension was performed to map *in vitro mos* cleavage sites at the nucleotide level. Reaction products from excision assays (2.3.5.1) were ethanol precipitated (2.3.2.4) and used as template. All reactions were carried out using a Hybaid OmniGene programmable PCR machine. For a 20µl reaction volume, template DNA was mixed with 10nmoles dNTPs, 0.5pmoles 5' ^{32}P end labelled primer (2.3.2.20), 1.0U *Taq*I polymerase, 2µl 10x PCR buffer, supplied by the manufacturers, (200mM Tris.HCl pH8.4; 500mM KCl), and 2µl 15mM $MgCl_2$. Reactions were overlaid with 40µl mineral oil (Sigma). Thirty cycles of denaturation at 94°C for 1min, annealing at 50°C for 1min and polymerisation at 72°C for 1min were then performed. Products were analysed by acrylamide gel electrophoresis, run against reference sequencing reactions.

2.3.2.22 Sucrose density gradient ultracentrifugation:

Approximately 500µg of *Drosophila* genomic DNA in TE to be separated was loaded on top of a 10ml, 10%-40% sucrose gradient prepared in 1M NaCl; 20mM Tris, pH 7.5; 5mM EDTA. The gradient was centrifuged for 21 hours at 110,000g. Fractions (250µl) were collected from the gradient starting at the bottom, and those containing *mariner* sequences identifed by Southern blot hybridisation.

2.3.3 Protein expression and detection

2.3.3.1 Expression of recombinant proteins in *E.coli*:

Plasmids containing the Mos1 ORF behind a suitable promoter were transformed into *E.coli* strains BL21(DE3) or BL21(DE3)pLysS. L-broth containing antibiotics as neccesary was innoculated with a single transformed colony and grown with shaking for 12-16 hours at 37°C. Ten-fold dilutions of culture into L-broth with antibiotic were made and grown at 37°C with shaking to an $O.D._{600}$ of 0.6-0.8. Cultures were then induced with IPTG to a final concentration of 1mM and incubated for a further hour. Cells were harvested and used immeadiately in purification procedures or stored at -70°C. Samples to be run on SDS polyacrylamide gels were resuspended in boiling mix.

2.3.3.2 SDS polyacrylamide gel electrophoresis (SDS PAGE):

Samples were loaded on a 10% acrylamide resolving gel, 3.75% acrylamide stacking gel and run in a BIORAD^{TM} mini protean II gel tank in 1xTGS for 40 minutes at 200V. To visualise protein, gels were immersed in Coomassie brilliant blue stain for 30 minutes at room temperature with shaking, then destained overnight. For more sensitive detection, gels were subjected to Western blot analysis.

2.3.3.3 Determination of protein concentration:

Protein concentration was estimated using the Bradford method. BSA standards in 1xPBS were used to calibrate a standard graph. 5µl of protein solutions to be quantified were made up to 1ml with 500µl of Coomassie protein assay reagent (Pierce) and 495µl of 1xPBS. The $O.D._{595}$ was measured and the protein concentration estimated from the standard graph.

2.3.3.4 Purification of GST-Mos1 protein:

GST-fusion proteins were purified under non-denaturing conditions by affinity chromatography on immobilised glutathione (Smith and Johnson, 1988).

E.coli strain BL21 (DE3) was transformed with plasmid pGEX-2T or pGEX-3X containing an in-frame fusion of the Mos1 ORF, and recombinant protein expression was induced as described in 2.3.3.1. Pellets from 1.5ml cultures were resuspended in

300μl of ice cold, 1xPBS and cells lysed using a probe sonicator. Insoluble material was pelleted by centrifugation for 5 minutes at 4°C. A 50% slurry of glutathione agarose beads (Sigma) was then mixed with the supernatant for 3 minutes at room temperature, before washing in 1xPBS. GST-fusion protein was eluted from the beads by addition of 50mM Tris pH8.0 containing 5mM reduced glutathione, aliquoted and stored at -70°C.

Factor Xa and thrombin were employed to cleave the GST moiety from GST-Mos1fusion proteins expressed from pGEX-3X and pGEX-2T respectively. For Factor Xa treatment, GST-Mos1was digested overnight at 4°C with 0.02units of FactorXa (Boehringer mannheim) in 50mM Tris pH7.5, 150mM NaCl, 1mM $CaCl_2$. Thrombin digestion was performed for twenty minutes at room temperature, using 1μg thrombin (Sigma) in 50mM Tris, pH8.0, 150mM NaCl, 0.5mM CaCl2, 0.1% 2-mercaptoethanol. Factor X_a and thrombin digestions were also performed in the presence of 0.005-0.1% (v/v) SDS.

2.3.3.5 Purification of 6-his-Mos1 protein:

Six histidine tagged fusion proteins were purified under denaturing conditions using Ni-NTA affinity resin (Qiagen).

Plasmid RSETC containing a translatable fusion of the Mos1 ORF to an N-terminal six histidine sequence was transformed into *E. coli* strain BL21(DE3)pLysS and expression of the recombinant protein induced as described in 2.3.3.1. Pellets from 250ml culture were lysed by sonication in buffer 1. Following centrifugation at 13,000rpm for 20 minutes at room temperature, the cell free extract was applied to a Nickel affinity resin pre-equilibriated in buffer 1. The column was washed in buffer 1, and a 6 to 0M urea buffer gradient applied to renature the protein. After washing the column, protein was eluted in imidazole buffer and dialysed against 3 changes of buffer 2. Protein fractions were aliquoted and stored at -70°C.

2.3.3.6 Mos1 transposase purification (This protocol was devised and performed by Dr. Angela Dawson:

Plasmid pBCP378-Mos1 containing the Mos1 ORF was transformed into *E. coli* strain BL21 (DE3) pLysS and expression of the recombinant protein induced as described in 2.3.3.1. One litre of induced culture was pelleted and resuspended in 5.0ml of resuspension buffer. 5mg of lysozyme was added to the resuspension and incubated at

room temperature for 5 minutes. 5ml of detergent buffer was then added, mixed and incubated at room temperature for a further 15 minutes. $MgCl_2$ to 10mM final concentration and 100 units of DNase I were further incubated in the lysate for 15 minutes before centrifugation at 12,000rpm for 20 minutes at 4°C to pellet insoluble matter. The pellet was resuspended in 10ml of wash buffer and pelleted at 12,000rpm, for 5 minutes at 4°C. This was repeated three times before washing in 2ml of urea buffer and resuspension in 500ml of guanidine buffer. After centrifugation for 5 minutes at 13,000rpm the supernatant was diluted into 50ml of diluent buffer, mixed by inversion and loaded onto a CM sepharose column, pre-equilibriated with column wash buffer. Renaturation of protein was then performed in a urea buffer gradient from 8 to 0M urea over 4 hours. Protein was eluted using a linear NaCl gradient from 100mM to 1M and 1ml fractions collected. Fractions were analysed by SDS PAGE, as described in 2.3.3.3 and fractions containing Mos1 concentrated as described in 2.3.3.5.

2.3.3.7 Preparation of Mos1 antibodies:

Pellets of *E.coli* overexpressing GST-Mos1 fusion protein were resuspended in boiling mix and subjected to SDS PAGE (2.3.3.2). Protein bands visualised by Coomassie brilliant blue staining. The band corresponding to overexpressed GST-fusion protein was excised from the gel using a sterile scalpel and rinsed several times in dH_2O. The gel slice was then added to 100µl 1xPBS, 1%SDS and rotated overnight at 4°C to elute the protein. Microsep™ microconcentrator vials of 30kDa molecular weight cut off were used to concentrate the protein to 2mg/ml, as determined by the Bradford assay.

Antibodies were raised against GST-Mos1 in New Zealand rabbits. After a pre-immune sample had been taken, 0.2ml of a solution containing 200µg of fusion protein, in a 1:1 ratio with Freund's Complete Adjuvant (Sigma), was injected at each of 4 sites. After 10-14 days a 5ml test bleed was taken. A second series of injections was performed 4 weeks after the first injections using a 1:1 ratio of protein to Freund's incomplete adjuvant (Sigma). A test bleed was taken after 10-14 days. This process was repeated for a third set of injections. Blood samples were allowed to clot, centrifuged and the supernatant removed and frozen in aliquots at -70°C. The supernatant was diluted 1/1,000 in block for use as the primary antibody in Western blots.

2.3.3.8 Western Blotting:

Proteins separated by SDS PAGE were transferred to PVDF membrane (Millipore) in western blot transfer buffer for 1 hour at 40V using a BIO-RAD Trans Blot^{TM} cell. Membranes were initially incubated in 1% block (Boehringer Mannheim) for 1 hour at room temperature with shaking. Protein of interest was detected by incubating the blot with primary antibody (1/1000 dilution in 0.5% block) for 1 hour. Blots were then washed in 1x TBS-T for 1 hour with frequent wash changes. Primary antibody was detected using a POD-conjugated goat anti-rabbit secondary antibody (diluted 1/1000 in 0.5% block) for 45 minutes, before washing repeatedly in 1xTBS-T for 30 minutes. The POD-conjugate reacts with the substrate luminol (Boehringer Mannheim Chemiluminescence Western blotting kit), in the presence of H_2O_2, to produce light which is detected on X-ray film.

2.3.3.9 Affinity purification of antibodies:

Antibodies were purified from crude serum using antigen immobilised on nitrocellulose membrane (Robinson *et al*, 1988). Preparative SDS PAGE of a crude lysate of E.coli cells induced for fusion protein synthesis was performed and protein Western blotted onto nitrocellulose membrane. The membrane was Ponceau S stained to visualise the protein pattern and the horizontal strip containing the fusion protein was excised and blocked in 7% Marvel in 1xPBS, with shaking for 1 hour. The membrane strip was then incubated with 10ml of blocking solution containing a 1:100 dilution of polyclonal serum for 2-3 hours at room temperature with agitation. The membrane was washed four times for 15 minutes each in 1x TBS-T. Bound antibodies were eluted by incubation in 2ml of 0.1M glycine-HCl pH2.5, 0.1% BSA (w/v) for 5 minutes followed by a second 2 ml of the same buffer for 10 minutes, both at room temperature. Each batch of eluant was immeadiately neutralised by the addition of 0.5ml 1M Tris pH7.5, the two batches were then pooled and dialysed against three changes of 1l each of 1xTBS at 4°C over 16-20 hours. The affinity purified antibody was stored at 4°C in the presence of 0.1% (w/v) NaN_3.

2.3.3.10 Extraction of proteins from *Drosophila*:

Ten *Drosophila* or *Drosophila* embryos were homogenised in 100µl of boiling mix, boiled for 10 minutes over a 100°C water bath and centrifuged at 12,000rpm for 10

minutes. The supernatant was then removed and loaded directly onto an SDS polyacrylamide gel or stored at -20°C.

2.3.4 *Drosophila* methods

2.3.4.1 Preparation of genomic DNA:

No. of flies	1	10	50
Solution A (ml)	100	200	500
8M KAc	14	28	70
Isopropanol (ml)	50	100	250
T.E. (ml)	10	20	100

Table 1. Volumes of solutions required for DNA extraction from *Drosophila.*

1-50 flies were placed in an Eppendorf tube at -70°C for 5 minutes, then homogenised in solution A (0.1M Tris.HCl, pH9.0; 0.1M EDTA, pH8.0; 1% SDS; see Table 1. for amount added per fly). After incubation at 70°C for 30 minutes, 8M potassium acetate was added and tubes 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. Isopropanol was added, and DNA left to precipitate for 10 minutes at -70°C. After a 5 minute centrifugation at 12,000rpm, the pellet was washed in 70% ethanol, dried under vacuum and resuspended in TE.

2.3.4.2 Collection of eggs for microinjection:

Egg collection chambers with a sliding food drawer were used for collecting eggs. The food drawer was filled with egg injection media and placed in the chamber along with several hundred 4-5 day old flies. The flies were allowed to lay eggs in the chamber at 25°C before dechorionation and injection as described in 2.3.4.3 and 2.3.5.2.

2.3.4.3 Establishment of transgenic lines by *P* element transformation (Rubin and Spradling, 1982):

DNA for transformation was prepared as described in 2.3.2.3, but DNA was resuspended in deionised water, not TE as this may adversely affect embryos. Injection buffer (5mM KCl; 100µM $NaPO_4$, pH6.8) containing 400ng/µl of plasmid (with the gene of interest) and 100ng/µl helper plasmid (pπ25.7wc) was back-loaded into the injection needle. Embryos from strain y^1 v^{36f}; ry^{506} (see chapter 7) were collected as described in 2.3.4.2 over 30 minute intervals, dechorionated in 50% bleach, and washed onto a gridded filter. After lining up embryos in the same orientation alongside a grid line, they were transferred to a coverslip with glue solution (from dissolving sticky tape glue in heptane) along one edge, and dessicated in a petri dish containing silica gel for approximately 6 minutes.

Embryos were covered with halocarbon oil, and DNA was injected into the posterior pole of preblastoderm embryos, using an eppendorf transjector 5246. Older embryos were killed by running the needle gently through them. After incubation at 18°C for 36-60hr, larvae were collected as they emerged and were left to develop at 25°C in vials of French fly food. Adults were backcrossed to y^1 v^{36f}; ry^{506} virgins, and their progeny examined for coloured-eyed flies. Transformants were crossed to y^1 v^{36f}; ry^{506} flies to produce males and females with the same insertion. Homozygous lines were then establised by sibling matings.

2.3.5 Activity assays using Mos1 transposase

2.3.5.1 *In vitro* excision assay (modified from Vos *et al*, 1996):

Approximately 200ng of supercoiled *pmos* plasmid, containing the *mos1* element from *D.simulans,* was incubated in 19µl of excision assay buffer containing 1µg each of BSA and poly d(I-C)-d(I-C) and 5mM MnAc or $MgCl_2$. 1µl of 250ng/µl purified Mos1 transposase was then added and reactions incubated at 37°C for 5, 15 or 30 minutes. Reactions were immeadiately placed at -20°C, or analysed by electrophoresis on 1% agarose gels as described in 2.3.2.7, followed by Southern blotting (2.3.2.13), hybridisation to ^{32}P-labelled Mos1 ORF sequence (2.3.2.15) and radioactive signal detection (2.3.2.16).

2.3.5.2 Germline transformation of *D. melanogaster* by purified Mos1 transposase (modified from Kaufman and Rio, 1991):

DNA to be transformed was prepared as in 2.3.2.3 and resuspended in dH_2O. Transposase used was prepared as described in 2.3.3.6 by Dr. A. Dawson. Transposase was diluted in excision assay buffer supplemented with 5mM $MnCl_2$, 25µg/ml BSA, 0.5µg/ml antipain, aprotinin and chymostatin (Sigma), 0.2mM sodium metabisulphite and 70 µg/ml PMSF. Equal volumes of DNA and transposase fractions were mixed (see section 6.2.3) and injected into w^{1118} embryos following the procedure described in 2.3.4.3. Adults were mated with w^{1118} fies and their progeny examined for coloured eyes.

Chapter 3

Overexpression and Purification of the Mos1 ORF

3.1 Introduction

A characteristic of autonomous transposable elements is the presence of one or more genes which encode proteins (transposases) essential for transposition.

Recently, the gene products of a number of eukaryotic transposable elements have been purified and used to reconstitute transposition reactions *in vitro*. The *Tc*1 element from *C. elegans* contains a single ORF comprising two exons, that are joined together by RNA splicing to encode a single protein, Tc1A (Vos *et al*, 1993). Recombinant Tc1A purified from *E. coli* has been demonstrated to be capable of supporting *Tc*1 transposition *in vitro* (Vos *et al*, 1996). Similarly, the gene product of the *mariner* element *Himar*1, from the hornfly, has been purified and shown to be sufficient for transposition of *Himar*1 *in vitro* (Lampe *et al*, 1996).

The active *mariner* element *mos*1, isolated from *D. mauritiana* contains a single ORF of 345 amino acids which is predicted to encode a protein of 40.8kD (Jacobson *et al*, 1986). The product of this ORF is required for transposition *in vivo* (Medhora *et al*, 1991). In this chapter, attempts to express and purify a functional Mos1 protein are described. The Mos1 ORF was initially subcloned into a number of different prokaryotic expression vectors for inducible expression in *E. coli*. Both Mos1-translational fusions and directly expressed Mos1 protein were then purified as determined by polyacrylamide gel electrophoresis. Recombinant Mos1 protein was then selected for use in a preliminary activity assay as described in chapter four, in which the ability of the protein to stimulate excision of *mos*1 DNA sequences from plasmids was examined.

3.2 Results

3.2.1 Production of a Mos1 translational fusion protein

3.2.1.1 Sub-cloning the Mos1 ORF into prokaryotic expression vectors for production of a translational fusion protein

Prokaryotic expression vectors pGEX-2T, pGEX-3X and pRSETC were employed to express the Mos1 ORF as a translational fusion in *E. coli*.

The pGEX plasmids contain a *tac* promoter and are designed for inducible, high level intracellular expression of genes as fusions with *Schistosoma japonicom* glutathione-S-transferase (GST) protein (Smith and Johnson, 1988). Fusion proteins are purified from bacterial lysates under non-denaturing conditions by affinity chromatography

using glutathione agarose beads. GST-fusion proteins expressed from vector pGEX-3X contain a blood coagulation factor X_a protease cleavage site between the GST carrier and the translational fusion allowing removal of the 26kD GST protein subsequent to purification. Vector pGEX-2T contains a thrombin cleavage site at this position.

Vector pRSETC contains a T7 promoter and is designed such that the polypeptide of interest is fused downstream of a sequence of six histidine residues. The six histidine tag allows efficient, one step purification of fusion protein on a Ni^{2+}-NTA affinity resin under both native and denaturing conditions.

Primers 337 and 338 were designed for amplification and sub-cloning of the Mos1 ORF. Primer 337 is homologous to the first 16 bases of the Mos1 ORF, but contains C rather than A in the *mos*1 translation initiation codon ATG, such that the initial methionine is omitted. In the absence of the *mos*1 initiation codon, translation starts at the first codon of GST. Primer 337 also contains a *Bam*HI site. Primer 338 contains the last 18 bases of the Mos1 ORF including the translation termination codon and an *Eco*RI site.

Primers 337 and 338 were employed in PCR reactions as described in 2.3.2.17 using plasmid p*mos* as template. The resulting PCR product of 1kb was gel purified and digested with *Bam*HI/*Eco*RI. Plasmids pGEX-3X, pGEX-2T and pRSETC were also *Bam*HI/*Eco*RI digested. The products of digestion were gel purified and quantified before ligating each plasmid backbone to the ORF insert in the presence of T4 DNA ligase for 16 hours at 16°C. Ligation of *Bam*HI/*Eco*RI restricted ORF sequence into the *Bam*HI/*Eco*RI digested pGEX and pRSETC plasmids produces in-frame insertions. Competent *E.coli* NM522 cells were then transformed and allowed to grow overnight on L-Amp plates. Colonies containing the ORF sequence were selected by picking colonies and performing diagnostic restriction digests on minipreparations of DNA.

3.2.1.2 Expression of GST-Mos1 fusion protein

The positive pGEX-3X-Mos1 and pGEX-2T-Mos1 constructs were transformed into *E.coli* strain BL21(DE3). L-broth containing ampicillin was innoculated with a single transformed colony and grown with shaking at 200rpm for 12-16 hours at 37°C. The following day ten-fold dilutions of culture into L-broth with ampicillin were made and grown at 37°C with shaking to an $O.D._{600}$ of 0.6-0.8. Cultures were then induced

with IPTG at a final concentration of 1mM and incubated for a further hour. Protein expression was analysed by SDS PAGE (2.3.3.2).

3.2.1.3 Purification of GST-Mos1 fusion protein

GST-fusion proteins were purified under non-denaturing conditions by affinity chromatography on immobilised glutathione (Smith and Johnson, 1988).

Pellets from 1.5ml cultures were resuspended in 300μl of ice cold 1xPBS and cells lysed for 30 seconds at 6 microns using a probe sonicator. Insoluble material was pelleted by centrifugation. A 50% slurry of glutathione agarose beads was then mixed with the supernatant at room temperature, before washing in 1xPBS. GST-Mos1 fusion protein was eluted from the beads by addition of 50mM Tris pH8.0 containing 5mM reduced glutathione

Purified GST-Mos1 fusion protein was subjected to SDS-PAGE and Coomassie stained. Figure 3.1, lane 1 shows the 67kD GST-Mos1 fusion protein after elution from glutathione agarose beads.

Attempts were made to cleave the GST tag from the fusion proteins. GST-Mos1 fusion proteins expressed from pGEX-3X were subjected to factor X_a protease treatment (2.3.3.4), and from pGEX-2T to thrombin digestion as described in (2.3.3.5). Cleavage of the GST moiety was not observed in either instance as analysed by SDS PAGE. On addition of a range of SDS concentrations, from 0.005-0.1%, to the Factor X_a cleavage reaction, cleavage products were obtained of the incorrect size for precise removal of GST and are presumably due to cleavage at secondary sites.

3.2.1.4 Preparation of GST-Mos1 antibodies

GST-Mos1 fusion protein was prepared for use as an antigen to raise antibodies against Mos1 transposase. Pellets of *E.coli* containing overexpressed GST-Mos1 fusion protein were resuspended in boiling mix and subjected to SDS PAGE (2.3.3.2). GST-fusion protein was visualised by Coomassie brilliant blue staining and bands corresponding to overexpressed GST-fusion protein excised from the gel. Protein was eluted from the gel slice as described in 2.3.3.7, before being concentrated and quantified.

Antibodies were raised against GST-Mos1 in New Zealand rabbits. A solution containing 200mg of fusion protein, in a 1:1 ratio with Freund's Complete Adjuvant,

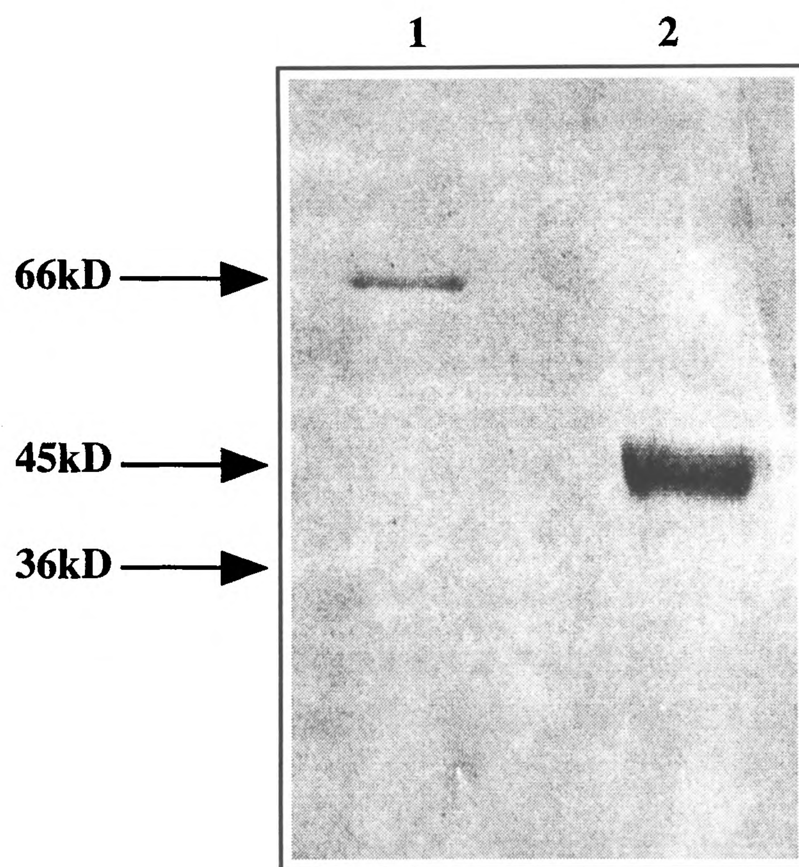


Figure 3.1. Coomassie stained SDS PAGE of purified Mos1-fusion proteins.

Samples fractionated on a 10% polyacrylamide gel at 200V as described in text. Molecular weight markers are SDS-7 (Sigma).

Lane 1......... GST-Mos1 fusion protein following elution from immobilised glutathione.
Lane 2......... 6-his-Mos1 fusion protein after elution from Ni-NTA resin in imidazole buffer.

was injected at each of four sites. A second and third series of injections were performed four and eight weeks later using a 1:1 ratio of protein to Freund's Incomplete Adjuvant. Blood samples were allowed to clot, centrifuged and the supernatant removed and frozen in aliquots at -70°C till required.

3.2.1.5 Expression of 6-his-Mos1 fusion protein

Plasmid pRSETC-Mos1 was transformed into *E. coli* strain BL21(DE3)pLysS and recombinant protein expression induced as described in 2.3.3.1.

Cell pellets containing expressed fusion protein were resuspended and sonicated before centrifugation to pellet insoluble material. Soluble protein remains in the supernatant. Analysis of the two fractions by SDS PAGE revealed the 6-his-Mos1 fusion protein to be present in the insoluble fraction.

3.2.1.6 Purification of 6-his-Mos1 fusion protein

Six histidine tagged fusion protein was purified from inclusion bodies under denaturing conditions using Ni-NTA affinity resin (Qiagen). Unlike the protocol for GST-fusion protein purification, 6-histidine tagged ORFs can be purified under denaturing conditions as the high affinity interaction between the Ni-NTA resin and the 6-histidines is retained in the presence of denaturants such as urea (Bush *et al*, 1991).

Pellets from 250ml cultures of expressed 6-his-Mos1 fusion protein were lysed by sonication in a buffer 1. Following centrifugation, the cell-free extract containing denatured, soluble protein was applied to a nickel affinity resin as described in 2.3.3.5. The column was washed and a urea buffer gradient applied to renature the protein. Fusion protein was then eluted in imidazole buffer and dialysed against three changes of buffer 2 to remove the imidazole. Protein fractions were aliquoted and stored at -70°C. Figure 3.1, lane 2 shows the soluble, purified 6-his-Mos1 fusion after elution from the nickel resin.

3.2.2 Direct expression of the Mos1 ORF

3.2.2.1 Sub-cloning the Mos1 ORF into prokaryotic expression vectors for direct translation

Expression vectors pET13a and pBCP378 are prokaryotic expression vectors designed to allow the direct translation of an ORF from either a T7 or a *trc* promoter in *E.coli.*

Both pET13a and pBCP378 contain an *Nde*I site (CATATG), incorporating the ATG start codon, at an appropriate distance from the ribosome binding site. This permits the cloning of a gene at its translation start point without altering the amino acid sequence of the synthesised protein.

Primer N6799, homologous to the first 17 bases of the Mos1 ORF and incorporating an *Nde*I site at the initiation codon ATG, was employed in a PCR reaction with primer 338. PCR product was analysed on an agarose gel before 'TA' cloning into vector pGEM-T as described in 2.3.2.9. to produce pGEM-T-Mos1. For proficient restriction of DNA *Nde*I requires the presence of at least seven bases either side of its recognition site (NEB catalogue). By direct cloning of the *NdeI*/*Eco*RI PCR product into a 'TA' cloning vector, subsequent cloning steps which require restriction at the *Nde*I site can be performed efficiently.

For subcloning the ORF into pET13a, the 1kb *NdeI*/*Eco*RI fragment was excised from pGEM-T-Mos1, and ligated into *NdeI*/*Eco*RI digested pET13a. Positive clones were selected by diagnostic restriction digestion.

For subcloning into pBCP378, pGEM-T-Mos1 was first linearised with *Eco*RI and end-filled with Klenow enzyme to create a blunt end at the 3' end of the Mos1 ORF. The Mos1 ORF was then digested from pGEM-T by *Nde*I and purified before ligation to *NdeI*/*Sma*I cut pBCP378. Positive clones were identified by colony hybridisation (2.3.2.14). This cloning step was performed by Dr. Angela Dawson.

3.2.2.2 Expression of Mos1 protein from pET13a

The pET13a-Mos1 construct was transformed into *E.coli* strain BL21(DE3)pLysS and protein expression induced as described in 2.3.3.1. Expression was allowed to continue for six hours with samples being taken at one hour intervals. Low levels of expression of Mos1 could be detected, as determined on Coomassie stained polyacrylamide gels. The effect of increasing the concentration of IPTG was studied. A six hour time course was repeated comparing Mos1 expression in cultures induced with, 1, 2.5 and 5mM IPTG. No noticeable alteration in expression was observed (data not shown).

3.2.2.3 Expression of Mos1 protein from pBCP378

Plasmid pBCP378 containing the Mos1 ORF was transformed into *E. coli* strain BL21 (DE3) and expression of the recombinant protein induced. High levels of induction

were observed. The majority of Mos1 protein is present in the insoluble fraction (see figures 3.2 and 3.3).

3.2.2.4 Purification of Mos1 protein (This protocol was devised and performed by Dr. Angela Dawson).

Purification of Mos1 protein was performed as described in detail in 2.3.3.6. One litre of induced culture was pelleted and resuspended in buffer before addition of lysozyme and detergent to lyse cells. The cell lysate was then treated with DNaseI to decrease the viscosity of the lysate. Insoluble matter containing Mos1 inclusion bodies was then pelleted and washed several times before resuspension in a buffer containing 6M guanidine. Protein solubilised in guanidine was diluted into a buffer containing 8M urea, to give a final guanidine concentration of 50mM. Soluble protein was then loaded onto a CM sepharose column before renaturation in a urea buffer gradient. Soluble, renatured Mos1 protein was eluted on application of a linear NaCl gradient to the column. Fractions containing Mos1 transposase were concentrated as described in 2.3.3.5. Mos1 transposase eluted at between 200-300mM NaCl.

Figure 3.2, depicts the different stages of protein purification, as visualised on a polyacrylamide gel by Coomassie staining. A comparison of lanes 1 and 2 demonstrates that the majority of overexpressed Mos1 is present in the insoluble fraction. Lanes 3-5 contain samples of supernatant following washing of the inclusion bodies in detergent and 6M urea, and lane 6, the load immediately prior to the chromatographic step. Lane 7 demonstrates that the protein purified to near homogeneity, migrates at 41kDa, which is the predicted molecular size for Mos1. This band was not found in a mock purification. A Western blot of the purification procedure using anti-Mos1 antibodies raised against GST-Mos1 was then performed to ensure the identity of the protein.

3.2.2.5 Affinity purification of antibodies

Antibodies were purified from crude serum using antigen immobilised on nitrocellulose membrane (Robinson *et al*, 1988). SDS PAGE of 6-his-Mos1 protein was performed and Western blotted onto nitrocellulose membrane. The membrane was stained with Ponceau S to visualise the protein and the horizontal strip containing the fusion protein cut out. This membrane strip was incubated with a dilution of polyclonal serum before washing and elution of the antibodies (2.3.3.9). The affinity purified antibody was stored at 4°C in the presence of 0.1% (w/v) NaN_3.

Figure 3.2. Coomassie stained polyacrylamide gel, demonstrating different stages of purification of the Mos1 ORF

Samples were fractionated alongside SDS-7 molecular weight markers (Sigma), on a 10% polyacrylamide gel as described in 2.3.3.2.

Lane 1........ Whole cell extract from BL21(DE3)pLysS cells overexpressing the Mos1 ORF

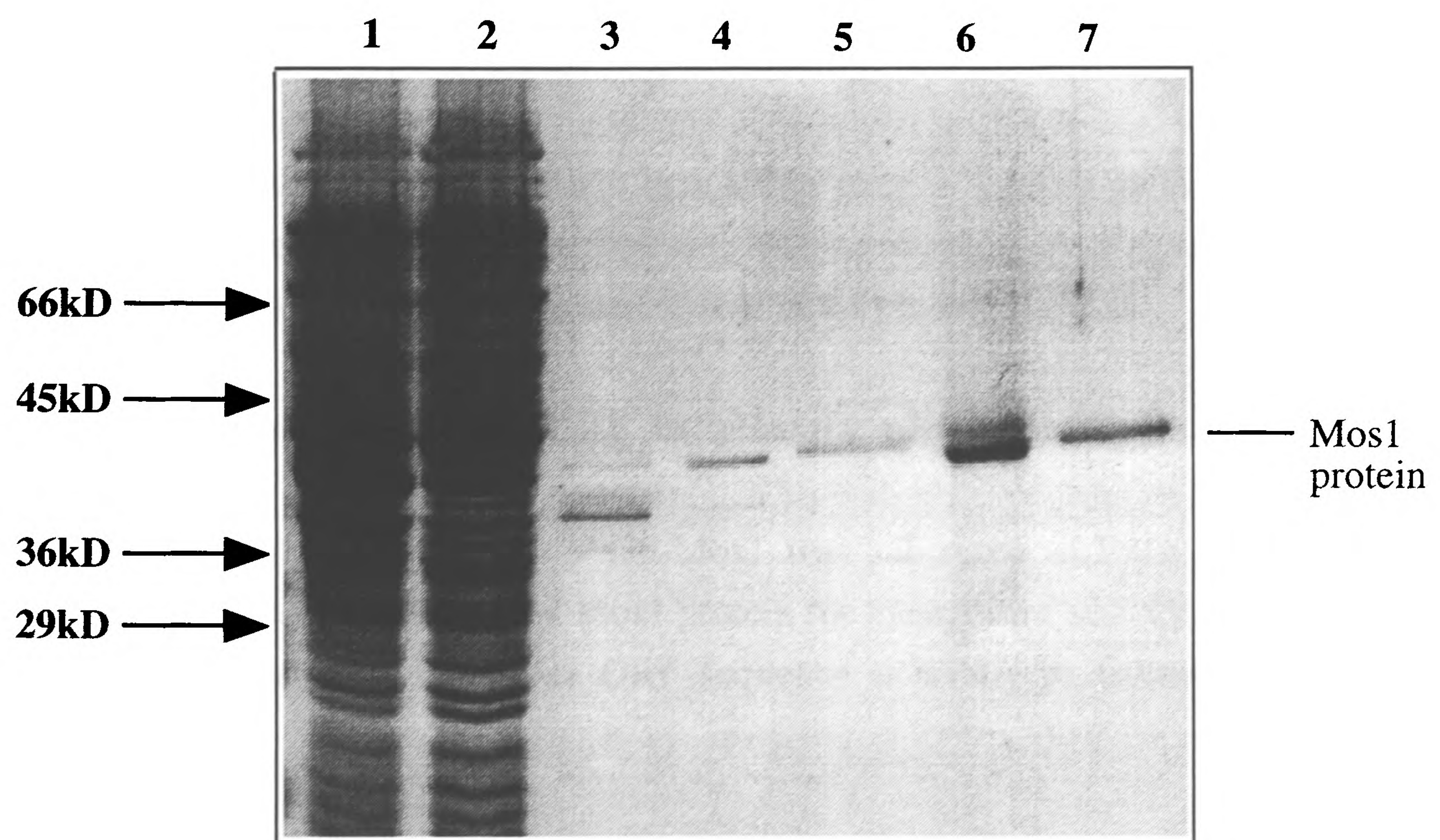
Lane 2........ Soluble cell extract from BL21(DE3)pLysS cells overexpressing Mos1

Lane 3-4...... Supernatant removed during first and second NP40 washes

Lane 5........ Supernatant removed during 6M urea wash

Lane 6........ Column load

Lane 7........ Column eluate at 200-300mM NaCl, comprising purified, soluble Mos1 protein.


1
2
3
4
5
6
7
66kD
45kD
36kD
29kD
Mos1
protein

3.2.2.6 Confirmation of the identity of the purified protein using anti-Mos1 antibodies

Protein fractions from different stages of mos1 purification were separated by SDS PAGE and transferred to PVDF membrane (2.3.3.8). Mos1 protein was detected by incubating the blot with affinity purified rabbit anti-Mos1 antibody, followed by POD-conjugated goat anti-rabbit secondary antibody. Signal was detected on X-ray film on addition of luminol. From figure 3.3, lane 1, it is apparent that the majority of the protein is present in the insoluble fraction. However, a strong cross-hybridising band is also observed in the lane containing soluble protein. This is likely to be an overestimation of the relative amount of soluble Mos1, since the centrifugation conditions used to pellet the inclusion bodies are probably not sufficiently stringent to pellet all the insoluble protein. Lanes 3-5, demonstrate that no Mos1 is lost due to solubilisation in detergent (NP40) and 6M urea washes, and lane 6 shows the guanadine solubilised Mos1 diluted into urea buffer prior to loading on the column. Lane 7 demonstrates that affinity purified anti-Mos1 antibodies cross react with the 41kDa purified protein. This confirms the identity of the protein as Mos1.

3.2.2.7 Sequencing the Mos1 ORF in plasmid pBCP378-Mos1

The ORF encoding Mos1 in pBCP378-Mos1 was subjected to DNA sequence analysis, prior to testing the purified Mos1 protein for transposase activity. This was performed to ensure that the Mos1 ORF sequence is wild-type following PCR amplification.

Primers 337, 338, P3855, P3856, M4269 and M4270 were employed in reactions such that the complete Mos1 ORF sequence could be read. Manual sequencing was performed as described in 2.3.2.18 using a Sequenase version 2.0 USB kit and α-^{35}S-dATP. A single base alteration was observed in the third position of a proline codon. As proline codons are degenerate at the position of the third base, the ORF sequence at the amino acid level of Mos1 remains wild-type.

Figure 3.3. Reaction of anti-Mos1 antibodies with extracts from the different stages of purification of the Mos1 ORF

Samples were fractionated on a 10% polyacrylamide gel as in figure 3.2, prior to transfer to PVDF membrane. Molecular weight markers are prestained protein markers, broad range (NEB). Anti-Mos1 polyclonal antibodies were used at 1/1000 dilution.

Lane 1........ Whole cell extract from BL21(DE3)pLysS cells overexpressing the Mos1 ORF

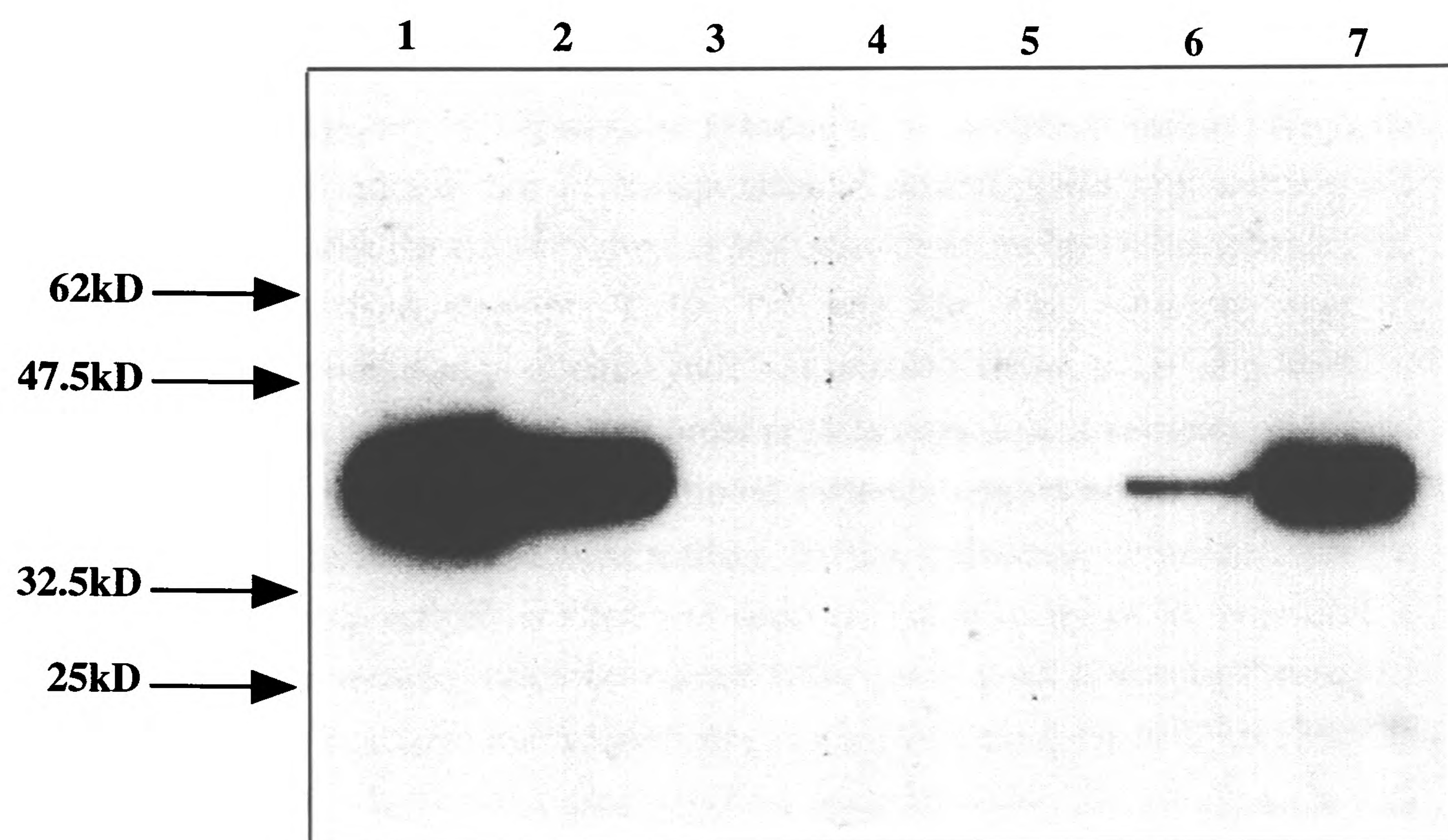
Lane 2........ Soluble cell extract

Lane 3-4...... Supernatant removed during first and second NP40 washes

Lane 5........ Supernatant removed during 6M urea wash

Lane 6........ Column load

Lane 7........ Column eluate at 200-300mM NaCl, comprising purified, soluble Mos1.


1
2
3
4
5
6
7
62kD
47.5kD
32.5kD
25kD

3.3 Discussion

The Mos1 protein has been overexpressed and purified from *E. coli* as both a translational fusion and as a directly expressed protein.

Expression of this protein as a translational fusion provides an efficient way to purify recombinant proteins in a single step. Purification can be performed under both native (GST- and 6-his) and denaturing (6-his) conditions. However, the fusion tags which allow rapid purification provide a bulky addition to the N-terminal of the Mos1 protein. GST in particular is large in size, contributing to more than one third of the molecular weight of the GST-Mos1 fusion protein.

During transposition, transposases recognise and bind to specific DNA sites at or near the ends of their element and co-ordinate the chemical reactions of strand cleavage and transfer. The arrangement of higher order synaptic nucleoprotein-complexes has been shown to be a pre-requisite for the transposition of several prokaryotic transposons (see 1.3.3.7), in particular the formation of such structures has been demonstrated for the well characterised elements Tn10, Tn7 and Mu. The chemical steps of transpositional recombination of eukaryotic transposable elements are also likely to proceed within the context of a higher order protein-nucleic acid complex. It is thus conceivable that the presence of additional residues tagged on to the end of a transposase may be detrimental to its activity. Indeed, it has been postulated that the absence of a nicking activity predicted to occur on the 3' strand at the ends of *Tc*1 DNA following incubation with 6-his-tagged Tc1A transposase, is that the C-terminal hisitidine tag may interfere with the assembly of a higher order protein-DNA complex (Vos and Plasterk, 1994). Attempts were thus made to cleave the purification tags from the Mos1 fusions.

On initial incubation of GST-Mos1 fusion protein with Factor X_a, or thrombin, in appropriate cleavage buffer, cleavage was not observed. It is possible that cleavage was not detected because the protease cleavage site is inaccessible to the proteases, Factor X_a in particular is bulky in size which may contribute to the problem. Low concentrations of SDS (0.005-0.1%) have previously been used to stimulate Factor X_a cleavage (Ellinger *et al*, 1991). This is thought to work by 'relaxing' the fusion protein so that the protease recognition site is no longer obscured. However, cleavage products obtained under these conditions were not of the predicted size. It appears that in the presence of SDS, cleavage by factor X_a occurs at secondary sites.

Substantial levels of Mos1 overexpression were observed on induction of recombinant protein expression from plasmid pCBP378-Mos1 in *E. coli* (figure 3.2, lane 1). Cells harvested from this induction were used as starting material for Mos1 purification. The transposase was purified to near homogeneity, as determined by SDS PAGE, through washing of inclusion bodies and a single chromatographic step. The identity of the protein was confirmed in a Western blot experiment using polyclonal anti-Mos1 antisera. Untagged, purified Mos1 protein was then selected for use in a preliminary activity assay.

Chapter 4

In vitro activity of purified Mos1 protein

4.1 Introduction

Following induction of high levels of *Tc*3 transposition in *C. elegans*, excised transposon sequences of 2.3kb can be detected. These molecules correspond to linear copies of transposon (van Luenen *et al*, 1993). Determination of the exact sequences at the ends of these elements has resulted in the suggestion that they are the active intermediates in *Tc*3 transposition (van Luenen *et al*, 1994). More recently, the *Tc*1 transposase, Tc1A from *C. elegans* has been over-expressed and purified from *E. coli* and used to reconstitute *Tc*1 transposition *in vitro*. During *in vitro* transposition of *Tc*1, both with purified transposase and with nematode extract, the appearance of excised *Tc*1 elements has been demonstrated (Vos *et al*, 1996). Complete excision of *Tc*1 and *Tc*3 suggests that transposition occurs by a 'cut and paste' process, similar to that of bacterial transposons such as Tn10. This is consistent with genetic data on double strand-break repair of donor DNA after *Tc*1 excision (Plasterk, 1991).

Cut and paste transposition is non-replicative. Cleavage takes place at both the 5' and 3' ends of the transposon, resulting in an excised copy of the element. The excised transposon then becomes joined to the target site by strand transfer between the 3' OH transposon end and the 5' phosphate target DNA, resulting in a simple insertion (see 1.3.3.3). *Mariner* elements have been predicted to transpose by a 'cut and paste' mechanism (van Luenen *et al*, 1994). This has recently been demonstrated for the *mariner* element, *Himar*1 from the Hornfly. *Himar*1 transposase cleaves *Himar*1 termini at both the 5' and 3' ends, consistent with a 'cut and paste' mode of transposition (Lampe *et al*, 1996). From this information it is conceivable that intermediates may occur during transposition of the *mariner* element *mos*1, and that these elements, like those of *Tc*1 and *Tc*3 may be detectable in a physical assay.

In this chapter, the ability of recombinant Mos1 protein, expressed and purified from *E. coli*, to stimulate the excision of *mos*1 elements from plasmids is described.

4.2 Results

4.2.1 *In vitro* excision of *mos*1 DNA sequences from plasmid p*mos*

Recombinant Mos1 protein has been purified from *E. coli* to near homogeneity as described in 3.2.2.4 and figures 3.2, 3.3. The ability of purified Mos1 protein to catalyse the excision of *mos*1 DNA from flanking sequences was examined. Plasmid p*mos* was used as a substrate for this reaction. p*mos* consists of a 5kb *Bam*HI/*Hind*III fragment of genomic *D. mauritiana* DNA, ligated into the polylinker of pBluescribe

(Medhora *et al*, 1991). The 5kb clone contains a full length (1.3kb), functional, *mos*1 element, situated 0.2kb from the *Hind*III site and 3.5kb from the *Bam*HI site. A large-scale preparation of p*mos* plasmid was performed using a Qiagen™ midi kit as described in 2.3.2.3 and quantified before storing at -20°C. Stock solutions, of HEPES pH 7.9, NaCl and glycerol, for use in the assay, were made using milliQ dH_2O and autoclaved. DTT and metal ion solutions were made up freshly each time as required.

One microlitre of supercoiled p*mos* plasmid (200ng) was initially incubated in a 19µl volume containing 25mM HEPES pH7.9; 100mM NaCl; 10% (v/v) glycerol and 2mM DTT supplemented with 1µg each of BSA and poly d(I-C)-d(I-C), and Manganese Acetate to a final concentration of 5mM. One microlitre of a 250ng/µl preparation of Mos1 protein was then added, followed by incubation for 30 minutes at 37°C. Poly d(I-C)-d(I-C) was included in the reaction to 'mop up' non-specific DNA binding proteins that may have co-purified with Mos1. BSA was added to increase the stablility of the purified Mos1 protein at low concentration.

The following controls were also used. Firstly, full assay conditions were observed with the single omission of Mos1 protein. In the second, p*mos* DNA was incubated in T.E. rather than excision assay buffer, transposase was also omitted from this reaction. The controls allow any activity observed to be attributed to either addition of Mos1 protein or to the presence of a contaminating nuclease activity in any of the constituents of the assay buffer. The excision assay and two control reactions were then repeated with the substitution of pBluescript (pBS), for p*mos*. pBS does not contain *mos*1 sequences.

Following incubation for 30 minutes at 37°C, all reactions were fractionated on a 1% agarose gel. On visualising the DNA under ultraviolet light, relaxation of the high mobility supercoiled p*mos* and pBS substrates could be detected. This was observed only in the reactions in which transposase had been incubated with the DNA (figure 4.1a lanes 4 and 8). Under UV light, high mobility bands corresponding to excised copies of *mos*1 could not be observed. However, a band of approximately 8kb, could be detected in lane 8. This is the mobility of linear p*mos* and it is possible that the band corresponds to molecules in which cleavage at a single transposon end has occurred.

To increase the sensitivity of the assay the reaction products were subjected to Southern blot analysis. DNA fragments from the agarose gel were transferred to nitrocellulose membrane as described in 2.3.2.13. ^{32}P labelled probe corresponding to

Figure 4.1. ***In vitro*** **excision of the** ***mos*1 transposable element from plasmid p*mos*.**

All lanes contain 200ng DNA. λ DNA (NEB) was restricted with *Hind*III to provide molecular weight standards. See text for assay details.

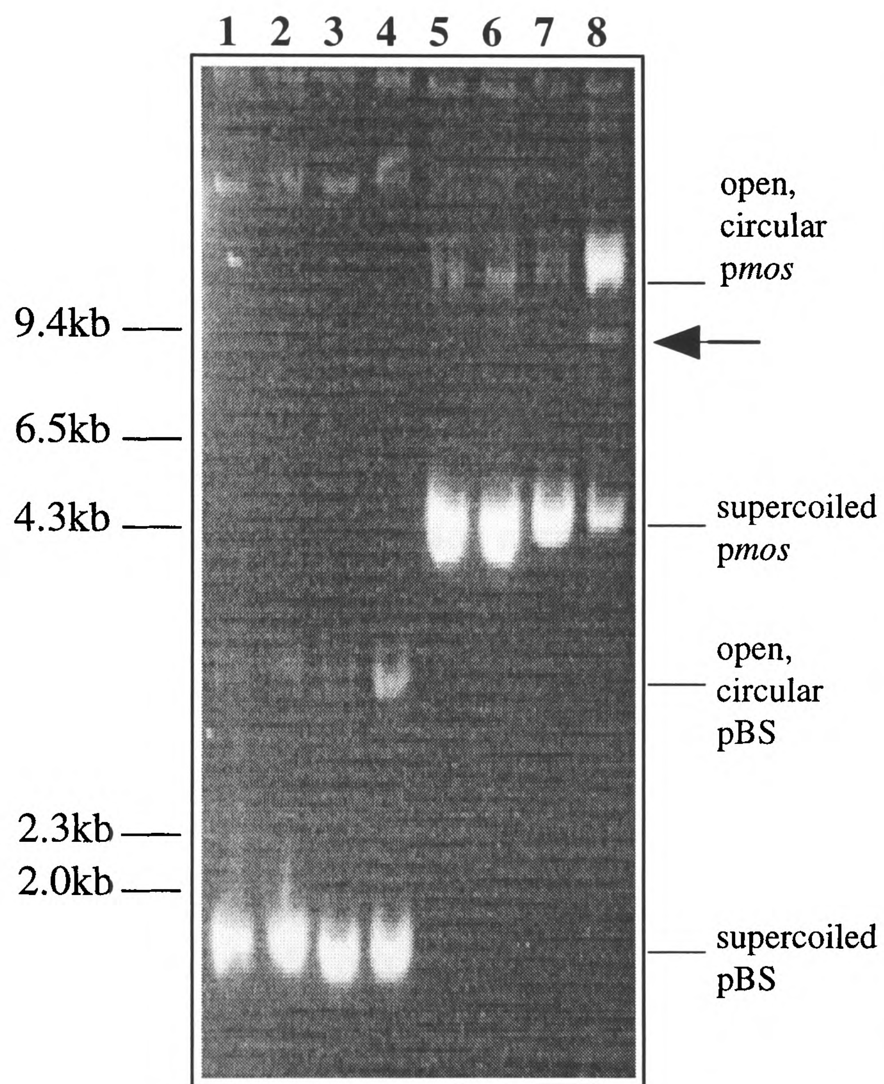
a) Ethidium bromide stained 1% agarose gel following fractionation of excision reactions.
b) Same gel following Southern blotting and hybridisation to ^{32}P labelled *mos*1 probe.

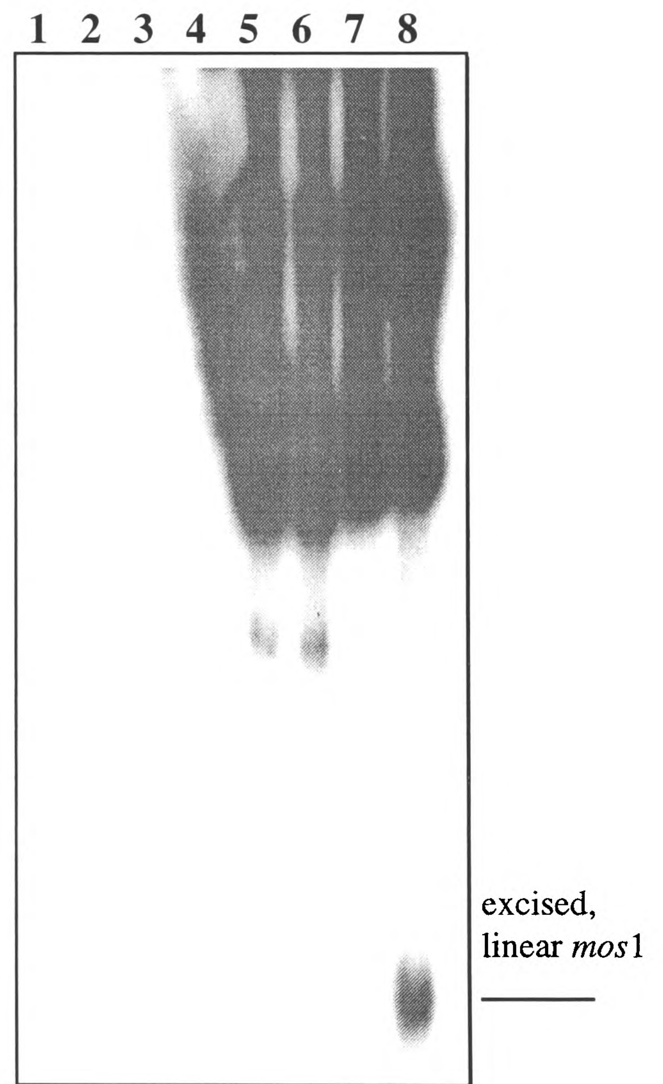
Lane 1........ pBS
Lane 2........ pBS incubated in T.E. - transposase
Lane 3........ pBS incubated in assay buffer - transposase
Lane 4........ pBS incubated in assay buffer + transposase
Lane 5........ p*mos*
Lane 6........ p*mos* incubated in T.E. - transposase
Lane 7........ p*mos* incubated in assay buffer - transposase
Lane 8........ p*mos* incubated in assay buffer + transposase

The arrow marks the position of the 8kb band obtained in the presence of transposase. This band may correspond to linear plasmid in which cleavage at a single end has occured.

a) Ethidium bromide stained agarose gel

b) Autoradiograph, following Southern blot analysis





the *mos*1 DNA sequence was added and hybridisation carried out overnight at 65°C (2.3.2.15). The membrane was then washed before exposure to X-ray film.

On analysis of the Southern blot autoradiograph, a 1.3kb band hybridising to the *mos*1 probe could be detected in lane 8 (figure 4.1b). This lane corresponds to the reaction in which all assay components were incubated. The *mos*1 hybridising band has a mobility which, within the limits of resolution of the gel, corresponds to that of a full length, excised, linear *mos*1 element. This experiment demonstrates that recombinant Mos1 protein (or a contaminating protein) is competent *in vitro*, to perform the endonucleolytic cleavage reactions which result in the excision of *mos*1 elements from flanking sequences.

4.2.2 A mock purification of transposase is unable to support excision of *mos*1 DNA from plasmid p*mos*

A mock purification of Mos1 protein was performed. *E. coli* BL21(DE3) competent cells were transformed with plasmid pBCP378. Conditions were then followed as described in 2.3.3.6, as though for overexpression and purification of recombinant Mos1 protein. As expected, fractions equivalent to those containing Mos1 protein in the previous purification failed to contain the 41kD band. Mock purified protein was then aliquoted and stored at -70°C. The mock purified protein was used to address the possibility that the excision event observed in figure 4.1b, lane 8, is due to the activity of an *E. coli* protein which had co-purified with Mos1 transposase.

A range of dilutions of transposase and mock extract were included in an excision assay. Excision reactions were carried out in standard buffer as previously described (4.2.1 and 2.3.5.1). The reactions were agarose gel fractionated and subjected to Southern blot analysis. Figure 4.2 shows the results of this experiment. On addition of transposase to the reaction, excision of a *mos*1 DNA hybridising band of the predicted size is observed. The intensity of excised product decreases as the concentration of transposase is reduced. On incubating mock purified transposase with the reaction no excision products are evident. This demonstrates that purified Mos1 protein, rather than an *E. coli* co-purifying contaminant is responsible for excision of the *mos*1 element from p*mos*.

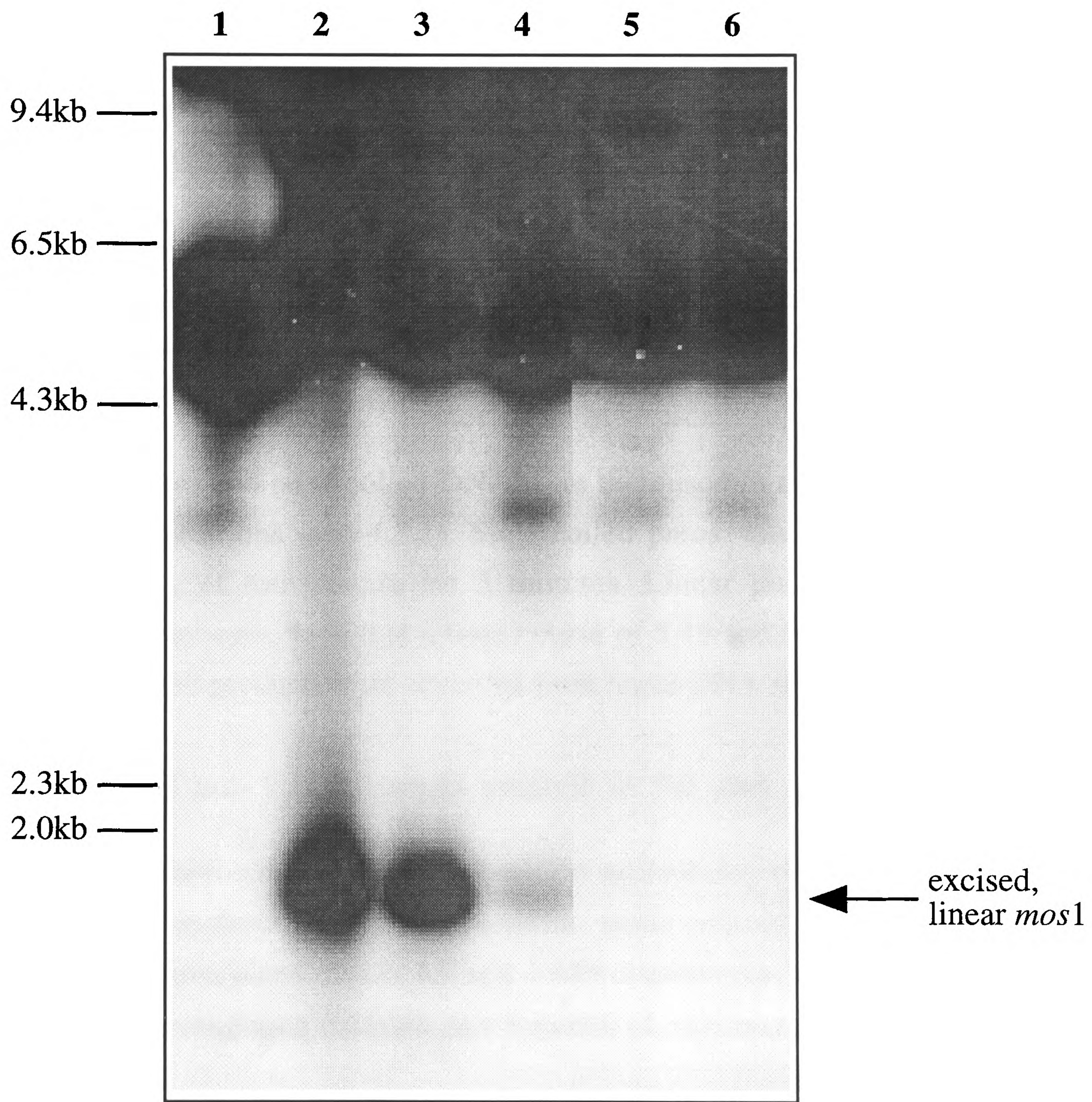
4.2.3 Mos1 transposase requires a divalent cation for excision activity

The requirement of Mos1 transposase for a divalent cation was analysed. Standard excision reactions were performed, as in 4.2.1, with the substitution of 5mM

Figure 4.2. Mock purified transposase is unable to support excision of *mos*1 from p*mos*.

All reactions contain 200ng of DNA. Fractionation was performed in 1% agarose prior to Southern blot analysis. λ *Hind*III molecular size standards used. See text for details.

Lane 1........ p*mos* incubated in assay buffer - transposase
Lane 2........ p*mos* incubated in assay buffer + 250ng transposase
Lane 3........ p*mos* incubated in assay buffer + 25ng transposase
Lane 4........ p*mos* incubated in assay buffer + 2.5ng transposase
Lane 5........ p*mos* incubated in assay buffer + mock extract (preparation 1)
Lane 6........ p*mos* incubated in assay buffer + mock extract (preparation 2)


1
2
3
4
5
6
9.4kb
6.5kb
4.3kb
2.3kb
2.0kb
excised,
linear *mos*1

MnAcetate for either 5mM $MgCl_2$, or 2.5mM EDTA. A time course for excision was then taken with samples being removed at 5, 15 and 30 minutes. No excision events were detected in the presence of 2.5mM EDTA, indicating that a divalent cation is essential for the endonucleolytic cleavage activity of the transposase. With Mg^{2+} rather than Mn^{2+} in the reaction buffer, a decrease in excision product is observed (figure 4.3). Quantification of this product using Molecular Dynamics Image Quant™ software has demonstrated that there is an 8-fold decrease of excised *mos*1 sequences when Mg^{2+} is present over Mn^{2+}.

4.2.4 Mos1 transposase requires a supercoiled DNA substrate for excision

To investigate the effect of altering the conformation of substrate on excision of *mos*1 elements, linear and supercoiled plasmids were employed in parallel reactions.

Plasmid p*mos* was digested with *Bam*HI for 1 hour at 37°C. *Bam*HI has a single restriction site in p*mos* which is 0.2kb from the right-hand end of the element. Following digestion, linear p*mos* was phenol extracted and ethanol precipitated before resuspension in TE and quantification.

200ng each of linear and supercoiled DNA were then used in an excision reaction, under standard conditions (see 4.2.1). Supercoiled p*mos* was incubated in assay buffer with 250ng of transposase for 5 minutes. Linear p*mos* was incubated, similarly, but in triplicate, such that a time course of 5,15 and 30 minutes could be followed. No excised products were observed from linear DNA substrate (figure 4.4).

4.2.5 Excision of *mos*1 is inversely related to the size of the element

Plasmid p*mos* contains a unique *Sal*I site between nucleotides 349-355. The *Sal*I site was employed to construct plasmids p*mos-white*, p*mos-yellow*, and pRJM348-*mos-tet*, which contain insertions of 4.0, 4.9 and 1.4kb respectively in *mos*1. A series of excision assays investigating excision as a function of increasing element size were then conducted.

The construction of p*mos-white* is described in detail in figure 6.1 and 6.2.2. P*mos-white* contains a 4kb insertion of the mini-*white* gene from *D. melanogaster*.

Plasmid p*mos-yellow* contains the *yellow* gene from *D. melanogaster*. Plasmid pBY was restricted with *Sal*I and agarose gel fractionated. A 4.9kb fragment containing the *yellow* gene was recovered from the gel by geneclean and quantified, before ligation

Figure 4.3. A divalent cation is required for *mos*1 transposase excision activity.

All reactions contain 200ng of DNA and were fractionated in 1% agarose prior to Southern blot analysis. λ *Hind*III molecular size standards used. See text for details.

Lane 1........ p*mos*
Lanes 2-4.... p*mos* + assay buffer + 250ng transposase + 5mM $MgCl_2$
Lanes 5-7.... p*mos* + assay buffer + 250ng transposase + 5mM MnAc
Lanes 8-10... p*mos* + assay buffer + 250ng transposase + 2.5mM EDTA

Reactions run in lanes 2,5 and 8 were incubated for 5 minutes; lanes 3,6 and 9 for 15 minutes and 4,7 and 10 for 30 minutes at $37^{o}C$. Samples were stored at $-20^{o}C$ prior to loading.

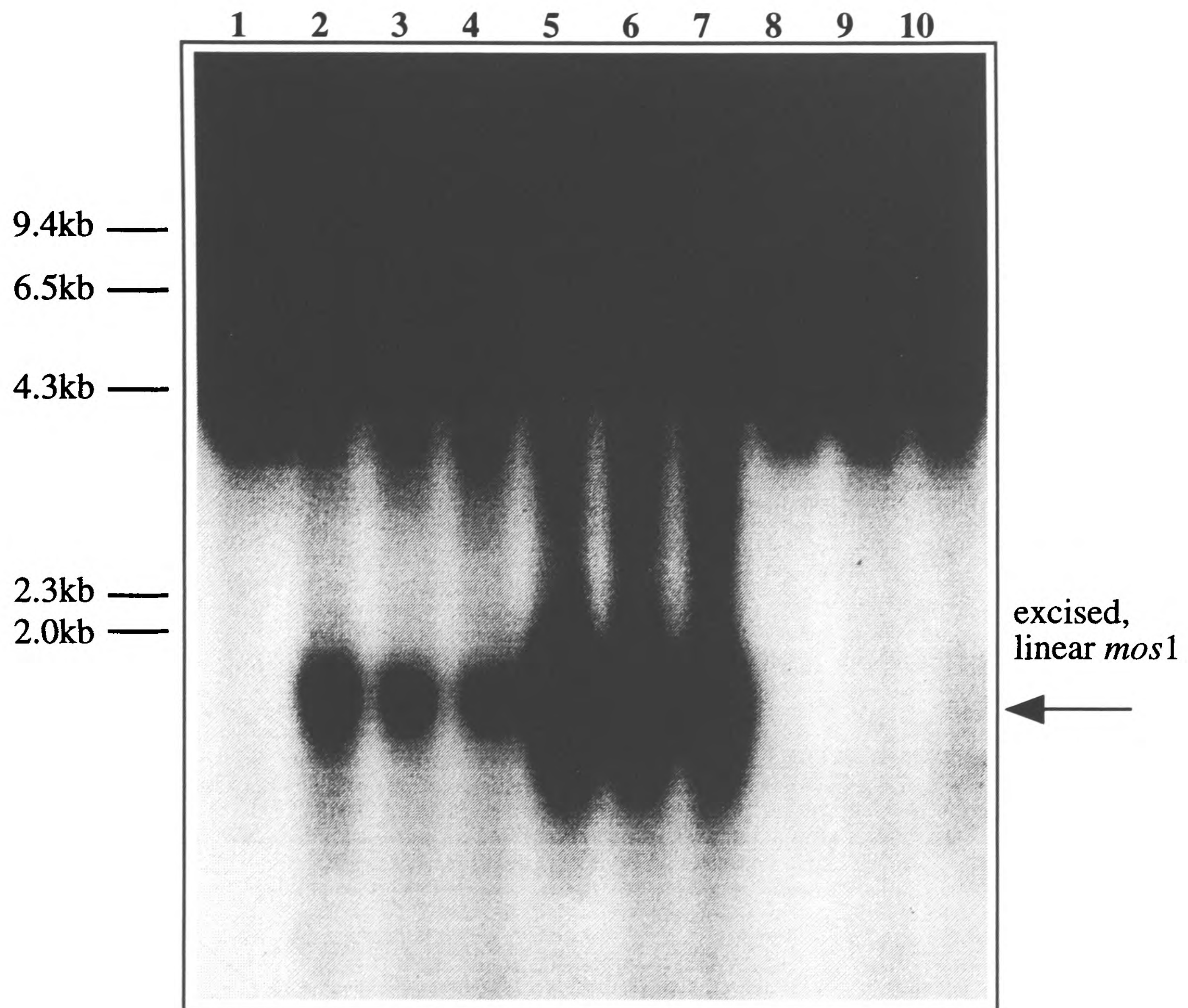
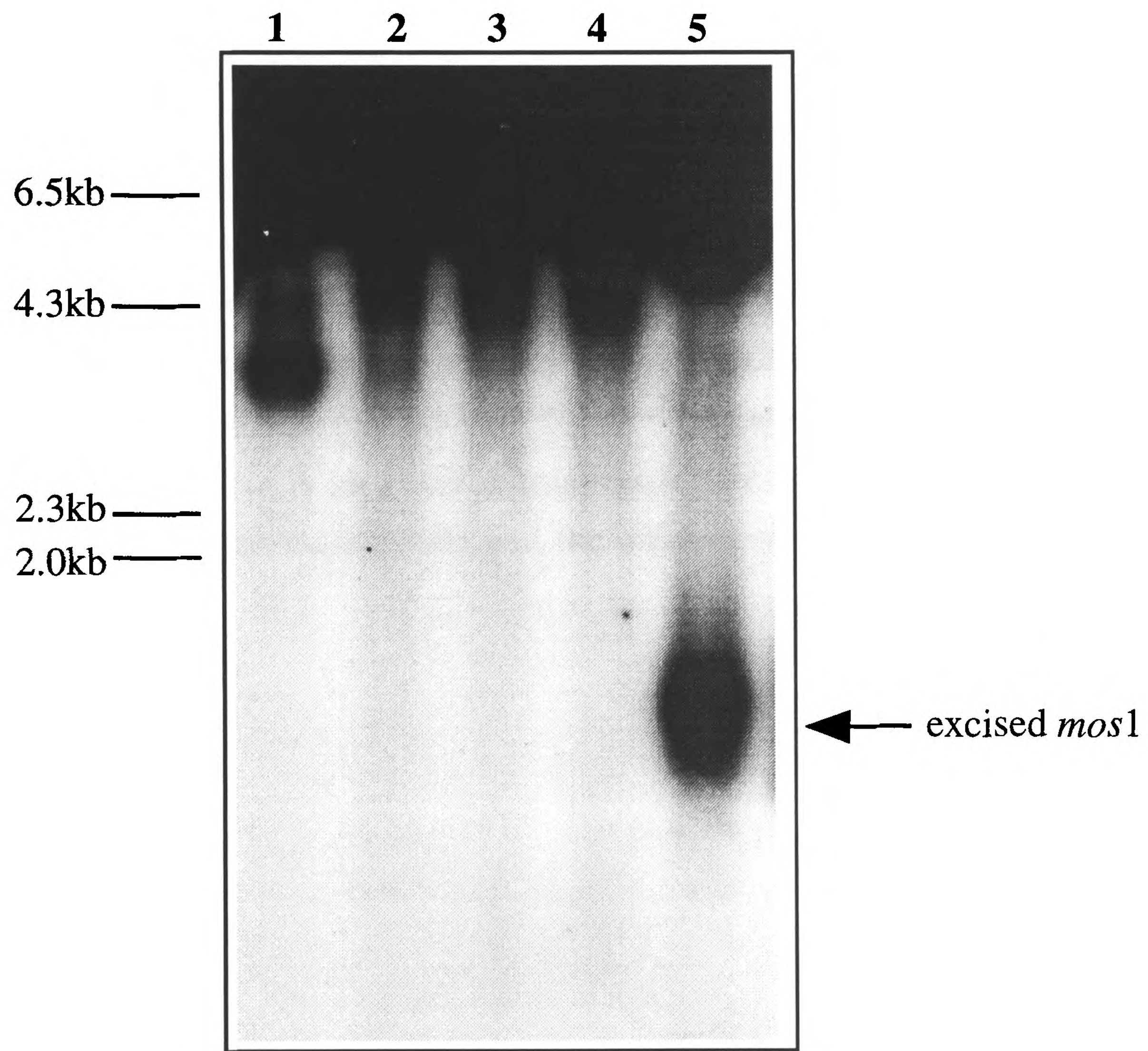

1
2
3
4
5
6
7
8
9
10
9.4kb
6.5kb
4.3kb
2.3kb
2.0kb
excised,
linear *mos*1

Figure 4.4. A supercoiled DNA substrate is required for transposase stimulated excision of the *mos*1 element.

All reactions contain 200ng of DNA and were fractionated in 1% agarose prior to Southern blot analysis. λ *Hind*III molecular size standards used. See text for details.

Lane 1........ p*mos*
Lanes 2-4.... linear p*mos* + assay buffer + transposase
Lane 5........ supercoiled p*mos* + assay buffer + transposase

Reactions run in lanes 2 and 4 were incubated for 5 minutes, lane 3 for 15 minutes and lane 4 for 30 minutes at 37°C. Samples were stored at -20°C prior to electrophoresis.


1
2
3
4
5
6.5kb
4.3kb
2.3kb
2.0kb
excised mos1

into *Sal*I digested, gel purified p*mos*. Competent *E. coli* NM522 cells were transformed with the ligation mixture and allowed to grow overnight on L-Amp plates. Colonies containing the *yellow* gene sequence were selected by picking colonies and performing *Sal*I digests on minipreparations of DNA.

Plasmid pRJM348-*mos-tet*, was made by Dr. Angela Dawson. A 5kb *Bam*HI/*Hind*III restriction fragment, containing *mos*1, from p*mos* was ligated into *Bam*HI digested pRJM348. The 1.4kb *Ava*I/*Eco*RI fragment containing the tetracycline resistance gene from plasmid pBR322 was then inserted into the *Sal*I site of *mos*1.

Excision assays using equimolar amounts of the different plasmids were then performed (0.04pmol, which corresponds to 200ng of p*mos*). A time course from 5 to 30 minutes was carried out as shown in figure 4.5. Excised products of the predicted mobility were observed from each construct. Molecular dynamics Image Quant^{TM} software was employed to analyse the percentage excision of the various *mos*1 constructs from plasmid at each of the three time points and a graph of percentage excision against element size plotted (figure 4.6). From analysis of the graph in figure 4.6, it appears that there is an inverse relationship between the size of the *mos*1 element used in the excision reaction and the amount of excision product detected.

Figure 4.5. Comparing the excision of different sized *mos*1 constructs.

All reactions contain 0.04pmol of DNA. p*mos* and pRJM348-*mos-tet* were fractionated in 1% agarose; p*mos-white* and p*mos-yellow* in 0.8% agarose, prior to Southern blot analysis. λ *Hind*III molecular size standards used. See text for details.

a)

Lane 1........ p*mos* + assay buffer
Lanes 2-4..... p*mos* + assay buffer + transposase
Lane 5........ pRJM348-*mos -tet* + assay buffer
Lanes 6-8..... pRJM348-*mos -tet* + assay buffer + transposase

b)

Lane 1........ p*mos -white* + assay buffer
Lanes 2-4..... p*mos -white* + assay buffer + transposase
Lane 5........ p*mos-yellow* + assay buffer
Lanes 6-8..... p*mos-yellow* + assay buffer + transposase

Reactions run in lanes 2 and 6 were incubated for 5 minutes, lanes 3 and 7 for 15 minutes and lanes 4 and 8 for 30 minutes at 37°C. Samples were stored at -20°C prior to loading.

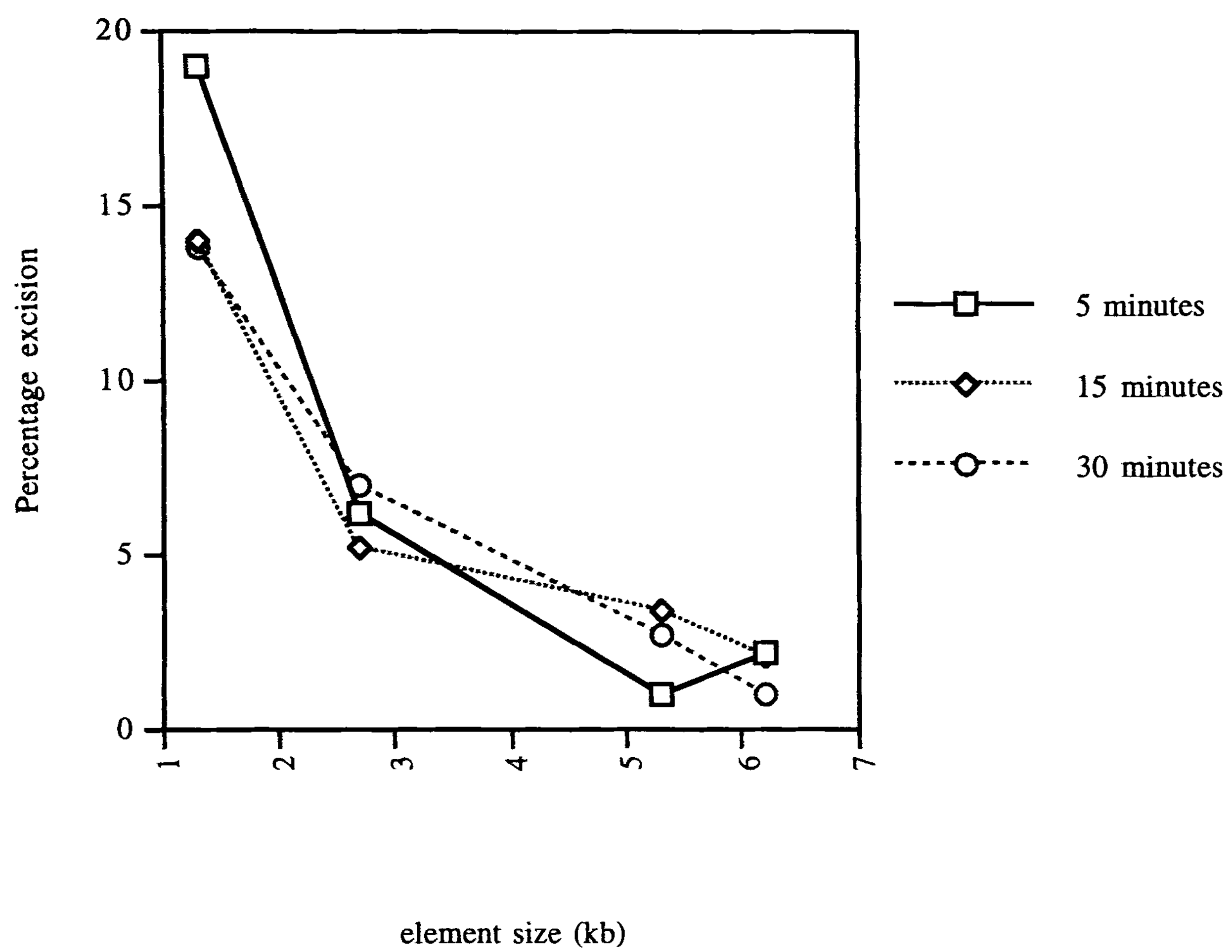


Figure 4.6. Percentage excised product as a function of element size, during *in vitro* excision of several *mos*1 constructs (see text for details).

4.3 Discussion

To test whether recombinant Mos1 protein is active *in vitro*, an assay was generated to examine the excision of *mos*1 elements from plasmids. On addition of Mos1 protein (transposase) to plasmid p*mos* in appropriate assay conditions, excised *mos*1 hybridising sequences corresponding in mobility to full length elements can be detected. Mos1 protein is both sufficient and necessary for the appearance of these excised bands (figure 4.1b, lane 8). This is apparent from figure 4.2, in which it is demonstrated that assay buffer containing mock purified transposase is unable to support excision.

The 8kb band detected in figure 4.1a, on addition of transposase to p*mos* may correspond to linear plasmid in which cleavage at a single transposon end has taken place. A similar result has been obtained on analysis of *Tc*1 excision *in vitro*. On addition of nematode extract containing *Tc*1 transposase to plasmids containing *Tc*1 elements, cleavage at either the left or right end of the transposon can be detected in addition to full length excised copies. Furthermore, deletion of one or other end of the transposon does not abolish cleavage at the remaining end (Vos *et al*, 1996).

The complete separation of transposon from flanking donor DNA has previously been observed for the prokaryotic elements Tn7 and Tn10. Molecules with single and double strand breaks between flanking DNA and the transposon left or right ends have also been detected (Benjamin and Kleckner, 1992; Bainton and Craig, 1993).

Mos1 transposase contains a nicking activity. This activity is observed in the relaxation of supercoiled pBS and p*mos* plasmids in figure 4.1a, lanes 4 and 8 respectively. Relaxation of supercoiled p*mos* is not observed on incubation with mock extract (data not shown) demonstrating that this activity is most likely attributable to transposase. *Mariner* elements generally insert into TA dinucleotides. The activity observed may reflect single strand cleavage events occurring at TA residues, in p*mos* and pBS, which are potential sites of integration for *mos*1.

The *Tc1/mariner* family of elements contain a D,D35E domain in their encoded transposases. This domain is also present in transposases of the prokaryotic elements Tn7, Tn10 and Mu transposase as well as retroviral and retrotransposon integrases (see 1.2.2.4). The acidic triad of amino acids are thought to be part of the catalytic site of the transposase and to be involved in binding divalent metal cations. In some of the crystals of the retroviral integrases metal has been found near these acidic residues (Dyda *et al*, 1994; Bujacz *et al*, 1996).

The requirement of Mos1 transposase for a divalent cation was studied. On substitution of Mn^{2+} for Mg^{2+} in the reaction buffer, an eight-fold reduction in excision product was observed. Excision was completely abolished in the presence of 2.5mM EDTA demonstrating that a metal ion is required. A similar decrease in activity on changing Mn^{2+} to Mg^{2+} is also observed for transposition of *mos*1 *in vitro* (Dawson, personal communication). A parallel reduction in transposase activity has been demonstrated for Tc1A on substitution of Mn^{2+} for Mg^{2+} in an assay examining the position of the 5' cleavage site of Tc1 (Vos and Plasterk, 1994). Excision and transposition reactions of both *Tc*1 and Himar1 elements *in vitro* require a metal ion (Vos *et al*, 1996; Lampe *et al*, 1996).

It is thought that Mg^{2+} is likely to be the relevant cofactor associated with the active site *in vivo*. However, some of the transposases so far studied only show detectable activity in the presence of Mn^{2+}. The effects of Mn^{2+} in reducing the stringency of reactions has been documented for a number of proteins including HIV integrase, the Rag1 and 2 recombinases, and Tn10 and *Himar*1 transposases (Engelman *et al*, 1991; Vink *et al*, 1991; van Gent *et al*, 1996; Eastman *et al*, 1996, Hanniford and Junop, 1996, Lampe *et al*, 1996).

Excision of *mos*1 is completely abolished if the substrate is linear. This suggests that co-ordinated cutting at both ends of *mos*1 is stimulated by supercoiling of the substrate. With *Tc*1, excision of complete elements is reduced two fold when the nature of the substrate is changed from supercoiled to linear. However, no decrease in efficiency of cleavage of a single end of this transposon is observed if the substrate is linear (Vos *et al*, 1996). The authors suggest that the majority of complete excision products observed with a linear substrate can be explained by non-coordinated cutting at either end. Mu and Tn10 require supercoiled substrate DNA for transposition. This topological form of DNA is believed to energetically favour a particular configuration which transposon ends must adopt in a synaptic complex (Craigie and Mizuuchi, 1986). In the absence of supercoiling, IHF can compensate, during Tn10 transposition (see Kleckner *et al*, 1996).

On analysing the molar ratios of excised *mos*1 to non-excised element in figures 4.5 and 4.6, it is apparent that in general an inverse relationship exists between increasing transposon size and excision frequency. A similar relationship has been observed on analysis of the frequency of transposition of *mos*1 and *mos*1 derivatives in *D. melanogaster*, in germline transformation experiments (Garza *et al*, 1991; Lidholm *et al*, 1993; Lohe *et al*, 1995).

An inverse relationship between insert size and transposition frequency has previously been reported for the *P* element in experiments transforming the germline of *D. melanogaster*. Despite sources of variation between experiments, the transposition frequency decreases on average as a function of size. On transformation of *Drosophila* with an 8kb transposon roughly 30-60% of G_0 tested contained the insert. An 18kb transposon was taken up in the germlines of between 5 and 14% of G_0, and 1-2% fertile Go resulted on transformation with a cosmid transposon of ~40kb (see Spradling, 1986).

Recombinant transposase used to stimulate the excision of *mos*1 elements from plasmids as described, is able to support *in vitro* transposition of a tetracycline marked *mos*1 element from plasmid pRJM-*mos-tet* to pBluescribe. Insertions obtained are flanked by TA target site duplications (A. Dawson, personal communication).

Chapter 5

Analysis of excised *mariner* elements

5.1 Introduction

Members of the *Tc1/mariner* family of elements are believed to transpose via a DNA mediated 'cut and paste' mechanism, in which an element is physically excised from the chromosome before insertion at a new genomic site (Plasterk, 1991; Plasterk and Groenen, 1992). Consistent with such a mechanism of transposition, extrachromosomal copies of several members of this family, including *Tc3* and *Tc1* from *C. elegans* (Ruan and Emmons, 1984; Rose and Snutch, 1984; van Luenen *et al*, 1993), and *minos* from *D. Hydei* (Arca *et al*, 1997) have been detected *in vivo*.

Linear *Tc3* forms have been observed after induction of *Tc3* transposase expression by heat shock in transgenic *C. elegans* lines. Determination of the exact sequences at the ends of these elements has resulted in the suggestion that they are the active intermediates in *Tc*3 transposition (van Luenen *et al*, 1994). On recovery from heat shock other extrachromosomal forms are also detectable (van Luenen *et al*, 1993).

Relaxed and supercoiled *Tc*1 circular monomers have been observed *in vivo* (Ruan and Emmons, 1984; Rose and Snutch, 1984; Radice and Emmons, 1993). *Tc1* circles represent a heterogeneous population of molecules. They are derived from joining of the transposon ends, some of which contain deletions in one or both inverted repeats. The predominant extrachromosomal species of *Tc*1 detected *in vivo* is a linear monomer (Ruan and Emmons, 1984). Linear, excised copies of *Tc*1 can also be detected *in vitro*. On addition of nematode nuclear extract, enriched for Tc1 transposase, to plasmids containing *Tc*1 sequences, the excision of *Tc*1 from flanking DNA is stimulated. Mapping of the *in vitro* cleavage sites at the nucleotide level has revealed that this species is the likely intermediate in transposition of *Tc*1 (Vos *et al*, 1996).

Following heat shock induction of minos transposase in *Drosophila*, extrachromosomal species of *minos* can been detected. Restriction analysis of these forms has demonstrated that they correspond to relaxed circular copies (Arca *et al*, 1997).

To investigate the mobilisation of *mariner* elements *in vivo*, free forms of the transposon were looked for. In this chapter the detection of extrachromosomal, circular copies of *mariner* elements is described and evidence regarding their precise structure presented. The structure of excised, linear *mos*1 elements obtained *in vitro*, on addition of recombinant transposase to plasmids containing *mos*1 is also described.

The structures of these two forms have been compared and a model for the transposition of *mos*1 and *mariner* elements in *D. simulans* is discussed.

5.2 Results

5.2.1 Analysis of *mariner* elements excised *in vivo*

5.2.1.1 Detection of extrachromosomal *mariner* sequences

To try and identify potential intermediates in *mariner* transposition *in vivo*, fractionation of DNA from strains of *Drosophila* containing active *mariner* elements was performed, such that separation of extrachromosomal elements from low mobility genomic DNA could be achieved.

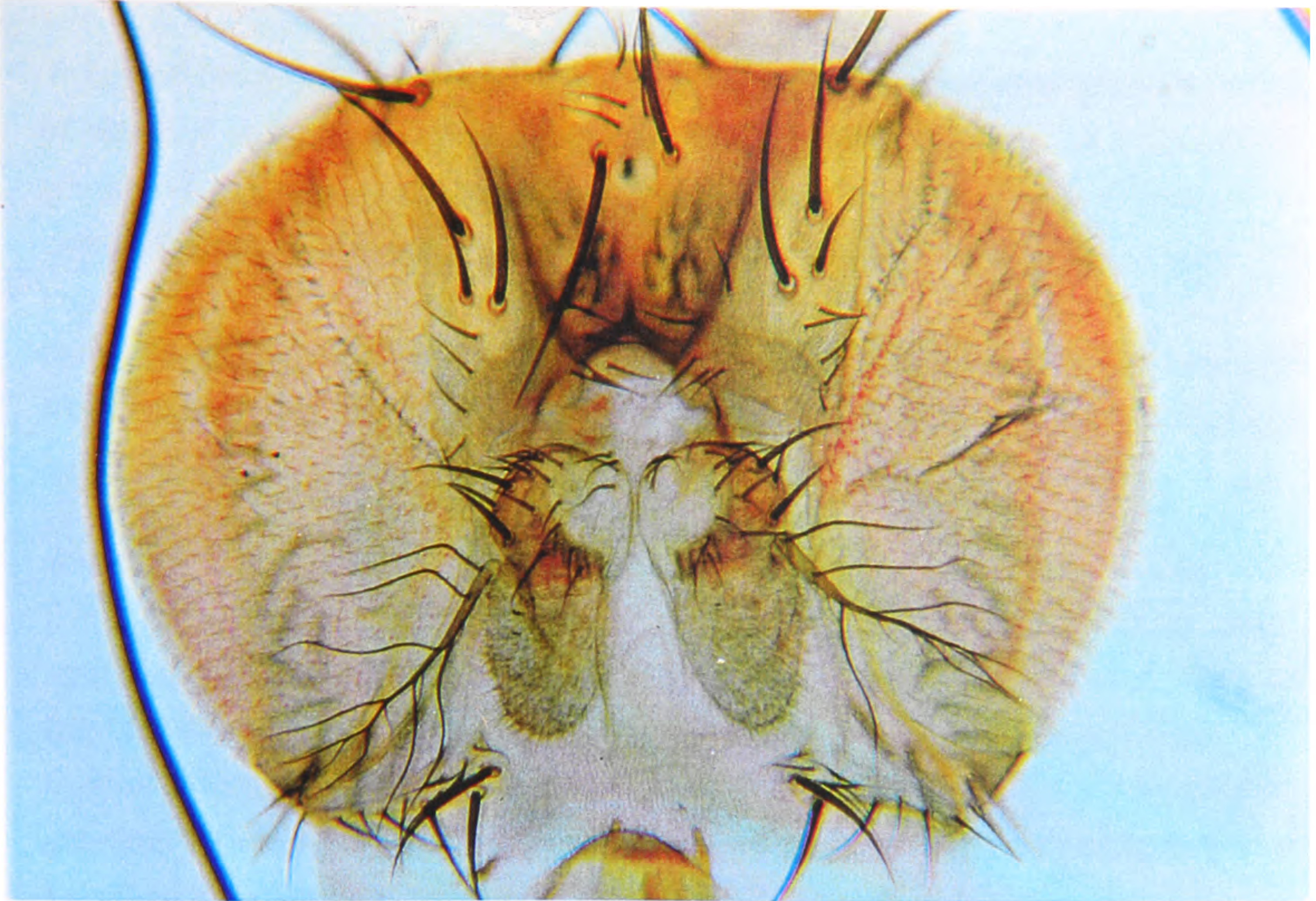
The w^{pch} allele from *D. mauritiana*, which contains the original *mariner* element inserted into the 5' untranslated region of the *white* gene, has previously been transferred to a $white^{-}$ strain of *D. simulans,* by a series of backcrosses (Bryan and Hartl, 1988). The *mariner* element from w^{pch} is non autonomous and can not transpose unless active transposase is supplied *in trans*. The w^{pch} allele is stable in this strain, number 244, as no other copies of *mariner* are present. Figure 5.1 demonstrates the eye colour phenotype from this strain.

D. simulans strain 238 contains several *mariner* elements, at least one of which is active and can complement *peach in trans.* This is evidenced in crosses between strains 238 and 244. Somatic excision of *peach* restores pigment production in the cells from which it has been removed and can be detected by the presence of wild type patches of eye colour on a peach coloured background. This effect is called somatic mosaicism and is shown in figure 5.1.

Whole genomic DNA preparations were made from *D. simulans* strains; 238, 244, somatic mosaics obtained by crossing 238 and 244, and *yw*, as well as *D. melanogaster* OreganR, which has no *mariner* elements (Maruyama and Hartl, 1991a). In general, *D. simulans* populations contain *mariner* elements (Capy *et al*, 1992), although it was not known prior to this experiment if they were present in *D. simulans yw*.

10μg of undigested DNA from each preparation was fractionated on 1% agarose gel to determine if high mobility species distinct from genomic DNA could be detected. Following electrophoresis, the DNA was Southern blotted onto nitro-cellulose and

1



2

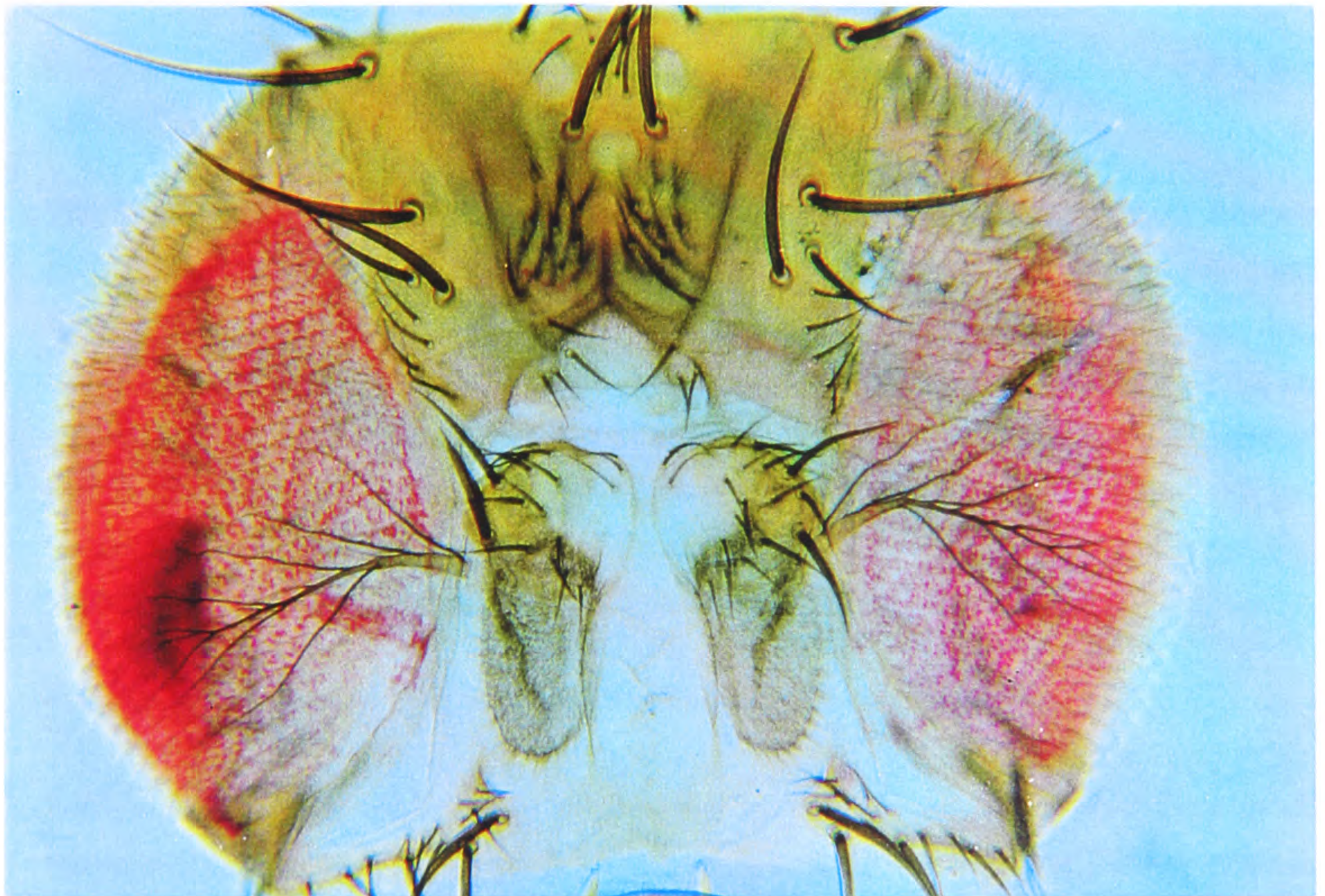


Figure 5.1 Eye colour phenotypes of *D. simulans strain* 244 and somatic mosaics.

1......... Peach coloured eyes observed in *Drosophila* homozygous for the w^{pch} allele

2......... Mosaic eyes due to excision of *peach* from the w^{pch} allele in some somatic cells. Wild type eye colour is restored in these cells. See text for details.

hybridised to a ^{32}P labelled *mos1* probe. As the DNA is undigested, genomic copies of *mariner* remain close to the wells as shown in figure 5.2. The intensities of the genomic DNA bands between lanes presumably indicate the relative differences in copy number of *mariner* between the strains. Females from strain 244 contain two copies of *peach* per genome and males one copy, as the w^{pch} allele is situated on the X chromosome. Strain 238 has a higher copy number of *mariner* elements than 244. Strain *yw* has less. As expected, no *mos1* cross hybridising band was observed in lane 5, which contains DNA from *D. melanogaster* OreganR.

Several, low molecular weight species that migrated on the gel ahead of the genomic DNA could be detected in lanes 1 and 2, figure 5.2. The high mobility *mos1* hybridising bands observed may correspond to extrachromosomal copies. They are only seen in genomic DNA preparations from somatic mosaic flies and strain 238, both of which contain active *mariner* elements. Strains 244 and *yw* contain only genomic copies of *mariner.* The control DNA from *D. melanogaster* OreganR did not hybridise to the *mariner* probe.

5.2.1.2 Separating extrachromosomal *mariners* from genomic DNA by centrifugation

The high mobility forms of *mariner* observed in figure 5.2 may contain circular or linear forms, or both. Sucrose gradient ultra centrifugation was performed to separate these copies of *mariner* away from the high molecular weight genomic DNA so that initial characterisation of the elements by diagnostic restriction digestion could be carried out. Fractions were assayed for the presence of low molecular weight *mariner* molecules by Southern blot analysis. However, the genomic DNA could not be satisfactorily separated from the low molecular weight forms of *mariner* (data not shown). A more sensitive, PCR based approach, was then adopted.

5.2.1.3 Detection of circular *mariner* elements by PCR

A PCR strategy was employed to amplify products from extrachromosomal copies of *mariner*. Primers were designed to anneal within *mos1*, but to prime DNA synthesis towards the ends of the element. Primers directed outwards in *mos1* will only amplify a subset of elements, namely those that contain nearby adjacent left and right ends. Such a subset includes circular, extrachromosomal copies of *mariner* and copies of *mariner* which are inserted close together in the genome. This is illustrated in figure 5.3.

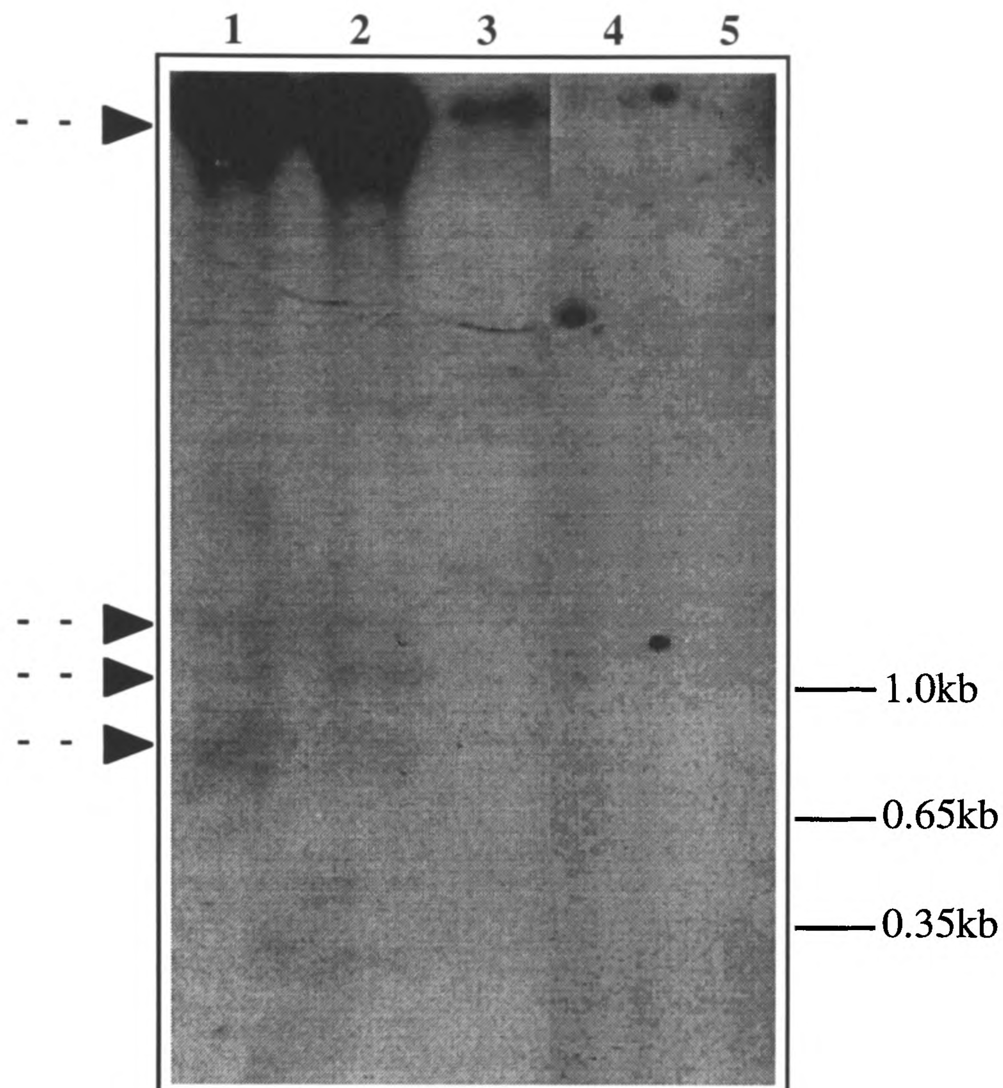


Figure 5.2. Southern blot of fractionated, whole genomic DNA preparations from several *Drosophila* strains, hybridised to a *mos*1 probe.

Dashed arrows show positions of *mos*1 hybridising bands. Size markers are from a *Sal*I partial digest of DNA corresponding to the *mos*1 sequence.

Genomic DNA prepared from:

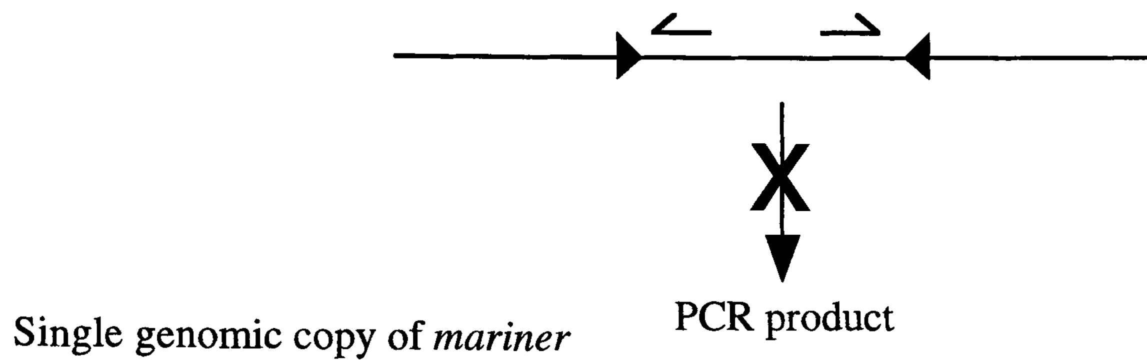
Lane 1........ Somatic mosaic *D. simulans*, which contain active *mariner* elements,

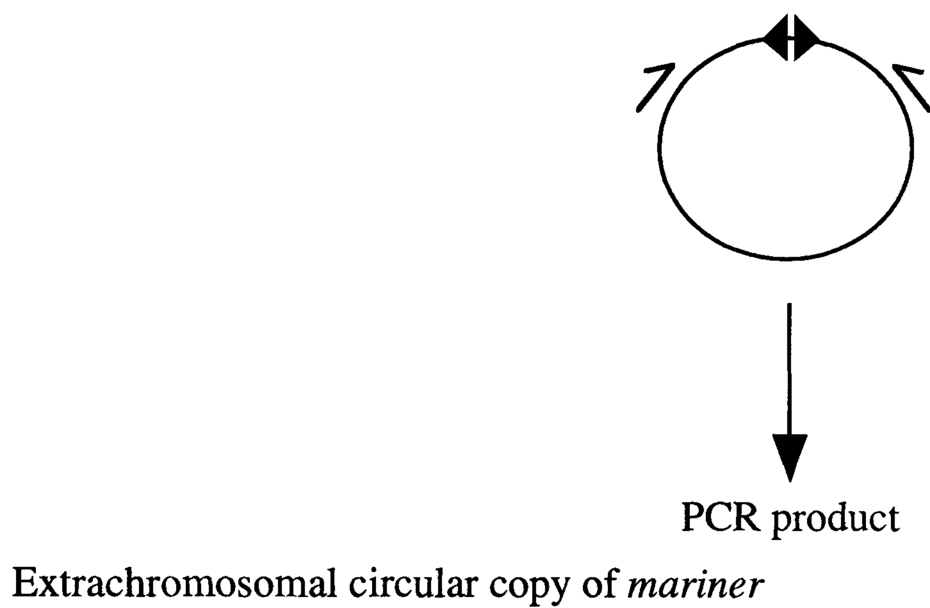
Lane 2........ *D. simulans* strain 238, also containing active *mariner* elements,

Lane 3........ *D. simulans* strain 244, which has only inactive *mariner* elements,

Lane 4........ *D. simulans yw*,

Lane 5........ *D. melanogaster* OreganR, which has no *mariners*.





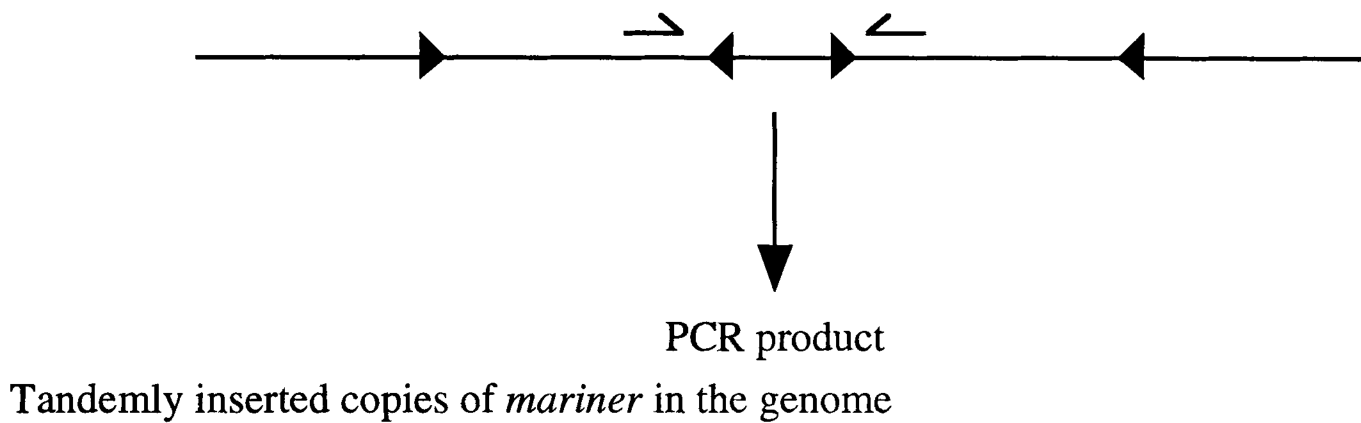


Figure 5.3 Amplification of a subset of *mariner* elements using primers designed to direct DNA synthesis towards the ends of the element. Only circular, extrachromosomal copies of *mariner* and copies of *mariner* which are inserted close together in the genome can serve as templates.

Primers P3855 and P3856, each approximately 250bp from the end of *mos*1 and directed outwards were employed in PCR reactions using 1μg of whole genomic DNA from *D. simulans* strains; 238, 244, *yw*, and somatic mosaics and *D. melanogaster* OreganR as template. The PCR cycle 94°C, 1 minute; 50°C, 1 minute; 72°C, 1 minute repeated for 30 cycles was followed. The size of PCR products predicted if complete, ligated, extrachromosomal copies of *mariner* are used as template is 500bp. If tandemly inserted genomic copies are used as template PCR products will be 500bp in length plus the distance between the adjacent ends. The results of amplification are shown in figure 5.4 after Southern blotting and hybridisation to ^{32}P labelled *mos*1 sequence. PCR products of 500bp, the size predicted for fragments amplified from free, circular forms were only observed with templates from strains 238 or somatic mosaics as seen in lanes 1 And 2, figure 5.4. This suggests that at least one of the high molecular weight bands in figure 5.2 contains circular *mariners*. No PCR products were seen using DNA from flies which contain inactive *mariner* (strain 244), very low copy number *mariner* (*D. simulans yw*), or no *mariner*, (*D. melanogaster* OreganR). Two additional bands were amplified from DNA from strain 238, the possible origins of these will be discussed later.

5.2.1.4 Amplification is across circular *mariner* elements not tandemly inserted genomic copies

PCR products of roughly 500bp would be expected if complete, extrachromosomal circular copies of *mariner* provided the template. When strain 238 was used as template two further PCR products were observed, one of lower and one of higher molecular weight than anticipated. These products may have resulted from amplification of circular *mariner* elements which had excised from the genome such that either some flanking DNA was also excised, or some *mariner* sequence was missing. Another explanation is that amplification from copies of *mariner* inserted close together in the genome has taken place. Indeed, *mariner* elements inserted adjacent to each other could also have provided the template for the 500bp PCR product since unfractionated DNA preparations were used to provide templates.

To determine whether *mariner* dimers in genomic DNA or extrachromosomal *mariner* elements were used as template, PCR reactions were performed on agarose fractionated DNA. Whole genomic DNA prepared from *D. simulans* was initially separated in 1% agarose as in figure 5.2. The gel was then divided into seven equal slices. Each slice was subdivided into three and used as a source of template in PCR reactions using three different sets of primers, as illustrated in figures 5.5 and 5.6.

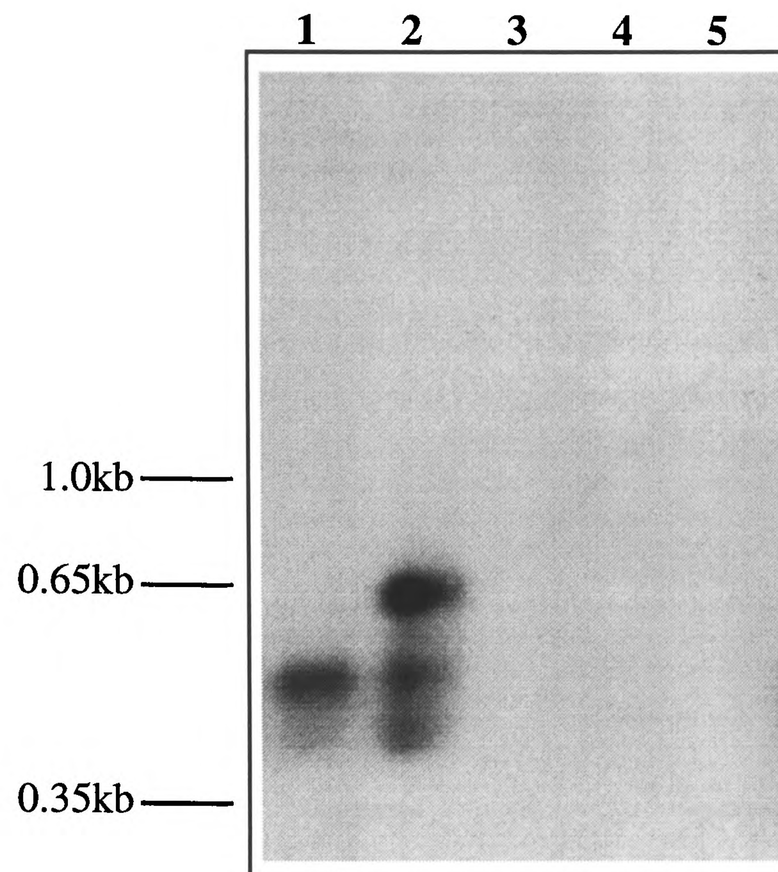


Figure 5.4. Products of amplification from different strains of *Drosophila*, using primers P3855 and P3856.

1μg of whole genomic DNA was used as template in each reaction. Products were fractionated on 1% agarose, Southern blotted and hybridised to a radiolabelled *mos*1 sequence. Size markers are from a *Sal*I partial digest of DNA corresponding to the *mos*1 sequence.

Lane 1......... PCR template is from somatic mosaic *D. simulans*,
Lane 2......... PCR template is from *D. simulans* strain 238,
Lane 3......... template is from *D. simulans* strain 244,
Lane 4......... template is from *D. simulans yw*,
Lane 5......... template is from *D. melanogaster* OreganR.

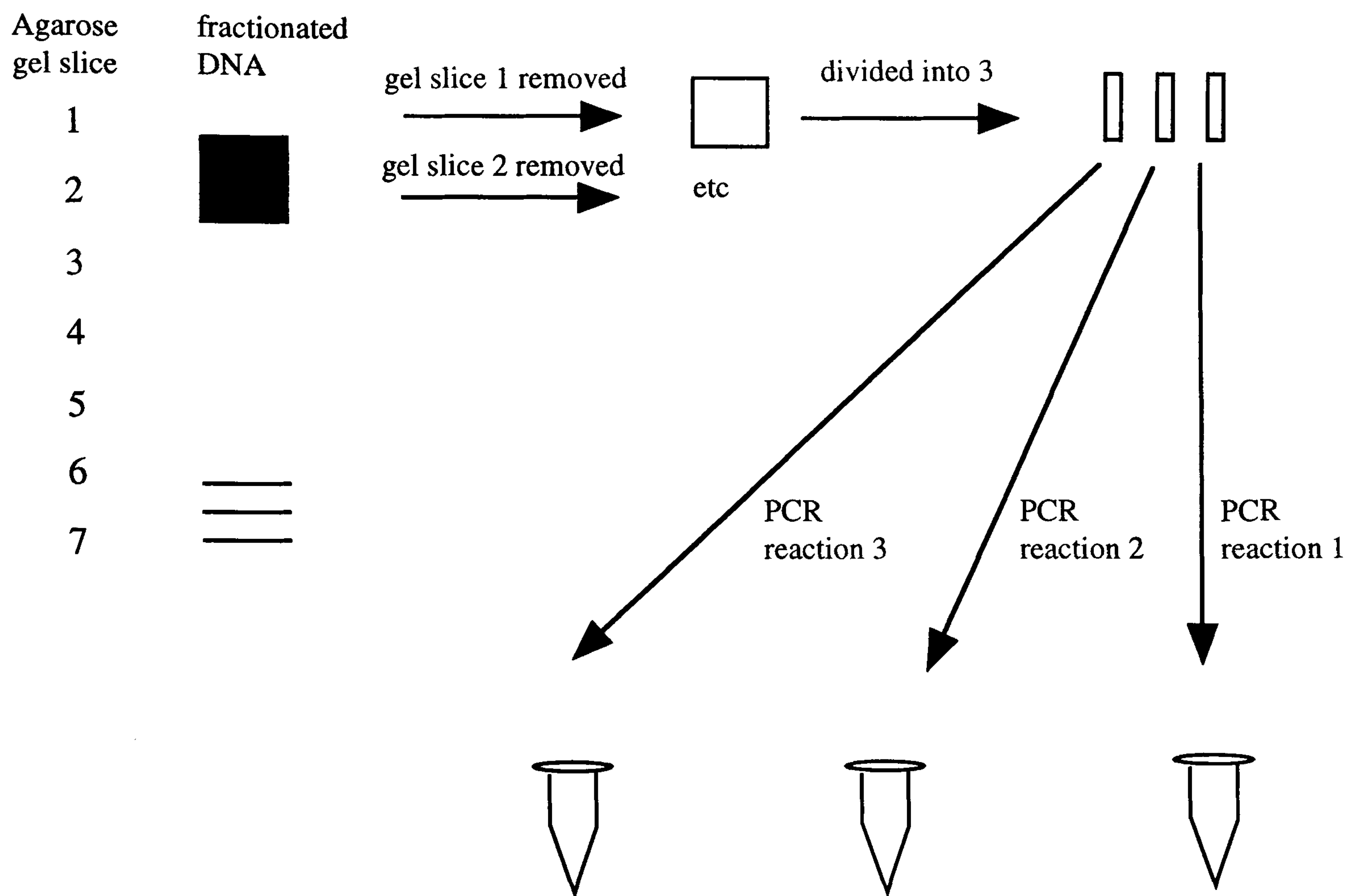
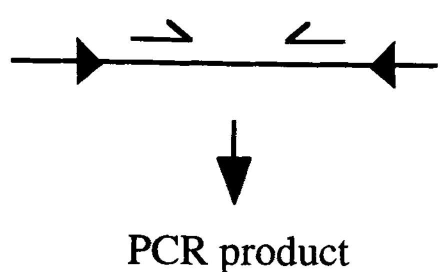
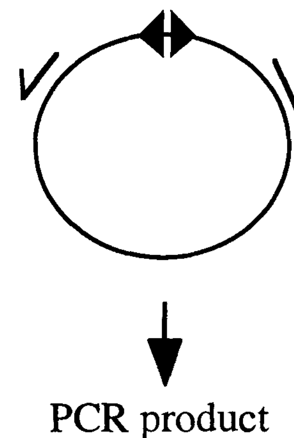
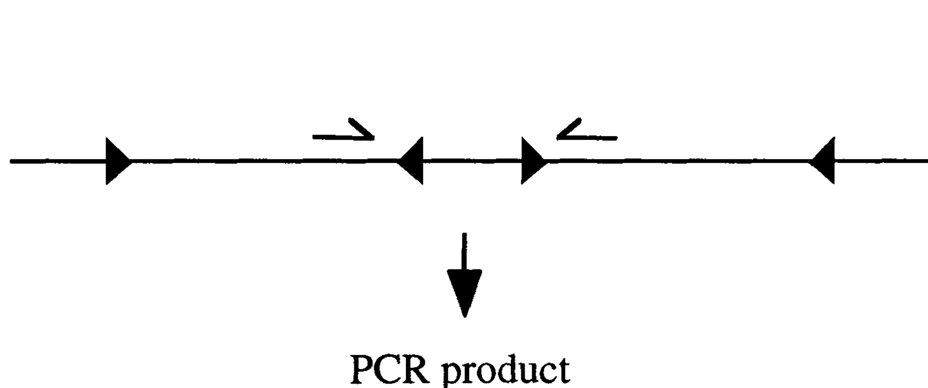


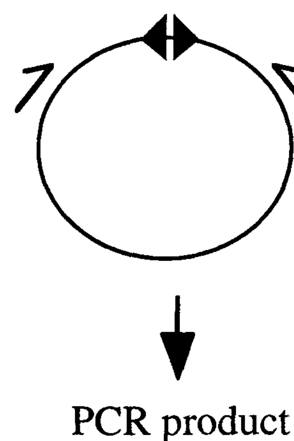
Figure 5.5. Demonstration of the removal of slices of agarose gel fractionated DNA, for use in PCR. See figure 5.6 for nature of PCR reactions.



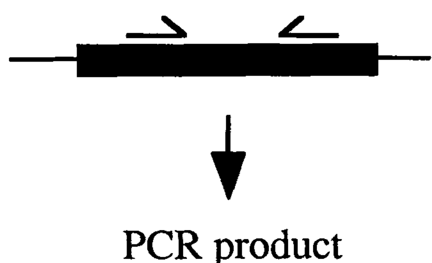


PCR 1. Using primers 337 and 338, directed inwards in *mariner*





PCR 2. Primers P3855 and P3856, directed towards the ends of *mariner*



PCR 3. Primers AJ1 and AJ2 for *atonal*

Figure 5.6. Diagram illustrating the DNA species that may act as template in PCR reactions with specific sets of primers.

PCR reaction 1 uses primers 337 and 338, directed inwards in *mariner*, and will amplify from all copies.
The second PCR reaction uses primers P3855 and p3856, directed outwards in *mariner*, these primers will only amplify from circular and tandemly inserted copies.
The third reaction uses primers AJ1 and AJ2 for *atonal*, and will amplify from all copies of *atonal*.

Primers 337 and 338 were employed in the first set of PCR reactions. The reaction conditions were 94°C, 1 minute; 50°C, 1 minute; 72°C, 1 minute repeated for 30 cycles. Primers 337 and 338 are designed to direct DNA synthesis inwards in *mariner* (see figure 5.6) and can amplify all *mariner* sequences whether linear or circular. The products of PCR reactions using each of the seven fractions were separated by agarose gel electrophoresis and then Southern blotted and hybridised to radio labelled *mos*1 sequence as shown in figure 5.7. From figure 5.7 it is apparent that *mariner* sequences are present in each agarose slice.

The second reaction used primers P3855 and P3856 directed outwards in *mariner*. These primers as described in 5.2.1.3 can amplify from both circular *mariner* molecules and DNA containing tandemly inserted *mariner* elements. Figure 5.8 demonstrates that PCR products are only observed when gel slices 6 and 7 are used as template (figure 5.8). Gel slices 6 and 7 correspond to a DNA template of low molecular weight. The source of the PCR products obtained with P3855 and P3856 is therefore likely to be circular copies of *mariner*, rather than copies which are close together in the genome.

It is possible that the population of circular *mariner* elements that exist *in vivo* in strain 238 are heterogeneous in size. The presence of molecules which contain either additions or deletions of sequence at the transposon termini, could serve as different sized templates for PCR, thus providing an explanation for the presence of extra PCR products obtained in lane 2, figure 5.4.

A third PCR reaction was performed, as a control, using each of the gel slices and primers AJ1 and AJ2 which amplify sequences from the *Drosophila* proneural gene *atonal*, which is not a transposable element. Fractionating the products of this PCR reaction revealed that the genomic DNA is retained close to the wells, since only slices 1, 2 and 3 contain template (figure 5.9).

5.2.1.5 Structures of joined *mariner* ends

The circular *mariner* elements detected by PCR could be intermediates in transposition or by products. A competent extrachromosomal transposition intermediate would be expected to contain all of the transposon sequence. In general, the *Drosophila* genome contains *mariner* elements which are homogeneous in length (Maruyama and Hartl, 1991) and thus most or all transposition events must generate a complete element. In order to determine whether extrachromosomal *mariner* elements consist of complete

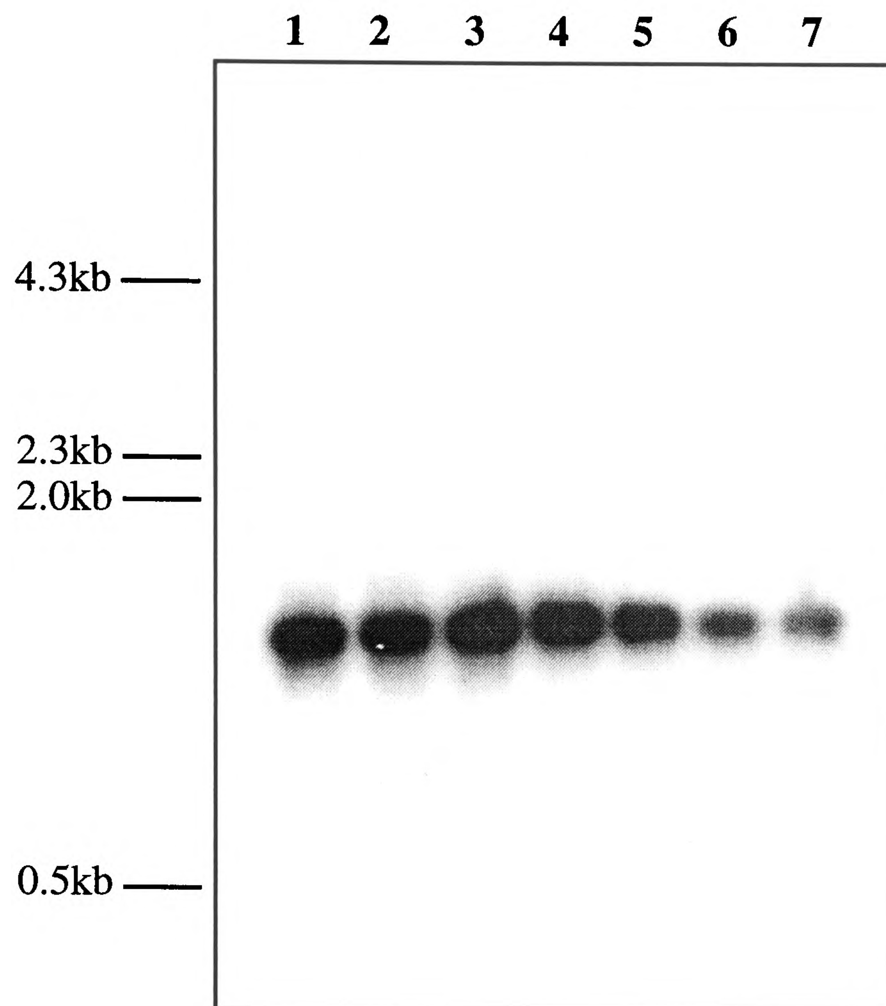


Figure 5.7. Amplification of DNA in agarose gel slices, using primers 337 and 338 directed inwards in *mariner*

Lane 1....PCR template contained in gel slice one (see figure 5.5)

Lane 2.... template contained in gel slice two

Lane 3.... template contained in gel slice taken three

Lane 7.... template contained in gel slice seven

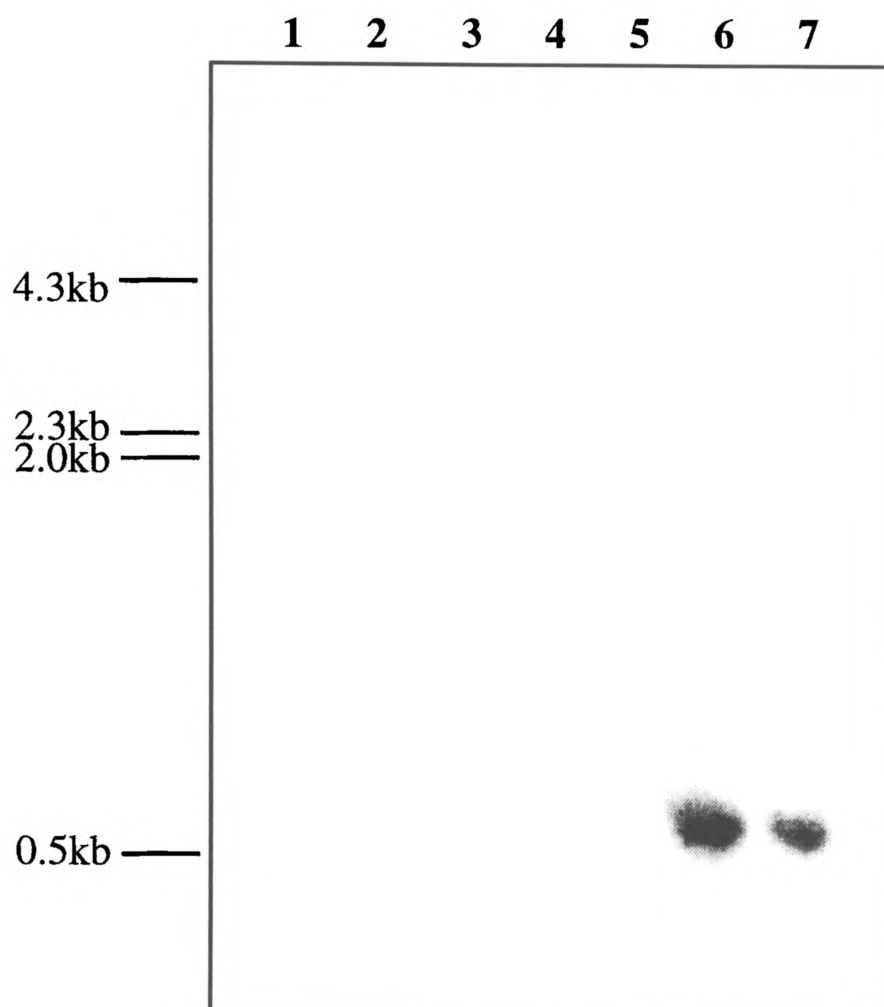


Figure 5.8. Amplification of DNA in agarose gel slices, using primers P3855 and P3856 directed outwards in *mariner*

Lane 1....PCR template contained in gel slice one (see figure 5.5)

Lane 2.... template contained in gel slice two

Lane 3.... template contained in gel slice taken three

Lane 7.... template contained in gel slice seven

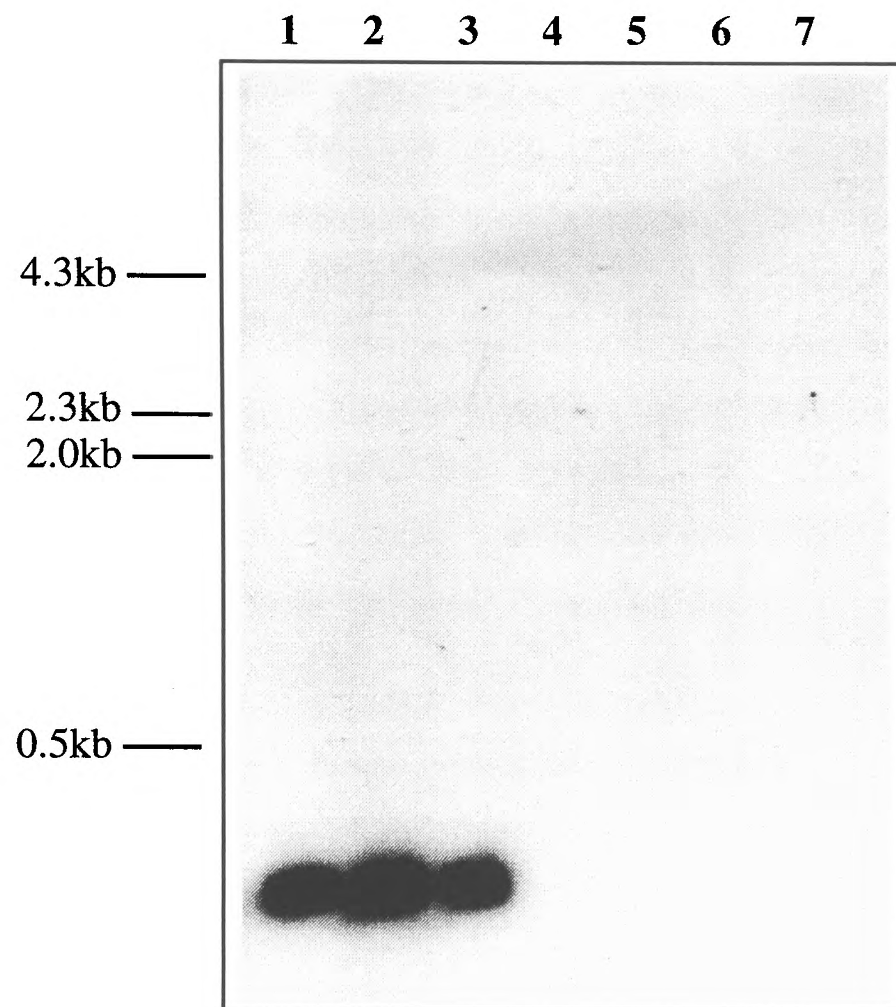


Figure 5.9. Amplification of DNA in agarose gel slices using primers AJ1 and AJ2, to amplify the proneural gene *atonal*.

Lane 1....PCR template contained in gel slice one (see figure 5.5)

Lane 2.... template contained in gel slice two

Lane 3.... template contained in gel slice taken three

Lane 7.... template contained in gel slice seven

circularised copies of *mariner*, PCR products containing the site of circularisation were cloned and sequenced.

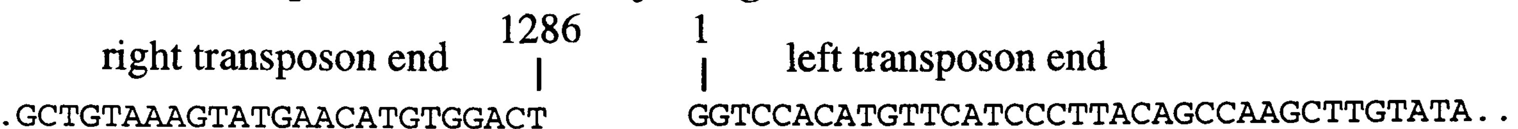
PCR products obtained using primers P3855 and P3856, were cloned into vector pGEM-T as described in 2.3.2.9, and blue/white selection used to detect recombinant colonies.

Primer P4795, which anneals approximately 50 bases from the left end of *mos*1 was then employed to determine the exact circularisation junction in 39 clones (table 5.1). Sequencing reactions were performed both manually, as described in 2.3.2.18 using a Sequenase version 2.0 USB kit and α^{35}S-dATP, and automatically, as in 2.3.2.19 using an ABI PRISM™ 377 DNA sequencer. Most clones were found to have lost terminal *mariner* sequences. The most complete copies had three nucleotides missing from either the left or right hand end. Twenty one clones were obtained which contained one complete end, but with various deletions from the opposite end, other clones lacked nucleotides from both ends of the element.

Four clones contained intact left ends of *mariner* plus extensions consisting of the TA dinucleotide which flanks all *mariner* insertions and additional non-*mariner* sequences. Primer V8523 which anneals roughly 40 bases from the right end of *mos*1 was used in sequencing reactions to determine the structure of the right hand termini of the element in these clones. Two of the clones had both ends intact with approximately 500bp of additional sequence between them. It is possible that these two clones are actually the same clone since multiple copies can be recovered from cloning PCR amplified fragments. The two remaining clones had lost three base pairs from the right hand end. These clones may also be a copy of each other.

The deleted and extended circular copies of *mariner* detected are presumably responsible for the PCR products obtained of higher and lower molecular weight than anticipated with strain 238 template DNA in figure 5.4.

The data indicate that circular, extrachromosomal *mariner* molecules are not homogeneous in structure, and usually contain nucleotides deleted from one or both ligated transposon ends. Since most of the circular copies of *mariner* isolated are missing at least 3 nucleotides from one end of the inverted repeat on both strands, it is unlikely that they are viable transposition intermediates. It is possible that extrachromosomal molecules of this structure have accumulated because of their inability to re integrate into the genome.



Conceptual head to head joining of *mariner* ends

right transposon end 1286 | 1 | left transposon end

.GCTGTAAAGTATGAACATGTGGACT GGTCCACATGTTCATCCCTTACAGCCAAGCTTGTATA. .

No. of clones	Molecules with one complete end	
13	GCTGTAAAGTATGAACATGTGG	GGTCCACATGTTCATCCCTTACAGCCAAGCTTGTATA
2	GCTGTAAAGTATGAACATGT	GGTCCACATGTTCATCCCTTACAGCCAAGCTTGTATA
1	GCTGTAAAGTATGAACA	GGTCCACATGTTCATCCCTTACAGCCAAGCTTGTATA
2	GCTGTAAAGTATGAACATGTGGACT	ACATGTTCATCCCTTACAGCCAAGCTTGTATA
1	GCTGTAAAGTATGAACATGTGGACT	TGTTCATCCCTTACAGCCAAGCTTGTATA
1	GCTGTAAAGTATGAACATGTGGACT	TTGTATA
1	GCTGTAAAGTATGAACATGTGGACT	TA
2	GCTGTAAAGTATGAACATGTGGACTTA......	TAGGTCCACATGTTCATCCCTTACAGCCAAGCTTGTATA
2	GCTGTAAAGTATGAACATGTGG.............	TAGGTCCACATGTTCATCCCTTACAGCCAAGCTTGTATA

No. of clones	Molecules without a complete end	
5	GCTGTAAAGTATGAACATGTGG	CCACATGTTCATCCCTTACAGCCAAGCTTGTATA
2	GCTGTAAAGTATGAACATGTGG	TCCACATGTTCATCCCTTACAGCCAAGCTTGTATA
1	GCTGTAAAGTATGAACATGTGG	CATGTTCATCCCTTACAGCCAAGCTTGTATA
1	GCTGTAAAGTATGAA	CAAGCTTGTATA
1	GCTGTAAAGTAT	CTTGTATA
1	GCTGTAA	A
1	deleted to 1203	CCACATGTTCATCCCTTACAGCCAAGCTTGTATA
1	deleted to 1182	CCACATGTTCATCCCTTACAGCCAAGCTTGTATA
1	GCTGTAAAGTATGAACATGTGG	TCCACATGTT at 35 seq starts 1...

Table 5.1 Sequences of *mariner* junctions in PCR clones. Dots between abutted inverted repeats represent several hundred bases of non-*mariner* sequence.

Molecules with two intact *mariner* ends might serve as competent transposition intermediates. Two of the 39 clones contained a complete *mariner* element. These copies are separated by several hundred bases of *Drosophila* DNA. Active *mariner* elements in general appear to be excised almost precisely (Coates *et al*, 1995) and new insertions result in duplication of a TA residue only (Bryan *et al*, 1990). The two clones containing intact elements are therefore also unlikely to be active intermediates in *mariner* transposition.

5.2.2 Analysis of *mos1* elements excised *in vitro*

5.2.2.1 Excised copies of *mos1* obtained *in vitro* are linear

A linear form of *mariner* may be the active intermediate in transposition. It is possible that at least one of the *mos1* hybridising bands observed in figure 5.2 contains linear DNA. However, since centrifugation proved unsuccessful in separating extrachromosomal forms from genomic DNA, further study of this possibility was impeded. A different approach utilising a functional *in vitro* *mos1* excision assay system was instead followed.

Addition of recombinant Mos1 transposase to plasmid p*mos* produces *mos1* hybridising bands which correlate in mobility to full length, linear *mos1* elements, as described in 4.2.1 and figure 4.1. To determine if these excised elements are solely linear in conformation, a PCR assay was performed similar to that employed in 5.2.1.3, for the detection of circular, extrachromosomal forms of *mariner.*

In vitro excision assays were performed, as described in 4.2.1. Assays were carried out in the presence of either 5mM MnAcetate or 5mM $MgCl_2$. The reaction products were then ethanol precipitated and resuspended in 20µl of dH_2O before using 1µl in a PCR reaction with primers P3855 and P3856. Positive control reactions for the primers and the DNA were also performed.

The products from the PCR reaction were fractionated on a 1% agarose gel and the DNA visualised under ultraviolet light (figure 5.10). From figure 5.10 lanes 2 to 4 it is apparent that no amplification from the excision assay takes place when primers designed to detect circular copies of *mos1* are employed. The control reactions for the DNA, lanes 6 to 8 demonstrate that the DNA from the excision assay is suitable for PCR and the control reaction in lane 1 demonstrates that primers P3855 and P3856 are functional. *In vitro*, all excision products therefore appear to be linear in conformation.

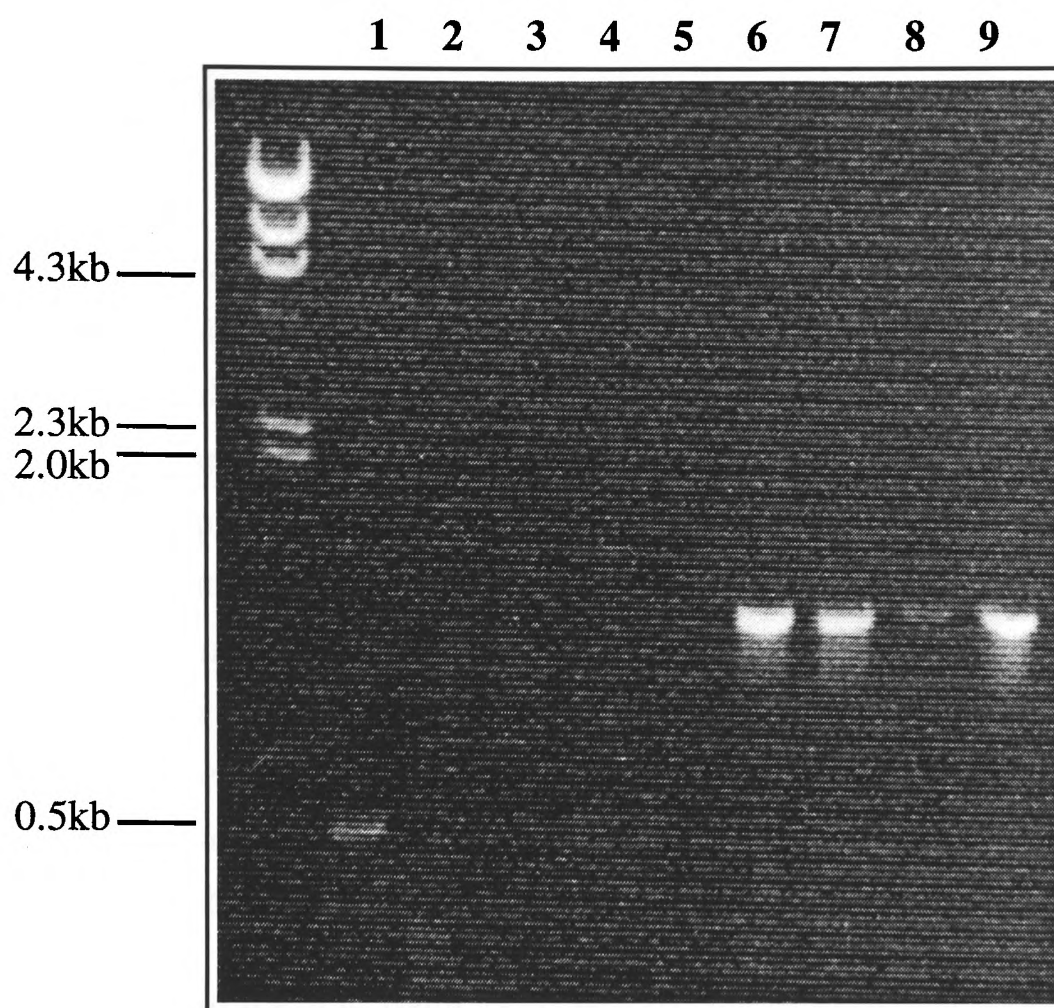


Figure 5.10. Only linear forms of *mos*1 exist following excision *in vitro*, as detected by PCR.

In lanes 1 to 5, primers P3855 and P3856 were used to detect circular copies of *mos*1. Lanes 6 to 9, employed primers 337 and 338 to amplify both linear and circular copies of *mos*1.

In lane 1 the PCR template is a clone from section 5.2.1.5 which is a positive control for primers P3855 and P3856. In lanes 2 + 6 the template is from an excision assay using Mn^{2+} as the cofactor. In lanes 3 + 7 the template is from an excision assay using Mg^{2+}. In lanes 4 + 8 the template is from a different excision assay, also using Mg^{2+}. Lane 5 contains no template and lane 9 uses p*mos*.

This contrasts with the situation *in vivo*, in which some or all of the excision products detected are circular in structure.

5.2.2.2 Mapping *in vitro mos*1 cleavage sites at the nucleotide level

To determine the positions of the cleavage sites following excision of *mos*1 *in vitro*, the structures of the termini of the excised *mos*1 elements were analysed by primer extension.

The DNA from *mos*1 excision assays dissolved in 20µl of dH_2O were used in primer extension reactions. Fifty pmol of suitable oligonucleotides were 5' end-labelled by incubation with polynucleotide kinase and α^{32}P-dATP, followed by purification on Sephadex G25 spun columns (see 2.3.2.20). Approximately 0.5pmol of each labelled oligo was then employed in 30 linear amplification cycles of 94°C, 1 minute; 50°C, 1 minute; 72°C, 1 minute, with 5µl of dissolved excision assay DNA as template. Manual sequencing reactions using each oligo and template p*mos* were also performed. Each primer extension reaction was then separated on a 6% acrylamide gel with sequencing reactions from the same oligo run in adjacent lanes. Running the primer extension and sequencing reactions in tandem, allows the last base of the excised element to be determined by direct reading from the autoradiograph.

Taq DNA polymerase has been reported to add a single template independent nucleotide at the end of sequences it has extended (Clark *et al*, 1988). Gibco Taq polymerase was tested to confirm this. Oligonucleotide V5054, which is homologous to sequences near the right end of *mos*1 was employed in a control primer extension reaction with *Rsa*I digested p*mos* template (figure 5.11). *Rsa*I restricts the DNA sequence CATG, between the A and the T residues. If primer extensions using Taq DNA polymerase did not extend the sequence an extra nucleotide, the *Rsa*I primer extension band in figure 5.11 would comigrate with a T residue. As can be observed though, it migrates with a sequencing band, which terminates in an A residue and is one nucleotide larger in size than the T.

V5054 was then used in a primer extension reaction to determine the terminal nucleotides on the 5' strand at the right hand end of excised *mos*1 elements. Figure 5.12 shows the results of primer extension from two *in vitro* excision assays, run alongside sequencing reactions from the same primer. The excision reactions differ in the divalent cation, Mn^{2+} or Mg^{2+}, that was present. The major primer extended bands comigrate with a sequencing band that terminates in a T residue. Bearing in mind that Taq DNA polymerase extends by an extra nucleotide, it is apparent that

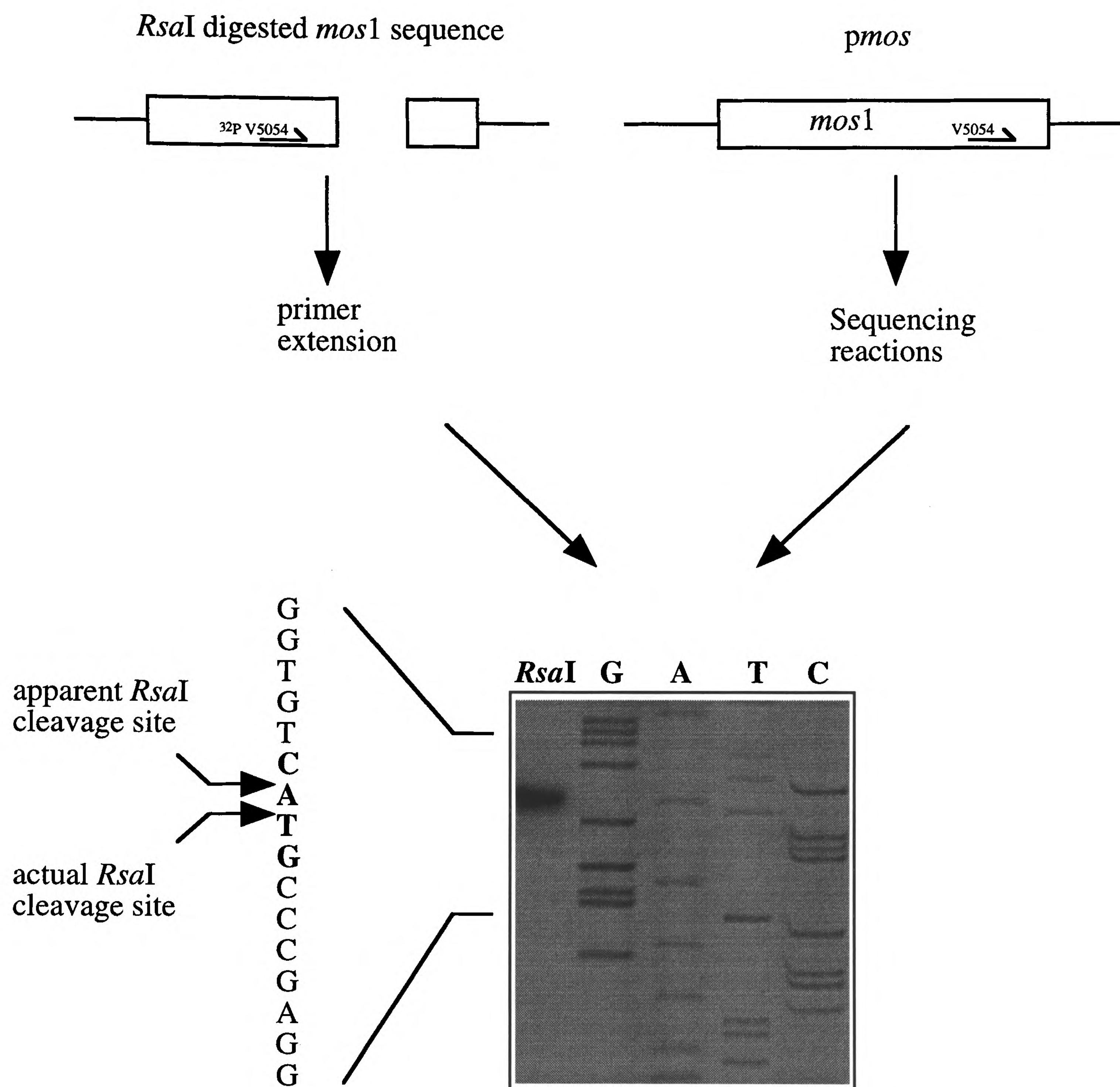


Figure 5.11. Taq DNA polymerase adds one base after copying template.

End labelled primer V5054 was employed to demonstrate the addition of 1 extra base by Taq DNA polymerase (from Clark, 1988). The *Rsa*I site is shown in bold type, and the site of cleavage marked. The primer extension band is co-migratory with the A residue from the *Rsa*I site, this band is 1 nucleotide larger than would be expected if Taq did not extend by an extra base.

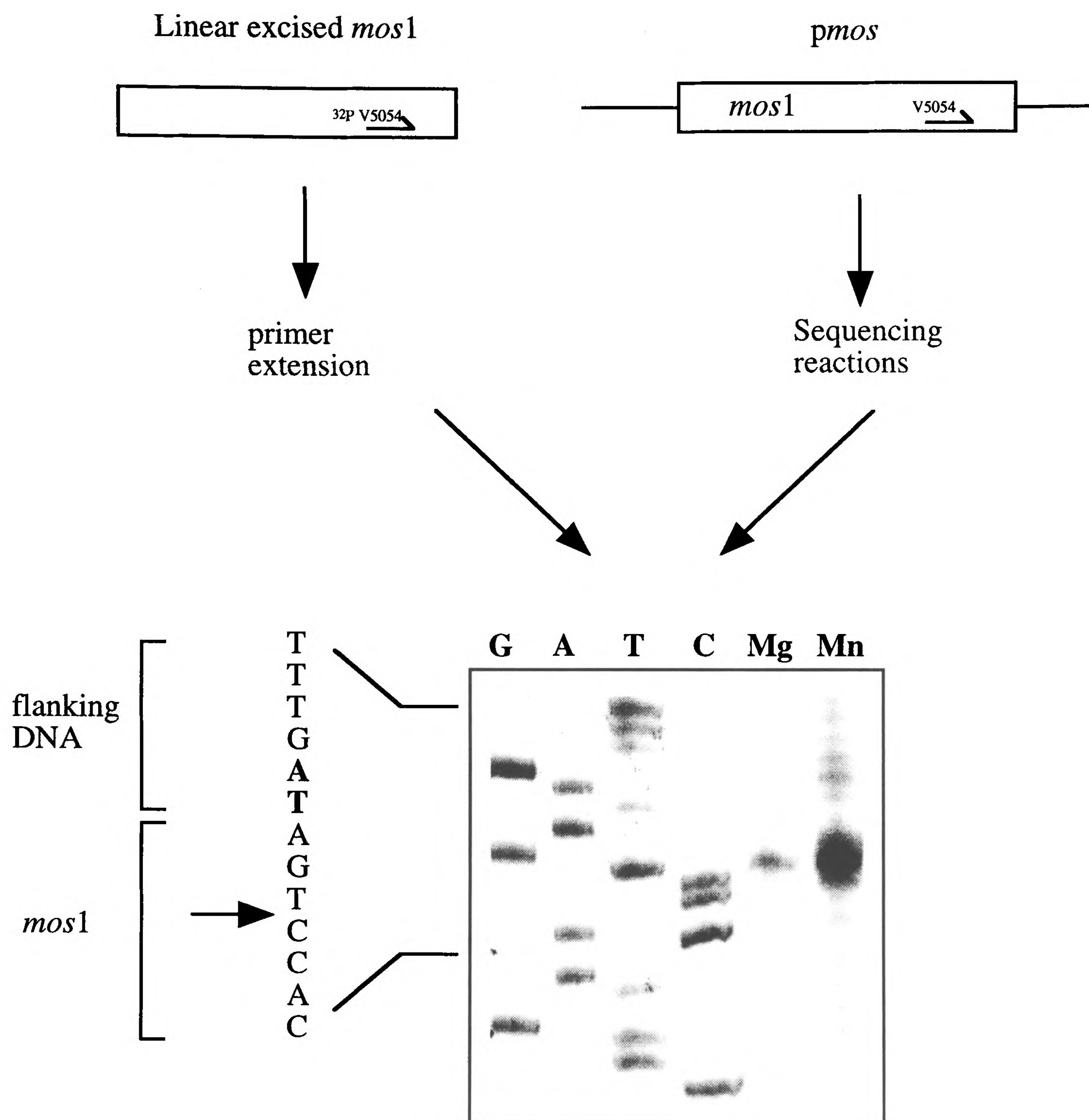


Figure 5.12. Determining the terminal most nucleotides on the 5' strand at the right end of excised *mos*1 elements by primer extension from oligo V5054.

The primer extension template is excision reactions peformed in the presence of either Mn^{2+} or Mg^{2+}. The vertical DNA sequence corresponds to the junction between the right hand end of *mos*1 and its flanking DNA in p*mos*. The TA target site duplication is shown in bold and the arrow marks the position at which cleavage has taken place. See text for detailed explanation.

cleavage of *mos*1 from plasmid p*mos*, occurs leaving three bases of *mos*1 sequence behind from the 5' strand at the right end of the element. Other minor points of cleavage at the right hand end of the element were observed in the presence of Mn^{2+} ions.

Oligo V6842 hybridises close to the left end of *mos*1 and was used to determine the terminal sequences of excised *mos*1 on the 5' strand at this end. The primer extension results in figure 5.13 demonstrate that three bases of *mos*1 sequence on the 5' strand at the left end of excised *mos*1 are also missing.

From figures 5.12 and 5.13, extension analysis has revealed that the three terminal most nucleotides of excised *mos*1 sequences are missing from each 5' strand. This result was obtained with either Mn^{2+} or Mg^{2+} in the excision reaction. Minor cleavage at other sites was observed in the presence of Mn^{2+}.

Due to the directionality of DNA synthesis, analysis of the sequences at the 3' ends of excised *mos*1 elements by primer extension could only be carried out by extending from sequences within plasmid p*mos*. By determining the last nucleotide on the flanking DNA from which *mos*1 excised, the sequence present on the 3' strand of excised *mos*1 elements can be deduced.

Oligo V8523, hybridises to a sequence in p*mos* flanking the left end of *mos*1 (see figure 5.14). This oligo was used in primer extension reactions to analyse the sequences remaining behind in p*mos* following excision of *mos*1. From this, the terminal most nucleotides of the 3' strand at the left end of excised *mos*1 were deduced, as shown in figure 5.14. In the presence of Mg^{2+} in the excision reaction, two cleavage products are obtained of equal intensity. The first product indicates that the terminal most nucleotide of *mos*1 remains behind in the plasmid on excision, the second that two nucleotides from *mos*1 remain behind. In the presence of Mn^{2+}, two major products of equal intensity are also observed and again one of the products demonstrates that one nucleotide, and the second that two, are missing from the 3' strand of *mos*1 at the left hand end of the element on excision. Minor primer extension bands indicating excision at nucleotides either side of the two main cleavage points were also detected when Mn^{2+} was used in the excision reaction.

Oligo V8524, which is homologous to sequences in p*mos* flanking the right hand end of *mos*1, was used to determine the structure on the 3' strand at the right end of *mos*1 elements following excision. From figure 5.15, in the presence of Mg^{2+}, three cleavage points were detected. These are such that two, three and five nucleotides from

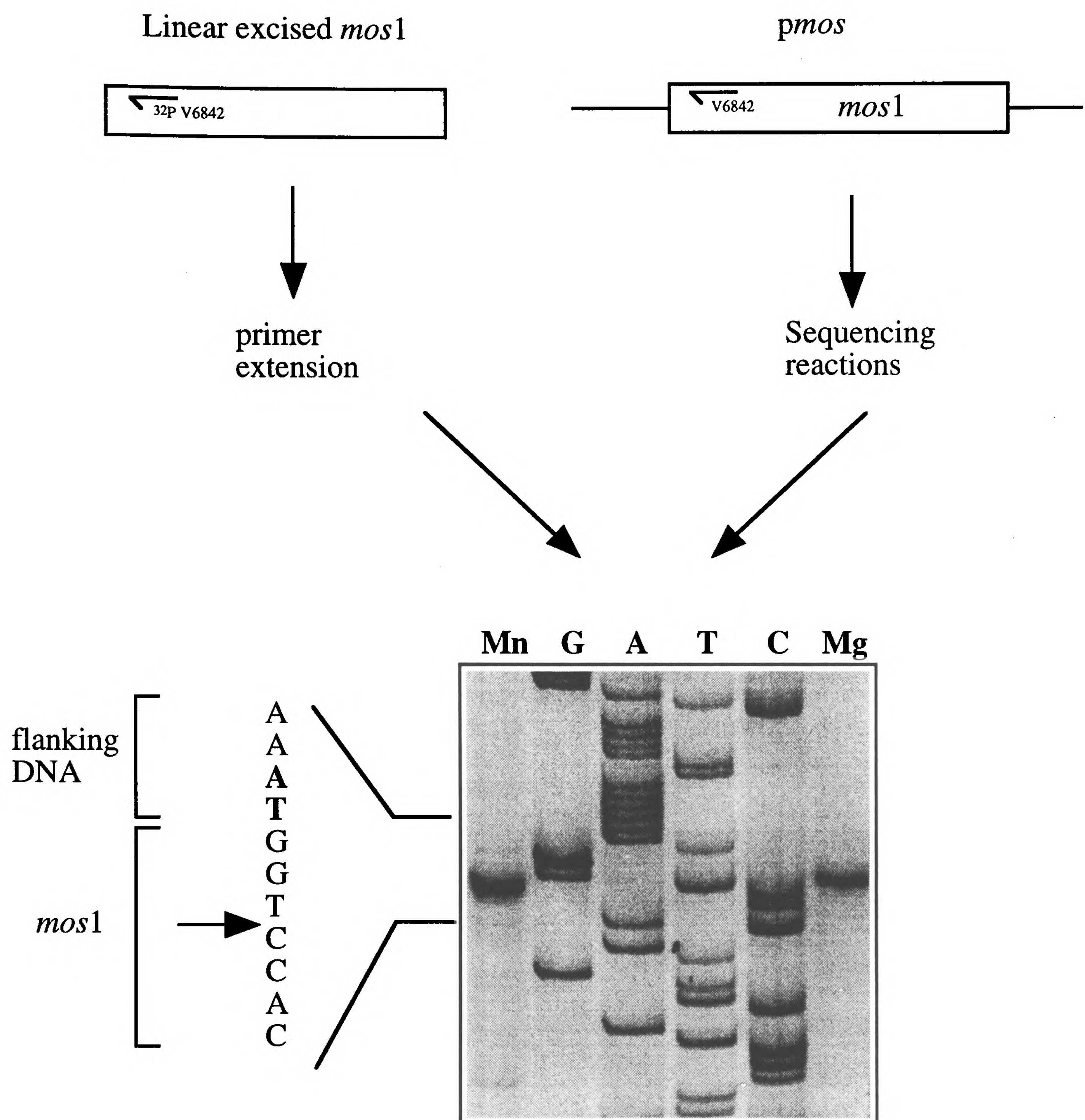


Figure 5.13. Determining the terminal most nucleotides on the 5' strand at the left end of excised *mos*1 elements by primer extension from oligo V6842.

The primer extension template is excision reactions peformed in the presence of either Mn^{2+} or Mg^{2+}. The vertical DNA sequence corresponds to the junction between the left hand end of *mos*1 and its flanking DNA in p*mos*. The TA target site duplication is shown in bold. The arrow marks the position at which cleavage has taken place, bearing in mind that Taq adds on one extra nucleotide. See text for detailed explanation.

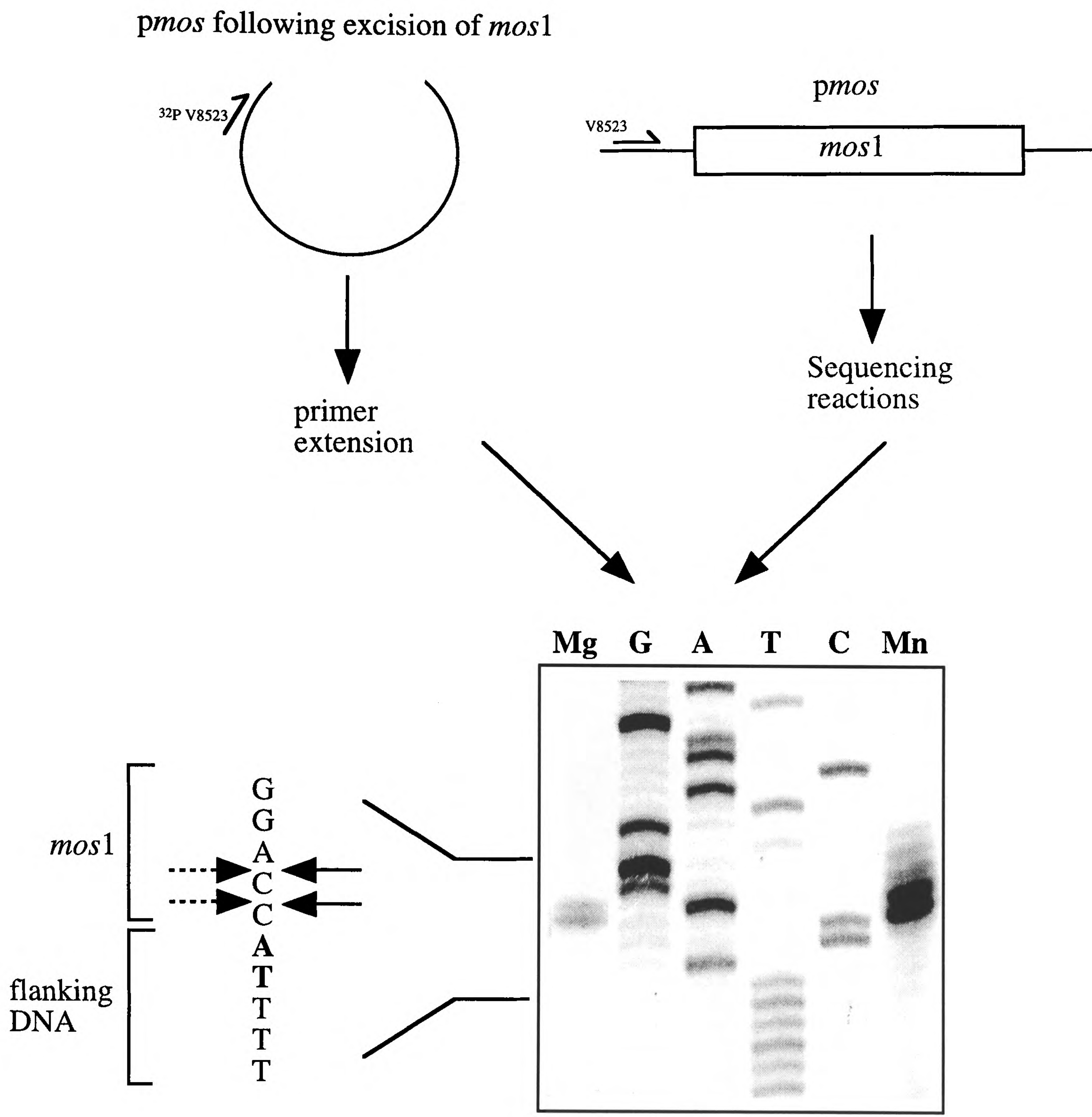


Figure 5.14. Determining the terminal most nucleotides on the 3' strand at the left end of excised *mos*1 elements by primer extension from oligo V8523, which hybridises outside the *mos*1 sequence.

The primer extension template is excision reactions performed in the presence of either Mn^{2+} or Mg^{2+}. The vertical DNA sequence corresponds to the junction between the left hand end of *mos*1 and its flanking DNA in p*mos*. The TA target site duplication is shown in bold. Sequences corresponding to *mos*1 and plasmid p*mos* are individually bracketed. Full arrows mark the positions at which cleavage has taken place when Mn^{2+} is present in the excision reaction and dashed arrows, when Mg^{2+} is used, bearing in mind that Taq adds on one extra nucleotide. See text for detailed explanation.

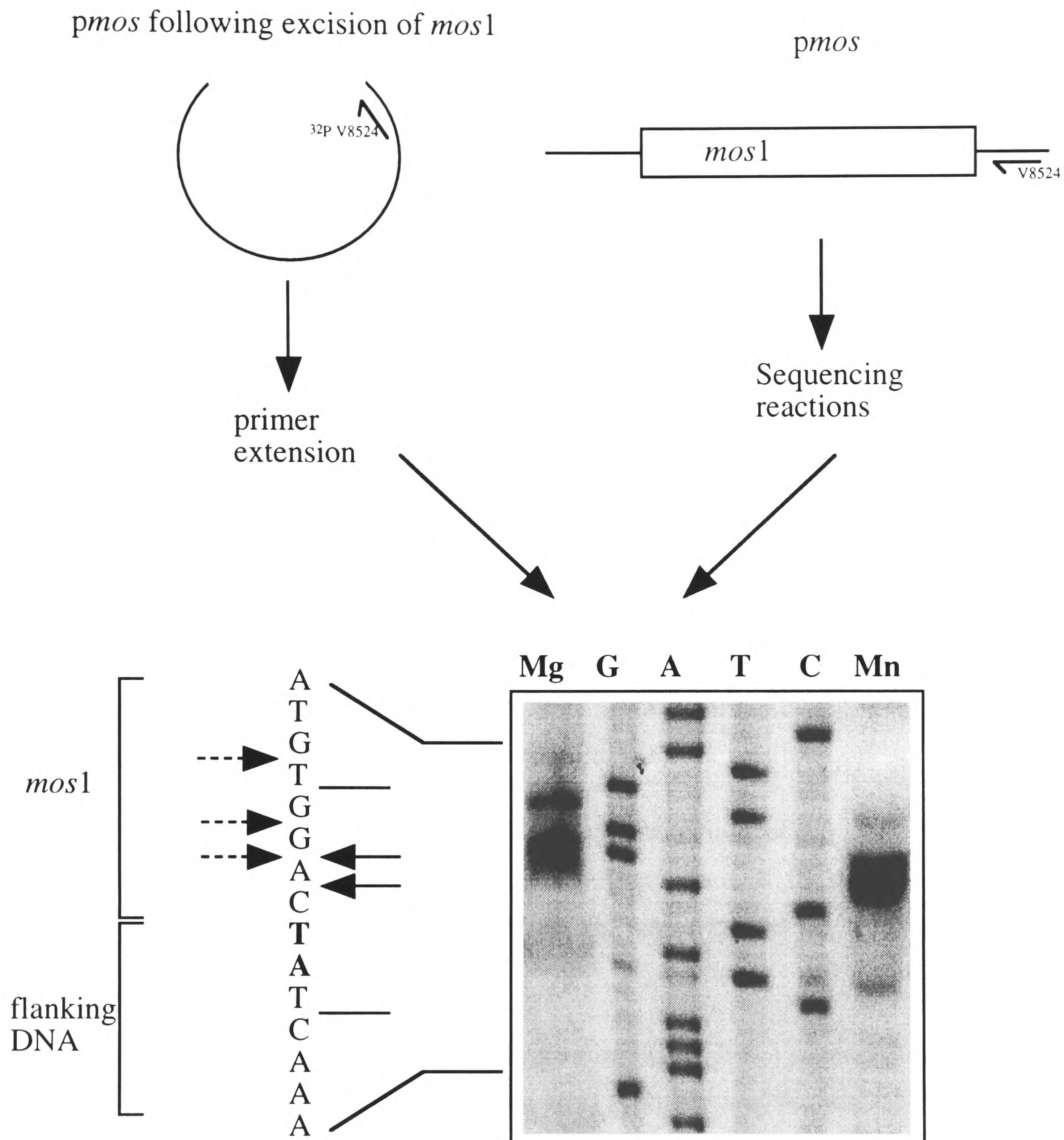


Figure 5.15. Determining the terminal most nucleotides on the 3' strand at the right end of excised *mos*1 elements by primer extension from oligo V8524, which hybridises outside the *mos*1 sequence.

The primer extension template is excision reactions peformed in the presence of either Mn^{2+} or Mg^{2+}. The vertical DNA sequence corresponds to the junction between the right hand end of *mos*1 and its flanking DNA in p*mos*. The TA target site duplication is shown in bold. Full arrows mark the positions at which cleavage has taken place when Mn^{2+} is present in the excision reaction and dashed arrows, when Mg^{2+} is used. The positions of minor cleavage sites are marked with lines. See text for detailed explanation.

the 3' strand of *mos*1 remain behind in the plasmid following excision. With Mn^{2+}, two major products, indicating that one or two nucleotides of *mos*1 are not excised from the plasmid. Two minor products were also observed which show that four nucleotides of *mos*1 remain in the plasmid, or that three bases of flanking DNA are also excised.

5.3 Discussion

5.3.1 Circular forms of *mariner*

The circular, extrachromosomal *mariner* molecules isolated from *D. simulans* are heterogeneous in structure, usually with nucleotides deleted from one or both ligated transposon ends. Such molecules would not be expected to successfully re integrate.

There are at least two possible mechanisms for the way in which these copies of *mariner* may have arisen. One explanation is that the elements have excised from the genome inaccurately before ligation of their ends. A second possibility is that the elements have excised incompletely, but in an identical manner and that further deleted circular copies are a result of exonuclease activity before ligation.

Class II transposable elements frequently leave 'footprints' at the site of their excision from the genome (e.g. Engels, 1989). The sequences of these footprints usually correspond to a duplicated target insertion site and parts of the transposon inverted terminal repeats.

The most common footprints following excision of the *Tc*1 and *Tc*3 elements in somatic cells of *C. elegans* consist of two base pairs from either end of the transposon, flanked by the TA duplication (Ruan and Emmons, 1987; Eide and Anderson, 1988; van Leunan *et al*, 1994). Footprints similar to this have been found upon loss of the related transposon *minos* from the genome of *D. hydei.* Most contain a six base pair sequence consisting of the four terminal nucleotides of either end of the transposon flanked by the TA target site duplication (Arca *et al*, 1997).

The two predominant footprints resulting from *mos*1 induced excision of *mariner* in *D. simulans* contain a TA dinucleotide followed by three nucleotides from the left or right end of *mariner* and a second TA dinucleotide (Bryan *et al*, 1990). Similar footprints have also been obtained using a plasmid based excision assay with *mos*1 in a variety of fly embryos (Coates *et al*, 1995), and following excision of the *mariner*-like *impala*

element from the chromosomes of the fungus *Fusarium oxysporum* (Langin *et al*, 1995).

The footprints remaining following excision are believed to result from a 'ligation-repair' mechanism originally proposed for *Tc*1 (Eide and Anderson, 1988). According to this model, for *mariner* elements, a three base pair staggered cut is introduced at each end of the transposon. This leaves three nucleotides of the inverted repeat at each end of the chromosome break and the complementary single-stranded tails at the ends of the transposon. Ligation of the chromosome ends to each other then produces a heteroduplex with a mismatch which, by repair or replication, gives rise to the two alternative footprints observed.

Recently heteroduplexes have been identified following transposase induced excision of *mos*1 from plasmids in insect embryos. Transformation of plasmid containing a heteroduplex at the *mos*1 excision breakpoint into *E.coli* results in two plasmids within the same *E. coli* cell, with different breakpoint sequences but which originate from one excision event (Coates *et al*, 1995). Finding the heteroduplex structures predicted by the model following excision, strongly suggests that *mos*1 transposase introduces staggered breaks at the transposon ends, leaving three base pairs protruding (figure 5.16).

Excision of *mariner* elements from the genome by a three base pair staggered cut at each transposon end, favours the second possibility for the way in which circular copies of the element have arisen. If *mariner* elements are excised linearly from the genome to give three base pair single strand ends, it is possible that heteroduplex formation may take place between the protruding single strand ends. This is analogous to heteroduplex formation at the site of a footprint, between the ends of the transposon that are not coexcised with the rest of the element. Mismatch repair of such a circularised element containing a hetereoduplex necessarily results in the deletion of three nucleotides from one end of the element. This would explaining why a number of the molecules detected in 5.2.1.5 are missing three bases from either one end of the element or the other. The end of the element that is complete presumably depends on which strand of the heteroduplex is used as a template during mismatch repair. The circular copies which contain further deletions are presumably due to exonucleolytic degradation before ligation of the remaining ends of the element.

In light of the footprints observed on excision of *mos*1 *in vivo* and the structures of the extrachromosomal *mariner* circles observed in 5.2.1.5, the model shown in figure

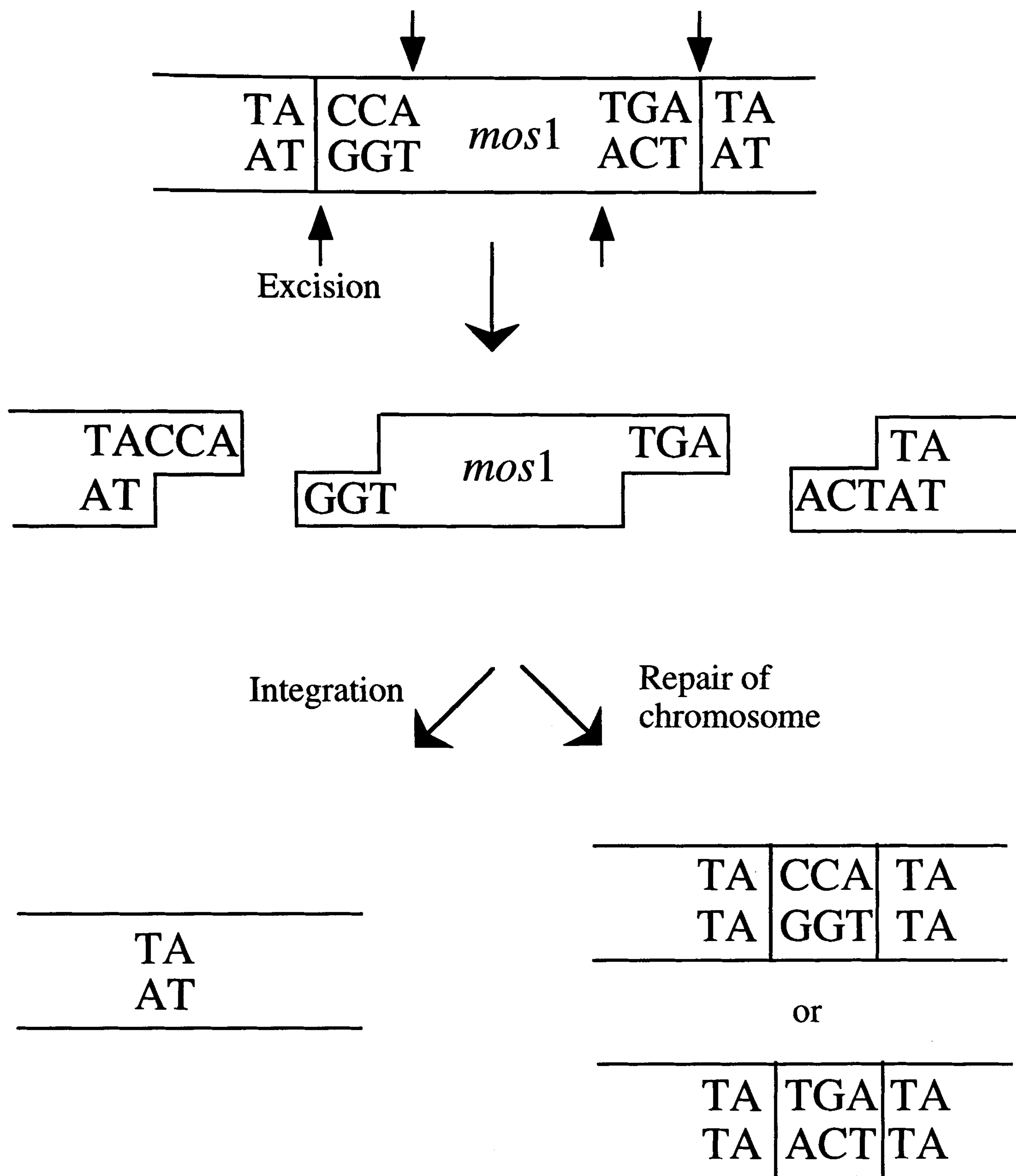


Figure 5.16. Proposed model for the transposition of *mos*1 and other *mariner* elements in *Drosophila* (adapted from van Luenen *et al*, 1994).

A schematic model for *mos*1 transposition in which double strand breaks with a three base pair stagger are generated, resulting in an excised element. Footprints may be generated at the excision site. The element integrates at a TA dinucleotide. Short arrows indicate the positions of the cleavage sites at the ends of the transposon. See text for further details.

5.16 can be proposed for transposition of *mos*1 and other *mariner* elements in *Drosophila*. The model is based on that originally proposed for transposition of the *Tc*3 element in *C. elegans* (van Leunan *et al*, 1994). The presence of 3' rather than 5' protruding ends suggests a simple mechanism for regeneration of the full sequence of the transposon after re-insertion and for generation of the TA target site duplication, as has previously been suggested for *Tc*3 (van Leunan *et al*, 1994).

The model in figure 5.16 proposes that double-strand breaks with a 3bp stagger are generated resulting in an excised element that contains the complete *mos*1 sequence at the 3'ends but lacks the terminal three nucleotides at the 5' ends. Integration of this element into a TA target sequence then takes place. The lesion in the chromosome can be repaired by pairing of the staggered ends and mismatch repair, leaving either the sequence of the first or last three nucleotides of the element. This model assumes a three base pair staggered intermediate and is therefore different to transposition of *Tc*3 and *Tc*1 and the *mariner* element *Himar*1, which move via a two base pair staggered intermediate (van Luenen *et al*, 1994; Vos *et al*, 1996; Lampe *et al*, 1996).

In addition to the *Tc1/mariner* family of elements, the presence of circular extrachromosomal copies have been described for transposon-like elements in protozoa, such as TBE1, Tec1 and Tec2 (Williams *et al*, 1993; Jarakzawsi and Jahn, 1993) and for bacterial transposons of the IS3 family, such as IS911 and IS1 (Polard *et al*, 1992; Turlan *et al*, 1995), all of which encode 'D35E' transposases. Transposable elements unrelated to this superfamily, such as the maize transposon Mu and the prokaryotic element Tn916, also generate circular forms (Sundarasan, 1987, Caparon, 1989). For most of these transposons, circular forms are not thought to be transposition intermediates, and only the covalently closed circular forms of the conjugative transposon Tn916 and the IS3 family member IS911 have been shown to act as intermediates in transposition (Caparon and Scott, 1989; Ton-Hoang *et al*, 1997).

5.3.2 Linear forms of *mos*1 obtained *in vitro*

The model in figure 5.16 proposes that the intermediate in *mariner* transposition is a linear copy of the element containing a three base pair staggered cut at each end. Primer extension analysis of *in vitro* excised *mos*1 elements was performed to determine if these structures could be detected. The 5' ends of excised, linear elements were looked at initially. The primer extension results in figures 5.12 and 5.13 demonstrate that in the presence of either Mn^{2+} or Mg^{2+} in the assay buffer an

excised element is formed which is deleted for 3bp on the 5' strand at each end. This is consistent with the model in figure 5.16.

However, on looking at the structure at each 3' end, the predicted point of cleavage, after the most terminal *mos*1 nucleotide was not observed. Several other points of cleavage were identified, these were different depending on the metal ion present during excision. Most of the cleavage events result in sequences from the 3' strands of *mos*1 at each end of the element remaining behind in the plasmid. A single, minor cleavage site was observed in the presence of Mn^{2+} in which the whole *mos*1 sequence on the 3' strand at the right hand end of the element was excised, however, 3bp of flanking DNA were also removed.

It is unlikely that any of the excised bands detected in the *in vitro* excision assays corresponds to an active *mos*1 transpositional intermediate, since the majority of copies are missing some nucleotides on both strands at each end of the element.

The structures of linear copies of *Tc*3 and *Tc*1 have been examined. The appearance of excised, linear *Tc*3 molecules is observed on induction of *Tc*3 transposition in *C. elegans* (van Luenen *et al*, 1993). The structures of the termini of the excised elements were analysed by restriction and co-fractionation on a polyacrylamide gel with synthetic oligonucleotides of known size. The 3' ends of excised elements were also determined in a different assay. Multiple dCTP residues were added to the 3' end of purified transposon by terminal transferase. The C-tailed extrachromosomal material was then used in a PCR reaction with a G-tailed and a *Tc*3 specific primer and the PCR product sequenced. These results demonstrated that *Tc*3 elements are excised after the last *Tc*3 nucleotide on each 3' strand, and two nucleotides into the element on each 5' strand (Van Luenen *et al*, 1994).

*Tc*1 elements can be mobilised *in vitro* on addition of nematode nuclear extract containing elevated levels of transposase to plasmids containing *Tc*1. The positions of the double strand cleavages of excised copies of *Tc*1 were determined using primer extension analysis. The 5' cut was found to be 2bp within the transposon, and the 3' cut to be at the transposon end (Vos *et al*, 1996). These linear, 2bp staggered forms of the elements are believed to be intermediates in transposition. They are consistent with the break repair of the genome which results in the formation of element specific footprints, following excision (Plasterk, 1991; van Luenen *et al*, 1994).

The nucleotide positions following cleavage of the inverted terminal repeats of the *mariner* element *Himar*1 have recently been determined. Incubation of radiolabelled

*Himar*1 sequence with recombinant transposase was found to result in a 2bp staggered cleavage pattern in which cutting after the most terminal transposon nucleotide on the 3' strand and 3 bases in on the 5' strand had occurred (Lampe *et al*, 1996). This suggests that cleavage of element termini by *Himar*1 transposase more closely resembles Tc1 and Tc3 transposases than Mos1 *mariner*.

The transposition intermediates of all transposable elements analysed to date contain the last element specific nucleotide on the 3' strand. The 5' end of the transposable element does not seem to play an important role in the transposition and is found at different positions in various elements as illustrated in figure 5.17.

It is possible that *mos*1 and other *mariner* elements in *Drosophila* do transpose via the predicted 3bp staggered form, but that elements of this structure are either not excised or not detected during cleavage reactions *in vitro*. It may be that the reaction conditions are unsuitable for correct cutting of the element since the assay was optimised for maximum excision, rather than transposition, in which, presumably, correct intermediates would be formed. This may have encouraged the excision of aberrant elements.

Cleavage with Mn^{2+} is generally more pronounced than with Mg^{2+} and results in additional minor cleavage products as observed in figures 5.12 and 5.14. Higher levels of cleavage in the presence of Mn^{2+} have previously been reported in *in vitro* assays examining the cleavage of *Tc*1 and *Himar*1 elements. The presence of extra cleavage products in the presence of Mn^{2+} has also been previously documented. (Vos *et al*, 1994; Lampe *et al*, 1996).

Although the excised *mos*1 elements are not intermediates in the transposition reaction, the primer extension results may perhaps indicate a specific order of cleavages. The 5' end single strand cleavage event is correct at both ends of the element. This may indicate that it occurs prior to 3' strand cutting. This would be the opposite of what has been reported for Tn10, in which the transferred strand is cleaved first (Bolland and Kleckner, 1995). It has also been suggested for *Tc*1, that 5' strand cleavage may occur first. On mutation of the terminal two nucleotides of the *Tc*1 inverted repeat, only 5' end single strand cleavage is observed (Vos *et al*, 1996). If the 5' and 3' cleavage reactions at each inverted repeat take place in an orderly manner and 3' cutting ability is removed (by mutation of the terminal two residues) then 5' cutting would not be expected. However, if 5' cutting is first, then only 3' cleavage will be prevented by mutation.

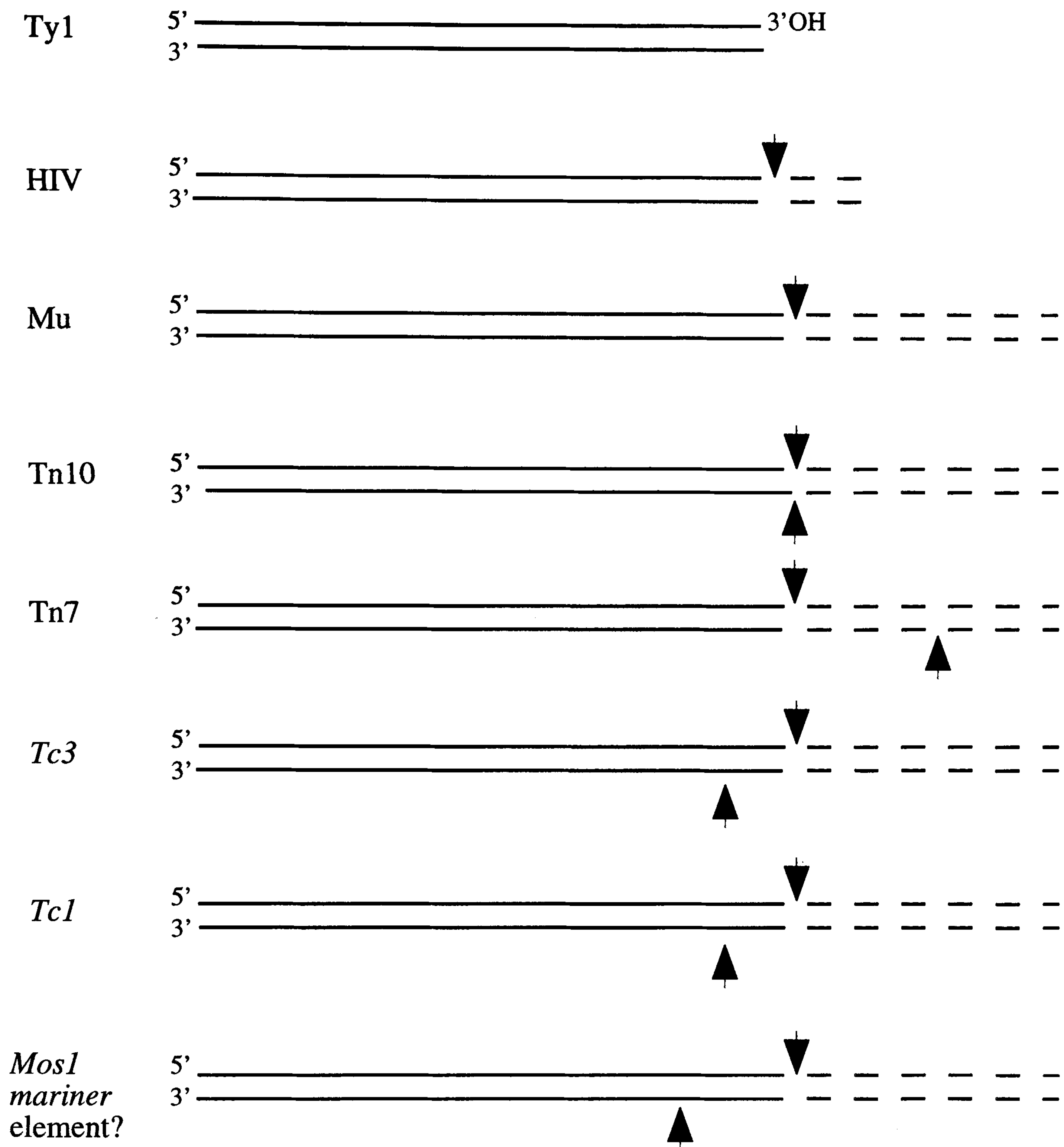


Figure 5.17. A comparison of cleavage sites between different transposable elements.

The right end of the indicated transposable elements are depicted together with the flanking sequence (dashed lines). The positions of the cleavage sites are indicated by arrows (adapted from van Luenen *et al*, 1994).

Cleavage of the non-transferred strand, i.e. the 5' strand, before the transferred strand would preclude the possibility of *mariner* or *Tc*1 promoted co-integrate formation as an intrinsic consequence of the reaction order. This is because the 3' termini required for strand transfer would have been generated in the same step that cleaved the second strand. During transposition of the 'cut and paste' element Tn10, the observed order of strand cleavage is compatible with a replicative mode of transposition. It has been suggested that the constraint on a replicative mode of transposition is imposed by a conformational block within the synaptic complex, which prevents stable interaction with the target DNA until after the double-strand cleavage has occurred at both ends (Bolland and Kleckner, 1995).

Chapter 6

Recombinant Mos1 transposase is active *in vivo*

6.1 Introduction

A major goal of current molecular entomology is the development of a broad range transformation system similar to that of the *P* element used extensively in *Drosophila melanogaster.* The technology for creating transgenic insects has many possible applications. One of these, for example, is to manipulate breeding stocks of the silkworm; *Bombyx mori,* this is analogous to the uses of transgenic technology in plant and livestock breeding programmes. Other proposed applications, such as replacing populations of pest insects with genetically engineered non-pest strains are more novel (Collins, 1994; Curtis, 1994, Spielman, 1994).

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 a 'helper' plasmid that provides a source of *P* transposase but cannot be mobilised *in trans*. The other is the construct to be integrated that is mobilised with the aid of transposase supplied by the helper. The construct to be integrated can contain non-transposable element sequences as long as the terminal *P* element sequences which are essential for transposition are present. Chimeric transposable elements capable of transposing can therefore be created.

The host range of *P* elements, however, appears to be limited to *Drosophila* and numerous attempts to transform heterologous organisms outside the genus of *Drosophila* have been unsuccessful (Engels, 1989). The application of *P* elements to serve as germline transformation vectors is thus restricted. *Mariner* elements however, are apparently active in many genomes and can transpose without the aid of host factors (Lampe *et al*, 1996; Dawson, personal communication). This is encouraging for the development of *mariner* based vectors. Indeed, the *mariner* element *mos*1 has been shown to be capable of germline transformation in *D. melanogaster*, a species in which it is not normally found (Garza *et al*, 1991; Lidholm *et al*, 1993).

A difficulty that may arise in the transformation of insects less well-characterised than *Drosophila,* is in the construction of suitable helper plasmids that enable transcription of the transposase gene in the experimental organism. There are relatively few genes characterised in most insects that could provide such promoters necessary for helper plasmids.

An alternative to the use of helper plasmids is the direct injection of purified transposase together with the construct to be integrated. Such a system works for the *P* element in *D. melanogaster* (Kaufman and Rio, 1991). As the *mariner* transposases *Himar*1 and *mos*1 do not require a host encoded function (Lampe *et al*, 1996; Dawson, personal communication), they provide an opportunity for developing a germline transformation system based on the direct injection of *mariner* transposase protein into insect embryonic germ cells. This method eliminates the need to transiently express the transposase gene in divergent insect hosts, and may result in higher levels of transposase activity in injected embryos, resulting in increased integration rates.

To examine whether purified, recombinant Mos1 transposase is able to stimulate transposition of a *mos*1 element *in vivo*, a 'mini-*white*' marked, non-autonomous *mos*1 element was constructed and co-injected with transposase into *Drosophila melanogaster* embryos.

6.2 Results

6.2.1 *D. melanogaster* is the ideal organism for investigating the *in vivo* activity of recombinant Mos1 transposase.

Natural populations of *D. melanogaster* do not contain *mariner* elements (Jacobson *et al*, 1986). The possibility of mini-*white* marked *mos*1 elements being cross-mobilised by the activity of endogenous *mariner* elements is therefore precluded. This will allow transposition events, if observed, to be attributed directly to the activity of recombinant transposase.

Previous attempts to transform the germline of *D. melanogaster* using a DNA based *mos*1 helper-plasmid system, based on that for the *P* element, have been successful. The transposition of *mos*1 elements ranging in size from 1.3-13.2kb has been detected (Garza *et al*, 1991; Lidholm *et al*, 1993). This demonstrates that *mos*1 and *mos*1-based elements are capable of transposition within a *D. melanogaster* background. Combined with the availability of suitable markers and fly strains, *Drosophila melanogaster* was conclude to be the ideal organism to examine the activity of recombinant Mos1 protein *in vivo*.

6.2.2 Construction of p*mos-white*, a germline transformation vector

Vector p*mos-white* was created by inserting the mini-*white* gene from *D. melanogaster* into a unique site in the ORF sequence of *mos*1 (see figure 6.1).

Plasmid p*mos* contains a full-length, functional *mos*1 element (Medhora *et al*, 1991). Restriction of p*mos* with *Sal*I was initially performed to linearise the plasmid within the *mos*1 ORF sequence. The digested plasmid was then incubated with Klenow enzyme and dNTPs at 23°C for 1 hour, to fill in recessed 3' ends, before gel purification and quantification.

Digestion of pCaSper with *Eco*RI was carried out to excise a 4kb fragment containing the mini-*white* gene from *D. melanogaster*. Following separation of mini-*white* from the plasmid backbone by agarose gel fractionation, the DNA was gel purified and treated with Klenow enzyme (2.3.2.10) to create blunt ends. The DNA was then phenol-extracted, ethanol precipitated, resuspended in T.E. and quantified.

Blunt ended p*mos* and mini-*white* DNAs were ligated together in a 1:3 molar ratio, in the presence of T4 DNA ligase and 1μM hexamminecobalt chloride for 20 hours overnight at 16°C (2.3.2.11). Competent *E. coli* NM522 cells were then transformed and allowed to grow overnight on L-Amp plates. Colonies containing the mini-*white* insert were isolated in colony hybridisation screens and their identity confirmed by diagnostic restriction digestion. The construction of plasmid p*mos-white* is illustrated in figure 6.1. The *Sal*I site employed in the cloning strategy is situated at nucleotide 349 in the *mos*1 sequence. Insertion of mini-*white* at this site causes a 4kb disruption of the *mos*1 open reading frame presumably preventing translation of the full-length, functional Mos1 transposase, whilst maintaining the integrity of the ends of the element which are required for transposition. A midi-preparation of p*mos-white* DNA was then performed and the DNA dissolved in dH_2O before storage at -20°C prior to injection.

6.2.3 Recombinant transposase mediated transformation of *D. melanogaster*

Purified, recombinant transposase was initially diluted into buffer (25mM HEPES pH7.9, 100mM NaCl, 10% (v/v) glycerol, 2mm DTT) supplemented with 5mM MnAcetate, 25μg/ml BSA and a cocktail of protease inhibitors (see 2.3.5.2) to a concentration of approximately 0.23fmol/μl (230pM). The addition of 1μl of 1μg/μl

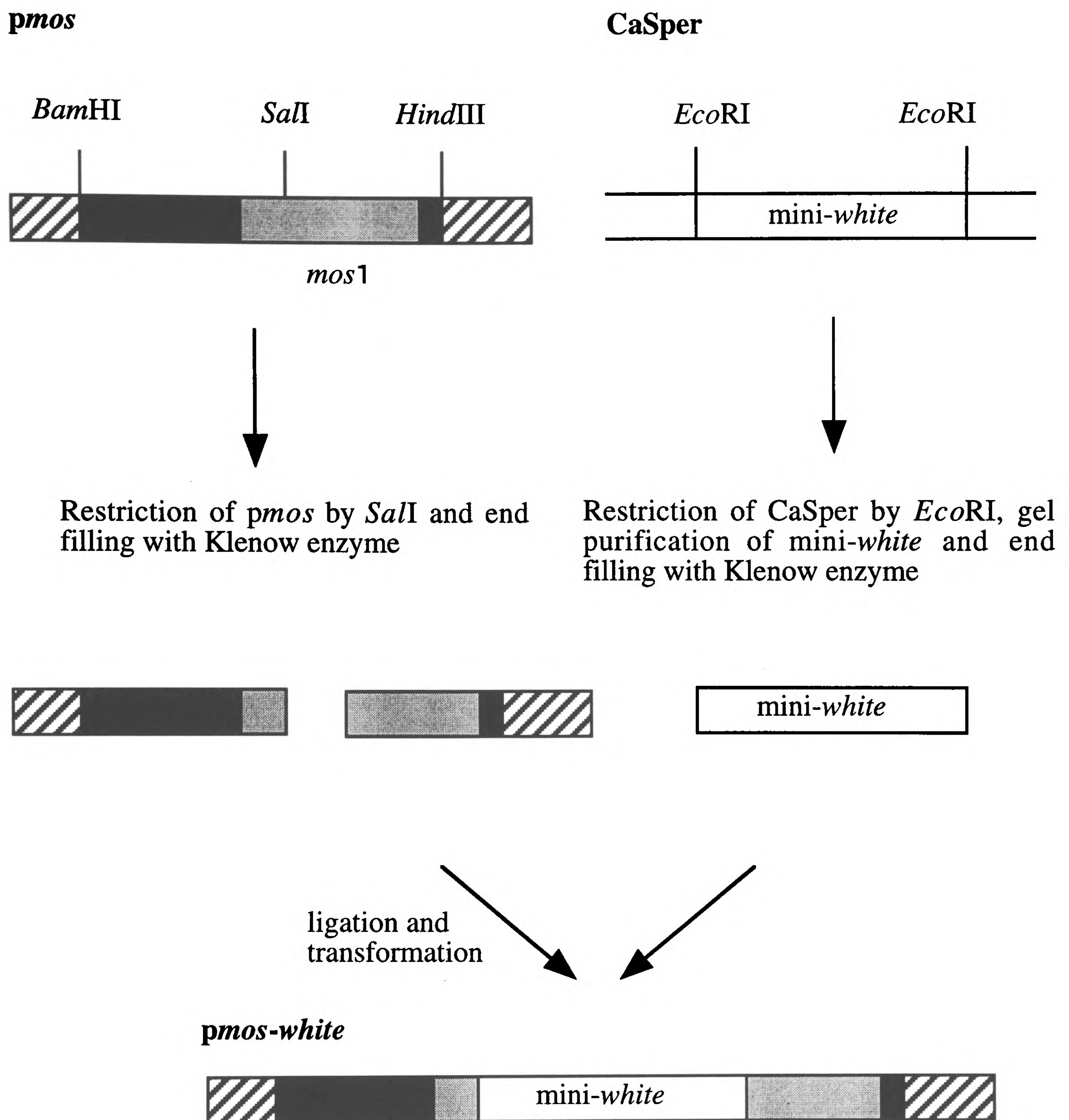


Figure 6.1. Construction of plasmid p*mos-white*

*Mos*1 sequence is pale grey, and *Drosophila* genomic DNA dark grey in shading, pBluescribe is represented by hatching.

p*mos-white* DNA in injection buffer (5mM KCl; 100μM $NaPO_4$, pH 6.8), to a 3μl transposase aliquot on ice was then carried out. The concentrations of DNA and transposase in this injection mixture are 0.25μg/μl (32nM) and 0.17fmol/μl (170pM) respectively, which gives a DNA:transposase molar ration of ~ 200:1. The transposase/DNA mixture was then used immediately to backload a microcapillary for embryo injection.

Embryos from *D. melanogaster* w^{1118} were collected as described in 2.3.4.2, before dechorionation in 50% bleach and alignment on a gridded filter. The embryos were then transferred to a glued cover slip and desiccated for 5-10 minutes (2.3.4.3). Embryos were covered with halocarbon oil, and the transposase/DNA mixture injected into the posterior pole of preblastoderm embryos. Embryos were allowed to hatch at 18°C, before transfer to French food at 25°C. G_0 flies were collected and mated individually to w^{1118} virgins. The resulting G_1 were assayed for a *white*$^+$ eye colour phenotype, which ranges from pale yellow to wild type brick red in a *white*$^-$ background (figure 6.2, table 6.1). Transformants were crossed to w^{1118} flies to produce males and females with the same insertion. Homozygous lines were then established by sibling matings.

A range of different transposase concentrations from 17-340pM were used to micro-inject embryos. A mock purification sample, consisting of identical fractions prepared from *E. coli* cells, was also injected together with DNA. The mock purification extract serves as a control for the possibility that an *E. coli* protein, co-purifying with Mos1 can provide transposase function. Embryos injected with mock transformed extract and p*mos-white* yielded no transformants, ruling out this possibility and demonstrating the absence of a trans-acting mobilising factor in the strain of *Drosophila* used.

6.2.4 Analysis of transformed lines

Transformed *Drosophila* were obtained only when transposase was injected at a concentration of 170pM. The progeny from *white*$^+$ *Drosophila* were used to establish lines, which were then screened for *mos-white* insertions by PCR. Genomic DNA preparations were made from transformed lines (2.3.4.1) and 1μg used in PCR reactions with primer SH1, which hybridises to sequences at the C-terminal of mini-*white* and primer 338, homologous to *mos*1 DNA. An initial denaturation step at 94°C for 5 minutes was performed, before 30 cycles at 94°C, 1 minute; 56°C, 1 minute;

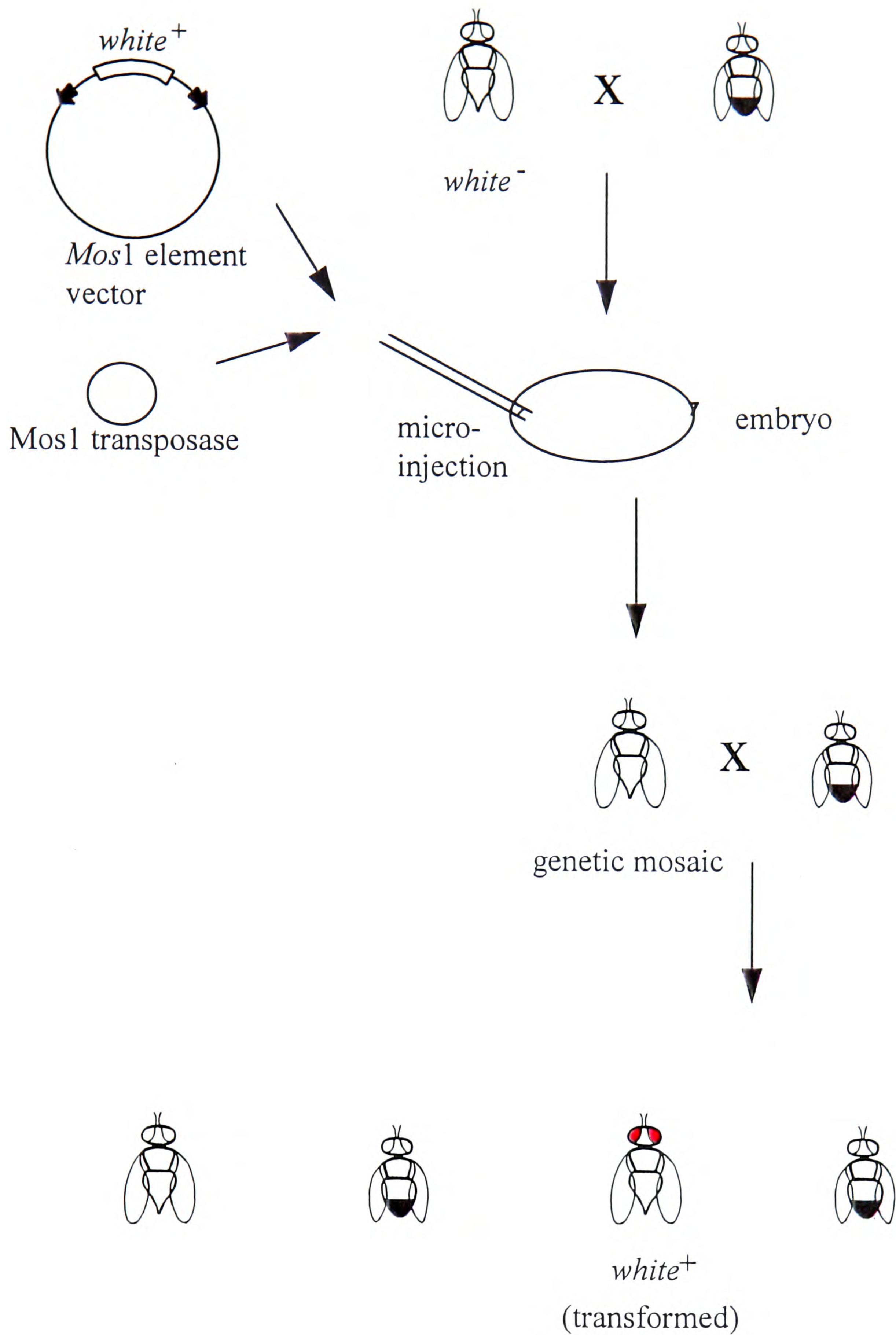


Figure 6.2. Schematic of the germline transformation of *D. melanogaster* by mini-*white* marked *mos*1, mediated by recombinant Mos1 transposase.

Transposase concentration (pM)	Number of embryos injected	Number of G_0	Number of fertile G_0	Number of transformed G_0	% fertile G_0 transformed
0 (Mock purification)	650	63	42	0	0
33	621	91	66	0	0
170	525	88	47	0	0
170	701	85	49	1	2
170	525	121	81	4	5
340	601	36	28	0	0

Table 6.1. Germline transformation data

72°C, 1 minute. A second round of PCR was then performed and the products fractionated on 1% agarose.

Figure 6.3, lane 6, shows that the control reaction for primers SH1 and 338 using p*mos-white* template, amplifies an 850bp region. This is of the predicted molecular weight. In lanes corresponding to template from transformed *Drosophila* lines PCR products of the predicted mobility were also obtained, indicating the presence of the mini-*white* marked *mos*1. Lane 7 is the negative control lane, in which the template DNA is from *D. melanogaster* w^{1118}, no amplification product is seen here as expected since this strain has a *white*⁻ phenotype.

To confirm the identity of the 850bp species, the PCR products were first subjected to gel purification, before restriction with each of *Ban*I and *Sac*I. The *Ban*I site is at position 528 within the *mos*1 sequence and digestion of the PCR product with *Ban*I is predicted to yield fragments of 670 and 180bp. Figure 6.4 demonstrates that species approximating to these sizes are obtained in the positive control lane, containing DNA amplified from p*mos-white*, and in the lanes containing the amplification products from the five transformed lines. *Sac*I is at position 787. Figure 6.5 shows that restriction with this enzyme yields DNA fragments of the predicted sizes of 450 and 400bp.

The results of the PCR and restriction analysis confirm that the five lines obtained with a *white*$^{+}$ eye-colour phenotype are bona-fide transformants.

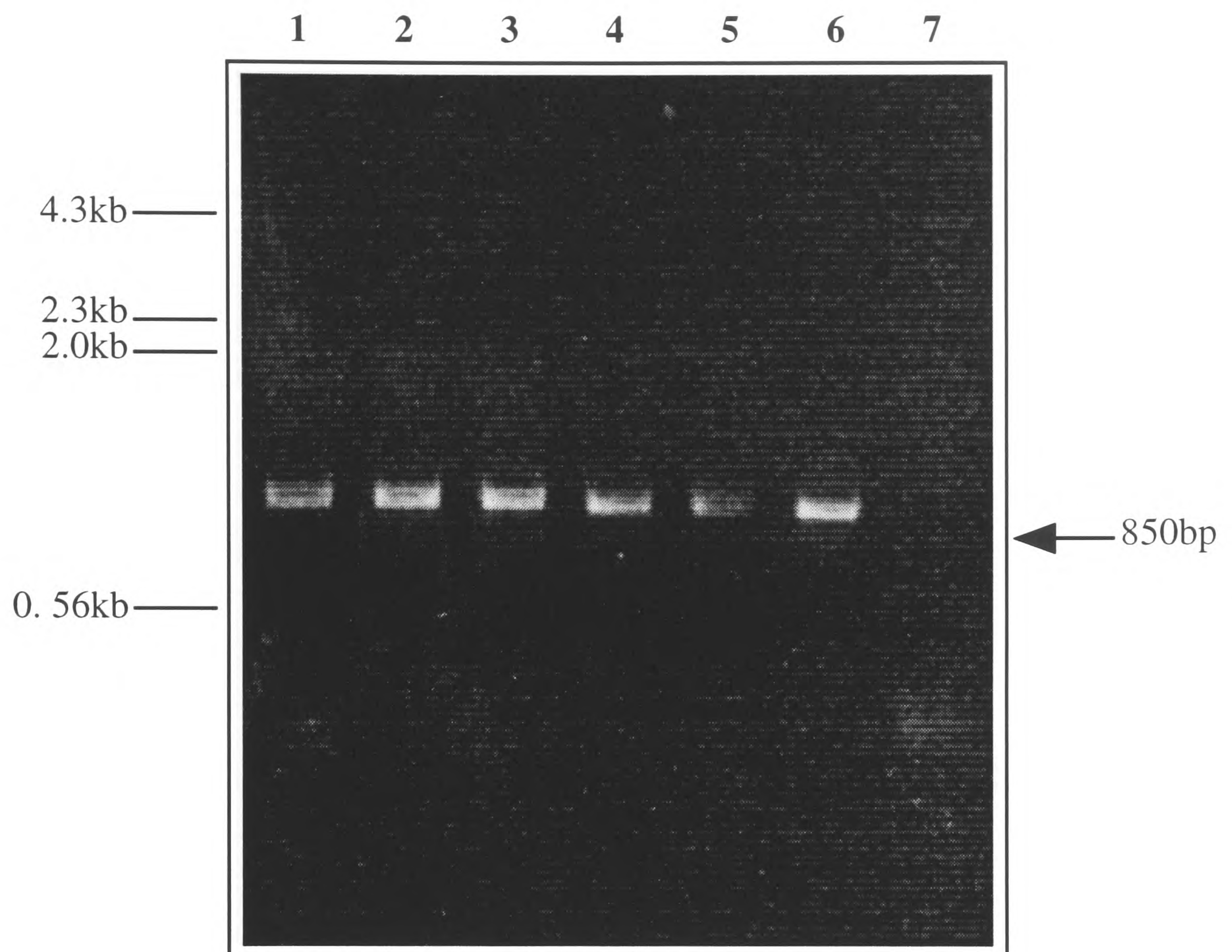


Figure 6.3 Mini-*white* marked *mos*1 insertions in *D. melanogaster*, as determined by PCR

Samples were analysed on a 1% agarose gel following two rounds of PCR amplification (see text for details).

Lanes 1-5...... Genomic DNA in PCR reaction from transformed lines 27, 74, 72, 8 and 3 respectively.

Lane 6........... p*mos-white* PCR template

Lane 7........... *D. melanogaster* w^{1118} template

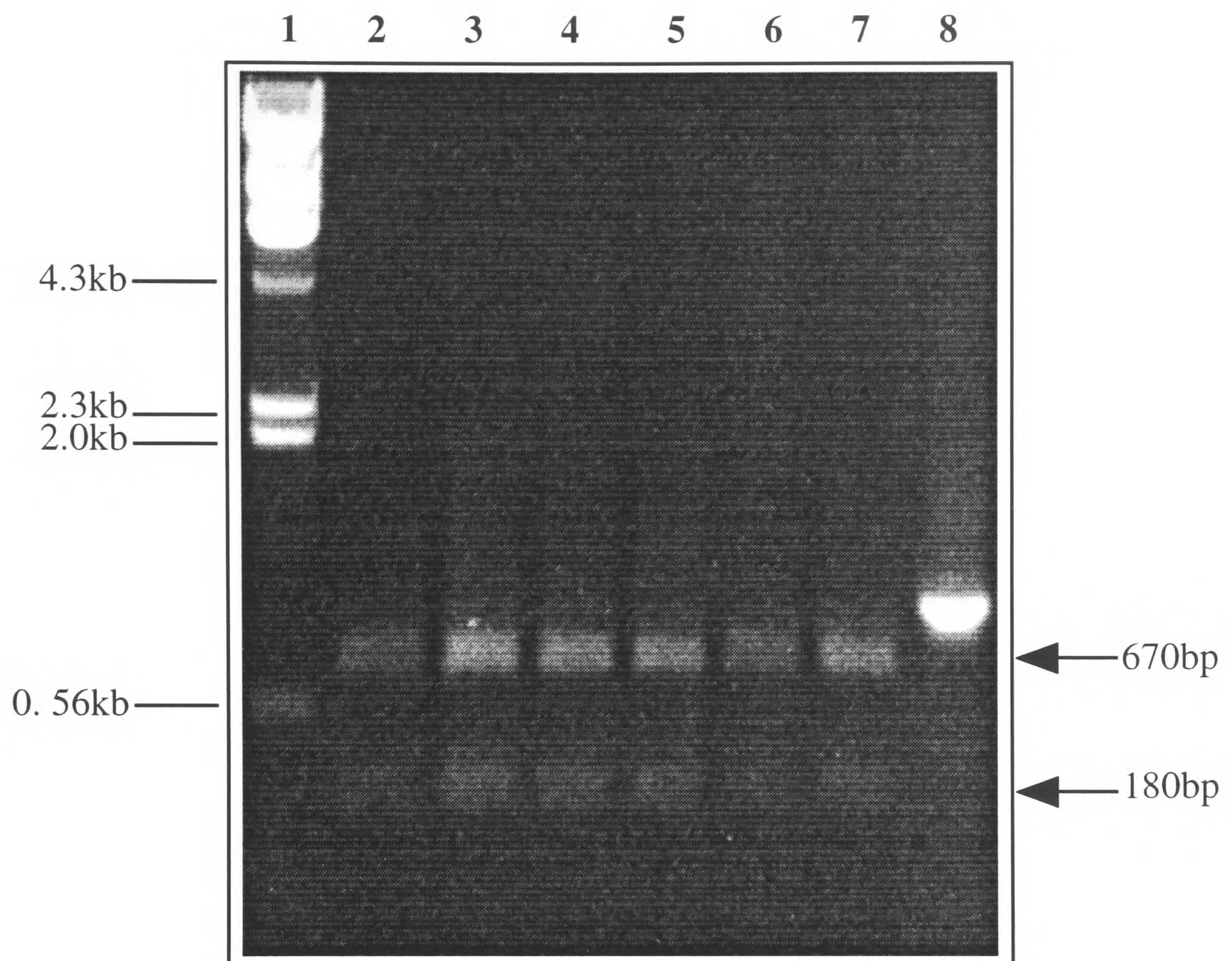


Figure 6.4 Analysis of PCR product by *Ban*I digestion

Samples were analysed on a 1% agarose gel following two rounds of PCR amplification (see text for details).

Lane 1........... *Hind* III restricted Lambda DNAmolecular weight markers

Lanes 2-6...... PCR product from transformed lines 27, 74, 72, 8 and 3 respectively following *Ban*I digestion

Lane 7........... *Ban*I digested PCR product obtained from p*mos-white* template

Lane 8........... Undigested p*mos-white* PCR product

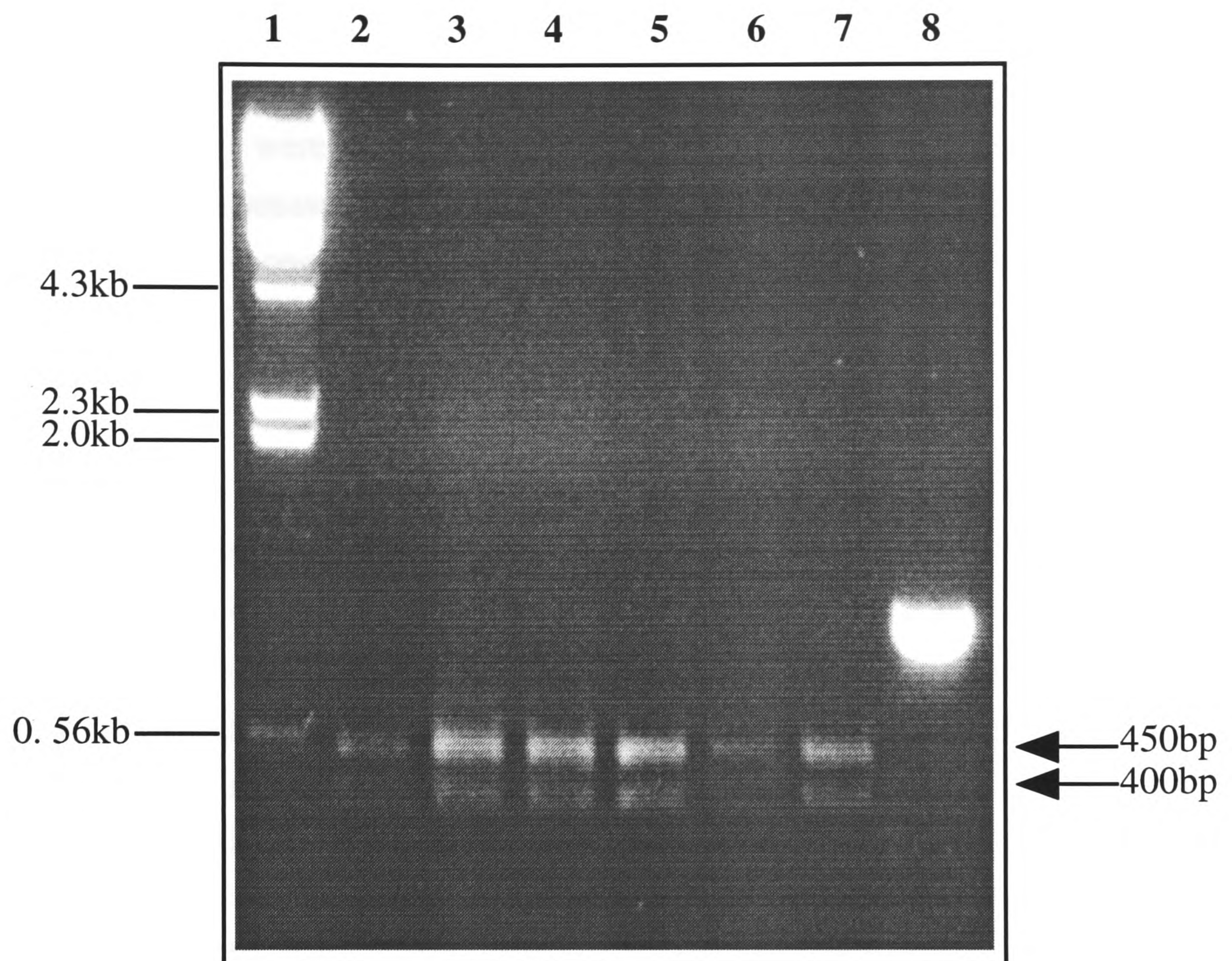


Figure 6.4 Analysis of PCR product by *Sac*I digestion

Samples were analysed on a 1% agarose gel following two rounds of PCR amplification (see text for details).

Lane 1........... *Hind* III restricted Lambda DNAmolecular weight markers

Lanes 2-6...... PCR product from transformed lines 27, 74, 72, 8 and 3 respectively following *Sac*I digestion

Lane 7........... *Sac*I digested PCR product obtained from p*mos-white* template

Lane 8........... Undigested p*mos-white* PCR product

6.3 Discussion

Here I have demonstrated that purified, recombinant *mos*1 transposase is biologically active and capable of supporting the transposition of a 5.3kb mini-*white* marked *mos*1 element in *Drosophila melanogaster* embryos. Transformants were recovered when transposase was injected at a concentration of 170pM together with plasmid p*mos-white*. No transformants were recovered with either higher (340pM) or lower (33pM) concentrations of transposase. The frequency of transformation at 170pM although detectable is low, and on one occasion no transformants were recovered. It is possible that sample sizes were too small to detect transformation events at other transposase concentrations.

On comparing the rates of transposition mediated by the direct injection of *P* and Mos1 transposases, it appears that in *Drosophila melanogaster P* transposase is able to stimulate higher levels of transposition. Caution must be taken however, in comparing germline transposition frequencies, since the quality of DNA used can affect the outcome of the experiment (Bownes, personnel communication). *P* transposase when co-injected with *rosy*$^{+}$ marked *P* element vector DNA was found to give rise to stable germline transformants. The three different molar protein:DNA ratios used in this experiment all resulted in a frequency of approximately 50% transformation of fertile G_0 flies (Kaufman and Rio, 1991). With Mos1 transposase, transformation of frequencies of two and five per cent are observed.

Previous studies using a *P* transposase encoding helper plasmid have demonstrated that *rosy*$^{+}$ marked *P* transposons stably transformed fertile G_0 flies at overall frequencies of 30-60% (Spradling, 1986). The levels of transformation of G_0 flies by the direct injection method appear to be similar to those typically observed with the DNA co-injection method.

The efficiency of integration of *mariner* elements appears to be inversely related to the length of the element. The integration of *mos*1 itself, into the *D. melanogaster* germline is estimated at 16% $\pm$ 5% among fertile G_0 flies (Lohe *et al*, 1995). Using a *mos*1 helper plasmid a 5.8kb *mos*1 element was observed to have a rate of transformation reduced several fold from the 1.3kb element. Among 320 fertile G_0 flies only two red eyed transformants were obtained (Lohe *et al*, 1995). On increasing the size of the *mos*1 construct to 13.2kb, a similar transformation rate was detected (Lidholm *et al*, 1993).

The rate of germline transformation of *white* marked *mos*1 with 170pM transposase is comparable with that found when a similarly marked element was injected with a helper plasmid as a source of transposase. In the two experiments the difference between the elements is in the site of insertion of the mini-*white* gene in *mos*1. One contains the mini-*white* gene inserted into the *Sac*I site at nucleotide 788 of *mos*1 (Lohe *et al*, 1995), and the other, the *Sal*I site at 349.

Thus the method of germline transformation via direct injection of *mos*1 transposase and transgenic *mos*1 elements looks promising as a transformation system. It will be interesting to determine if it is possible to transform species outside of the *Drosophila* species sub group using this technique.

Although this technique has potential as a basis for being a broad based transformation system in insects some limits exist which I would like to discuss:

Mariner would appear to be a good vector for the delivery of DNA since *in vitro* experiments indicate that it can transpose without the aid of host factors (Dawson, personal communication) and it is already established in the genomes of a wide variety of species. However, the broad distribution of *mariner* may also cause problems as functional, endogenous *mariner* elements may reside in the genomes of insects to be transformed. This may compromise the stability of the transgene, making the maintenance of a stable transgenic line impossible. However, initial indications are that this may not be a significant problem, since integrated *mos*1 elements in *D. melanogaster* appear to be relatively stable despite the introduction of various *mos*1 constructs expressing functional transposase (Lidholm *et al*, 1993; Lohe *et al*, 1995).

A second consideration is the sensitivity of transposition to element length. It is a possible that *mariner* based vectors may only be useful as efficient gene vectors in insects if the amount of DNA inserted between the ends of the element is kept to a minimum.

A third feature of this system is that it may be difficult to deliver the vector to the cells of interest in many insects and arthropods. Vector sequences are delivered directly to the developing germ cells by introducing DNA into the pole plasm of preblastoderm embryos by injection. Direct introduction of DNA in this manner will not be possible for some insects because of unique properties of the egg such as impenetrable chorions or because of the reproductive physiology of the insect. it would not be possible with

viviparous insects such as Tsetse flies for example, or those with minute eggs. Direct injection into the gonads or haemocoel may be possible alternatives for some species (Presnail and Hoy, 1992).

Chapter 7

Introduction of the *C. elegans* transposable element *Tc*3 into *D. melanogaster*

7.1 Introduction

The transposable element *Tc*3 from *C. elegans* is 2335bp long and has terminal inverted repeats of 462bp (Collins *et al*, 1989). Within the inverted repeats is a transposase gene composed of two exons, which encodes a protein of 38kD (van Leunan *et al*, 1993).

*Tc*3 belongs to the *Tc1/mariner* family of transposons. One of the distinctive features of this family is their spread among many different species: fungi, ciliates, planarians, arthropods, nematodes and vertebrates (Langin *et al*, 1994; Tausta and Klobutcher, 1989; Garcia-Fernandez *et al*, 1993; Brezinsky *et al*, 1990; Franz and Savakis, 1991; Henikoff, 1992; Robertson, 1995; Smit and Riggs, 1996). This has not been observed for any other class of transposons.

In vitro transposition using purified, recombinant transposase has recently been demonstrated for a number of members of this family (Dawson, personal communication; Lampe *et al*, 1996; Vos *et al*, 1996). The independence of species-specific factors may explain why members of the *Tc1/mariner* family are dispersed over so many different phyla, possibly by horizontal transfer. This is in contrast to *P* elements which are restricted to *Drosophila* species. A possible candidate for a species-specific host factor in *P* transposition is the inverted repeat binding protein, IRBP (Beall *et al*, 1994).

The *Tc*3 transposable element may be an ideal tool for the genetic manipulation of many different species. This chapter describes how this idea was tested. Two constructs based on the *Tc*3 element were introduced by *P* element mediated germline transformation into *Drosophila melanogaster*, a species in which it is not normally found. One construct contains a *Tc*3 element marked with a *Drosophila* gene. The second contains the *Tc*3 transposase gene behind the *Drosophila* hsp26:Sgs-3 cassette. This should enable transcription of the *Tc*3 ORF in the female germline. Crossing the two transformed strains allows the mobilisation of *Tc*3 to be studied in a foreign genome.

7.2 Results

7.2.1 Choosing suitable genetic markers

Three *Drosophila* genes; *vermillion* (*v*), *rosy* (*ry*) and *yellow* (*y)* (Lindsley and Zimm, 1992) were employed as genetic markers, to enable both excision and transposition of *Tc*3 to be studied in *Drosophila melanogaster.* The gene products of *vermillion* and

rosy are required for wild-type eye colouration. In the absence of a functional *vermillion* allele, eyes are bright orange in colour and in flies lacking a wild-type *rosy* allele, eyes are dark red. In *Drosophila* lacking functional copies of both *vermillion* and *rosy* alleles, eyes have a pale orange phenotype. The *yellow* locus controls the melanotic pigment pattern of the adult fly. Flies lacking a functional *yellow* allele differ from wild-type in having a cuticle which is yellow-brown in colour rather than black.

7.2.2 Plasmid construction

Plasmids pRP1406 and pRP1436, for transformation into *Drosophila* were constructed by Dr. Henri van Leunan.

Plasmid pRP1406 (figure 7.1) contains a *Tc*3 element marked with the *yellow* gene from *Drosophila.* A plasmid containing *Tc*3 inserted into the *C. elegans unc*22 gene was initially subjected to partial *Xho*I digestion. A 4.9kb *Sal*I fragment containing the *yellow* gene from plasmid pBY was then ligated into the *Xho*I site in *Tc*3, disrupting the ORF. A 7.3kb *XhoI/Sal*I fragment containing *yellow* marked *Tc*3 and flanking *unc*22 gene sequence was then excised and ligated into the *Sal*I site in the multiple cloning site of plasmid pYC1.8 (Fridell and Searles, 1991), a *P* element transformation vector containing the *vermillion* allele within the *P* element inverted repeats, to create pRP1406.

Plasmid pRP1436 (figure 7.1) was created by initially ligating a *Kpn*I fragment containing the hsp26:Sgs-3 cassette immediately upstream of the Tc3 ORF in pBluescribe. The hsp26:Sgs-3 cassette contains two copies of the female germ-line specific transcriptional regulator of the *Drosophila* hsp26 gene upstream of the *Drosophila* Sgs-3 gene promoter (see Frank *et al*, 1992). A 7.2kb *Hind*III fragment containing the $rosy^{+}$ gene was then ligated into the *Hind*III site in the pBluescribe polylinker, before addition of an *XbaI/Not*I fragment containing *P* element termini into the pBluescribe polylinker.

7.2.3 Construction of *Drosophila melanogaster* strain $y^{1}v^{36f}; ry^{506}$

(This work was performed in conjunction with Dr. Marie-Christine Chabboissier).

In order to use the *vermillion, rosy* and *yellow* genes as genetic markers, it was necessary to construct a strain of *Drosophila* containing non-functional copies of these genes. Two different strains of *Drosophila melanogaster* containing a combination of null alleles of the three genes were available; $v^{36f}; y^{506}$ and y^{1}, ry^{506}. Alleles v^{36f} and y^{1} are spontaneous mutations and ry^{506} is a γ-ray induced mutation (see Lindsley

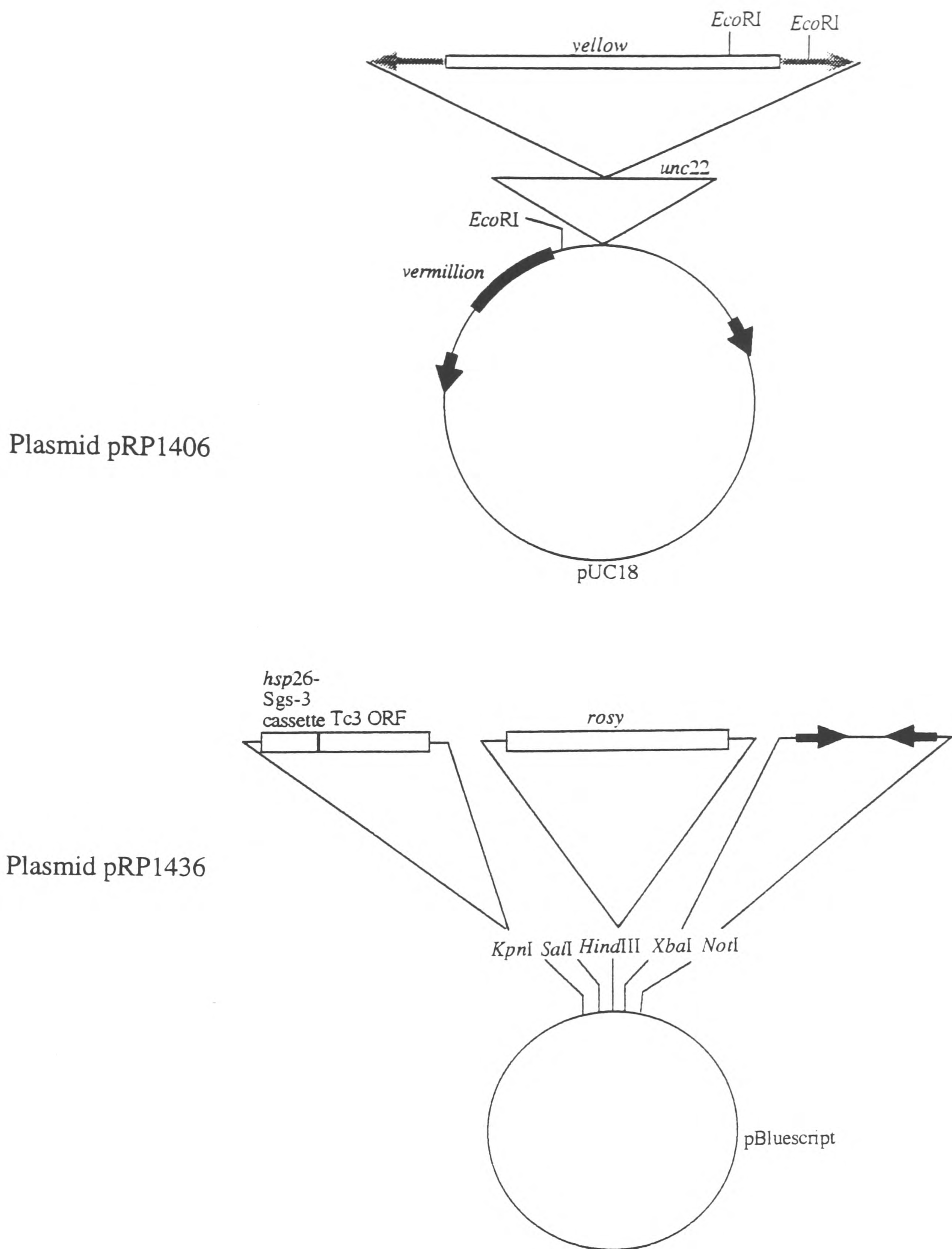


Figure 7.1. Plasmids pRP1406 and pRP1436.

Inverted repeats from *P* elements are shown as black arrows and from *Tc*3 as shaded arrows. The arrows are oriented towards the termini of the elements. *Drosophila* marker alleles and the Tc3 ORF are labelled.

and Zimm, 1992). The crosses shown in figure 7.2 were carried out to introduce all three null alleles into the same genetic background. Recombination between X chromosomes in females resulting from the first cross, brings the y^1 and v^{36f} alleles onto the same chromosome. Sibling crosses were then performed to produce a homozygous strain of *Drosophila* of genotype y^1v^{36f}; ry^{506}.

7.2.4 The creation of transgenic lines by *P* element mediated mutagenesis

P helper plasmid pπ25.1wc, pRP1406 and pRP1436 were prepared as in 2.3.2.3. and the DNA dissolved in dH_2O before quantification and storage at -20°C.

pRP1406 was initially diluted to 400ng/μl in injection buffer and the helper plasmid pπ25.1wc to 100ng/μl before injection into dechorionated *D. melanogaster* y^1v^{36f}; ry^{506} embryos, as described in 2.3.4.3. Embryos were incubated at 18°C till hatching, and then at 25°C. The G_0 were backcrossed to y^1v^{36f}; ry^{506} virgins and progeny examined for a yellow$^+$vermillion$^+$ phenotype (wild-type body colour, dark-red eyes). Transformants were crossed to y^1v^{36f}; ry^{506} flies and homozygous lines established by sibling matings. Four transformed lines were obtained; Tc3y^+v^+ 18, 25, 95 and 96.

Plasmid pRP1436 was similarly co-injected with *P* helper plasmid into y^1v^{36f}; ry^{506} embryos. Transformants were selected by their bright orange eye colour. Only one transformed line was obtained; Tc3ry^+.

Lines of the transformants produced were screened for insertions in Southern transfer experiments. Genomic DNAs prepared from transformed lines Tc3y^+v^+ 18, 25, 95 and 96 and from untransformed y^1v^{36f}; ry^{506} were subjected to *Eco*RI digestion before fractionation on a 1% agarose gel and Southern blotting. Restriction with *Eco*RI should yield fragments of 1.6kb and 5.8kb, which contain the *yellow* gene. The DNA was then hybridised to a ^{32}P labelled, 4.9kb *Sal*I fragment corresponding to the *yellow* gene from plasmid pBY. In figure 7.3, lanes 1-5, which contain DNA from transformed flies, two bands were observed to hybridise to the probe. These fragments, one of 1.6kb and one of 5.8kb, are of the predicted size, indicating that lines *Tc*3y^+v^+ 18, 25, 95 and 96 are true transformants. In the control lane containing DNA from strain y^1v^{36f}; ry^{506}, no bands hybridising to the probe were observed, as expected.

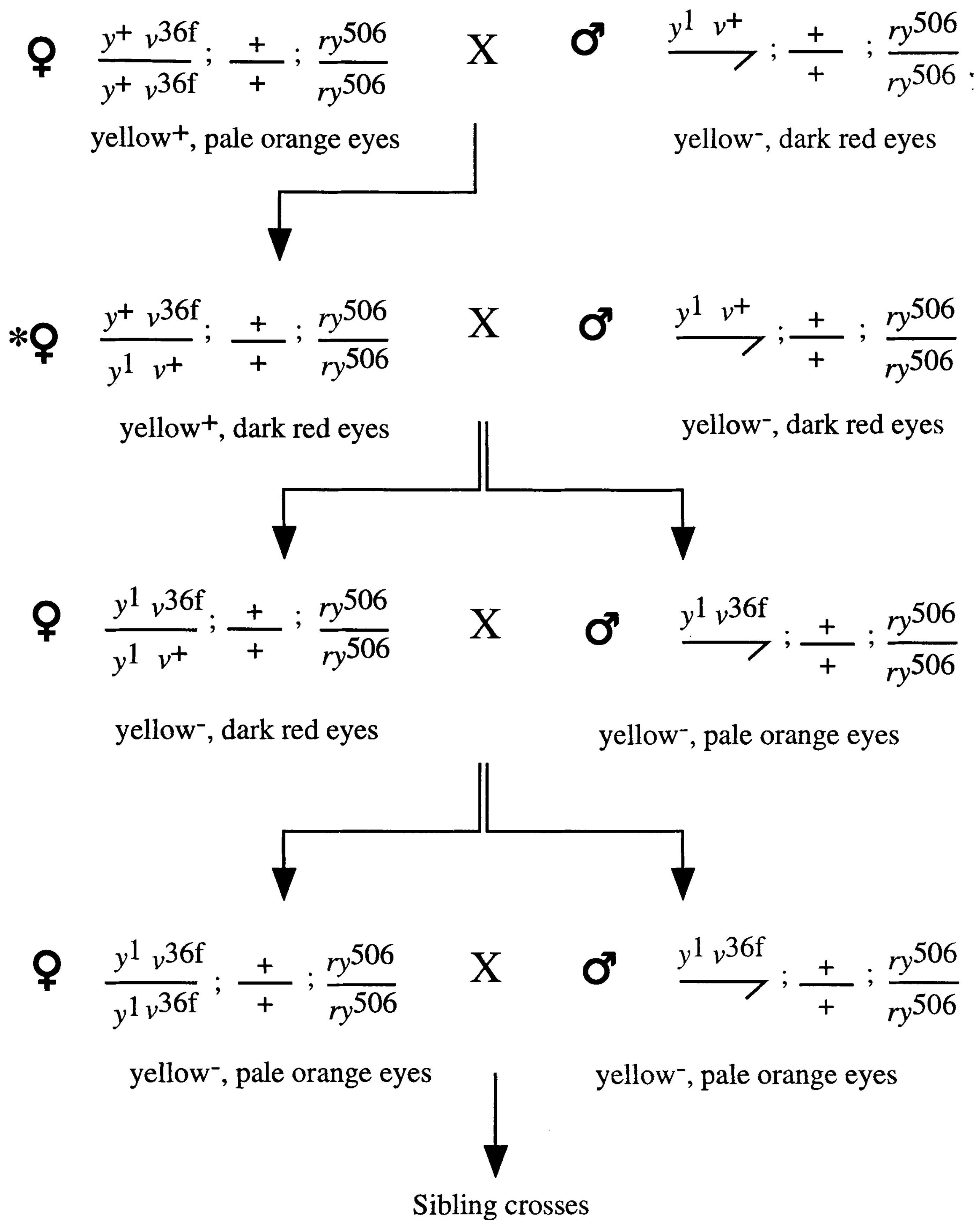


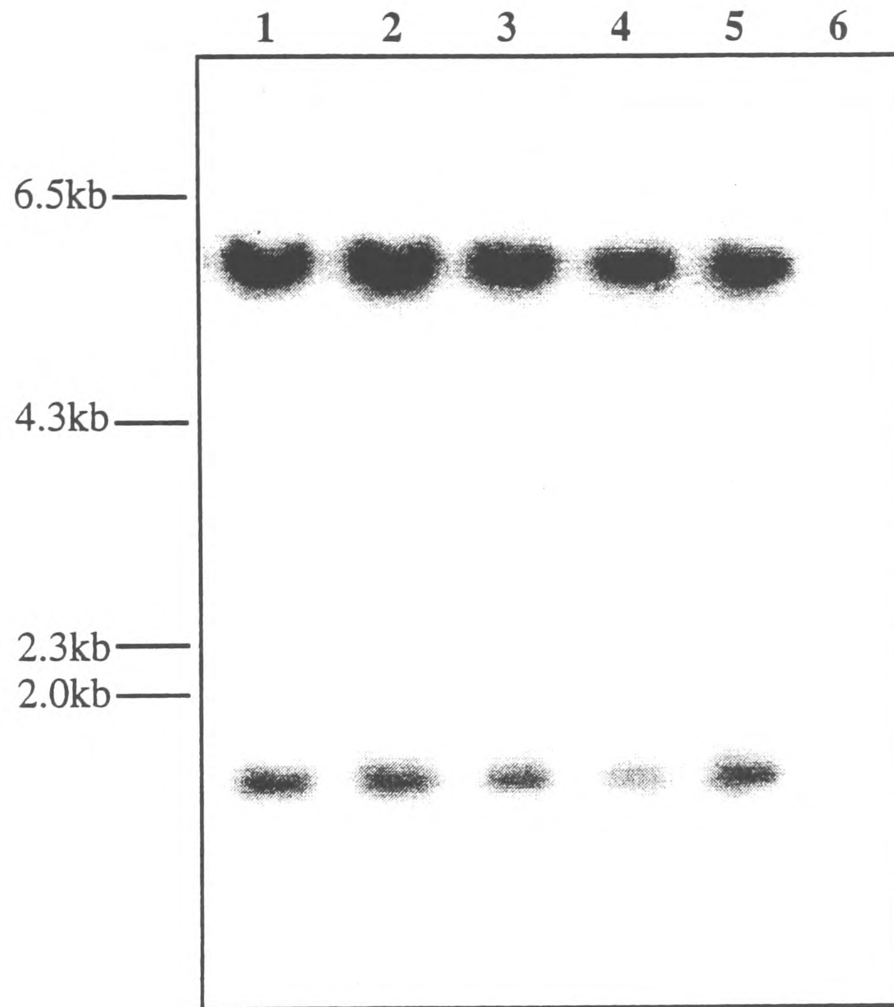
Figure 7.2. Constructing *Drosophila melanogaster* strain y^1 $v36^f$; ry^{506}.

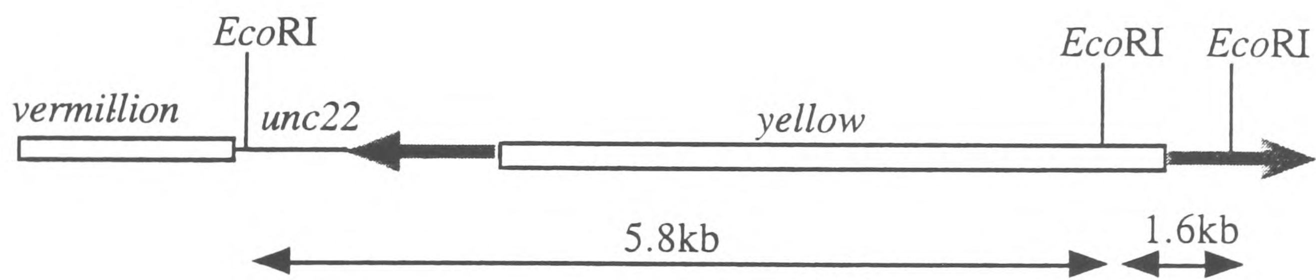
Recombination in the female marked *, is required to bring the y^1 and v^{36f} alleles onto the same X chromosome.

Figure 7.3. Genomic Southern to show *P* element insertions of *Tc3y*$^{+}$*v*$^{+}$ lines produced by transformation.

DNA was digested with *Eco*RI before fractionation on 1% agarose, Southern blotting and hybridisation to a ^{32}P labelled *yellow* probe. In transformed flies, fragments of 1.6kb and 5.8kb are predicted to contain the *yellow* gene sequence following *Eco*RI digestion. Bacteriophage lambda DNA digested with *Hind*III was used as a molecular size marker.

Lanes 1-5..... contain ~ 10µg of *Eco*RI digested genomic DNA from transformed lines *Tc3y*$^{+}$*v*$^{+}$ 18, 25, 95_2, 95_3 and 96 respectively.
Lane 6........ contains ~10µg of *Eco*RI digested genomic DNA from the original Drosophila strain y^1 v^{36f}, ry^{506}.


1
2
3
4
5
6
6.5kb
4.3kb
2.3kb
2.0kb


EcoRI
EcoRI
EcoRI
vermillion
unc22
yellow
5.8kb
1.6kb

Genomic DNA prepared from the transformed line Tc3*ry* + was digested with *SalI/NotI* and Southern blotted. Plasmid pRP1436 was also *SalI/NotI* digested and a ~3.5kb restriction fragment containing the Tc3 ORF gel purified before ^{32}P labelling. This radiolabelled fragment was then used to probe the Southern blot. As expected a single band of ~3.5kb is recognised, see figure 7.4, which is absent in the control lane, confirming that the bright orange eyed line contains the Tc3 ORF.

7.2.5 Localising the chromosome of insertion

A series of genetic crosses, detailed in figure 7.5, were performed to determine the chromosome site at which the *yellow* marked *Tc*3 element had inserted in the four transformed lines; *Tc*3y^+v^+ 18, 25, 95 and 96.

To determine if the insert is present on the X chromosome *Tc*3y^+v^+ males, containing the *yellow* marked *Tc*3 element, were crossed with virgin females from strain y^1v^{36f}; ry^{506}. This cross results in female progeny which have dark red eyes and a yellow$^+$ phenotype, and males which are yellow$^-$ with pale orange eyes, if the insert is on the X chromosome. The same cross results in females that are yellow$^-$ with pale orange eyes and males that are yellow$^-$ with dark red eyes, if the insert is on the Y chromosome. If the insert is present on an autosome then the yellow$^+$ vermillion$^+$ phenotype will not segregate with a single sex.

The *Tc*3y^+v^+ transformed lines were also crossed with *Drosophila* strains containing second or third balancer chromosomes with dominant markers as shown in figure 7.5. The marker for the second chromosome balancer used was *Curly of Oster* (*CyO*) in which flies have wings curled upwards. Flies which were yellow$^+$vermillion$^+$ and contained the balancer were selected and crossed with the original y^1v^{36f}; ry^{506} strain. Insertion on the second chromosome gives an F_2 progeny of phenotype either yellow$^+$ vermillion$^+$ CyO$^+$, or yellow$^-$ vermillion$^-$ CyO$^-$.

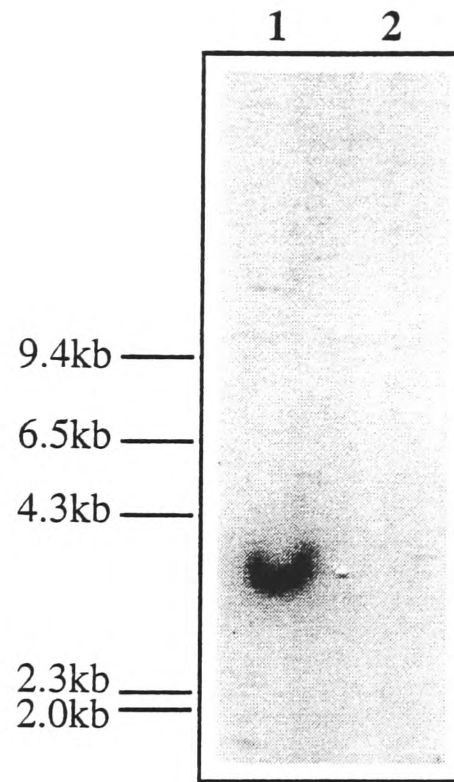
The marker for the third chromosome balancer used was *Stubble* (*Sb*), which is manifested as bristles which are less than half the normal length. Following the cross illustrated in figure 7.5 insertion on the third chromosome results in F_2 flies of phenotype yellow$^+$ vermillion$^+$ Sb$^+$ or yellow$^-$vermillion$^-$ Sb$^-$.

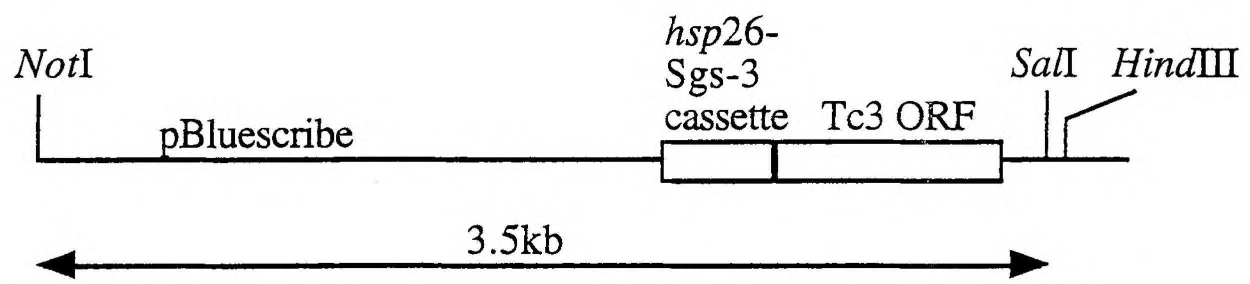
Following these crosses allowed the *Tc*3y^+v^+ insert to be mapped to the X chromosome in line *Tc*3y^+v^+ 18, to chromosome II in line *Tc*3y^+v^+ 25 and to chromosome III in lines *Tc*3y^+v^+ 95 and 96.

Figure 7.4. Genomic Southern to show *P* element insertions of *Tc3ry*$^+$ lines produced by transformation.

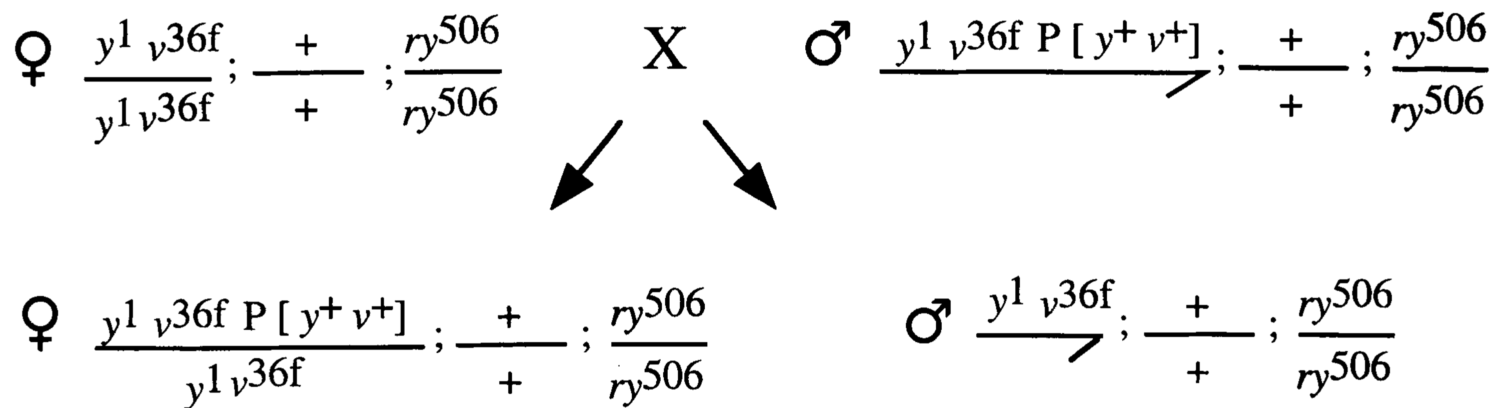
DNA was digested with *Not*I/*Sal*I before fractionation on 1% agarose, Southern blotting and hybridisation to a ^{32}P labelled Tc3 ORF sequence probe. The *Not*I/*Sal*I fragment from transformed flies, predicted to contain the Tc3 ORF sequence is 3.5kb. Bacteriophage lambda DNA digested with *Hind*III was used as a molecular size marker.

Lanes 1....... contains ~ 10µg of *Not*I/*Sal*I digested genomic DNA from the transformed line *Tc3ry*$^+$.
Lane 2........ contains ~10µg of *Not*I/*Sal*I digested genomic DNA from the original Drosophila strain y^1 v^{36f}, ry^{506}.


1
2
9.4kb
6.5kb
4.3kb
2.3kb
2.0kb

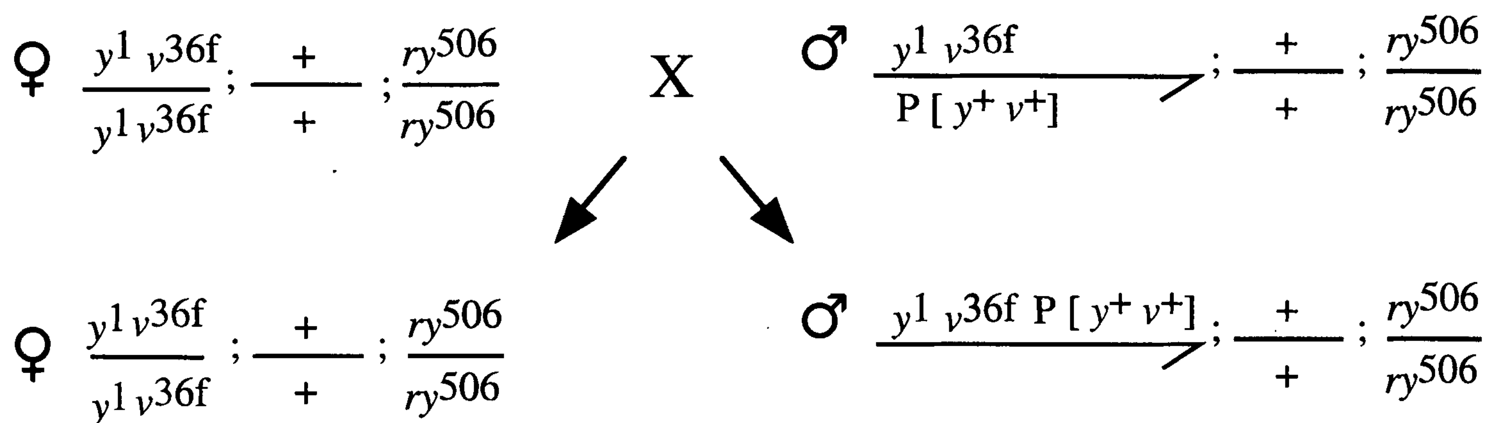

NotI
pBluescribe
hsp26-
Sgs-3
cassette
Tc3 ORF
SalI
HindIII
3.5kb

I. On the X chromosome



all females are yellow$^+$ with dark red eyes   all males are yellow$^-$ with pale orange eyes

II. On the Y chromosome



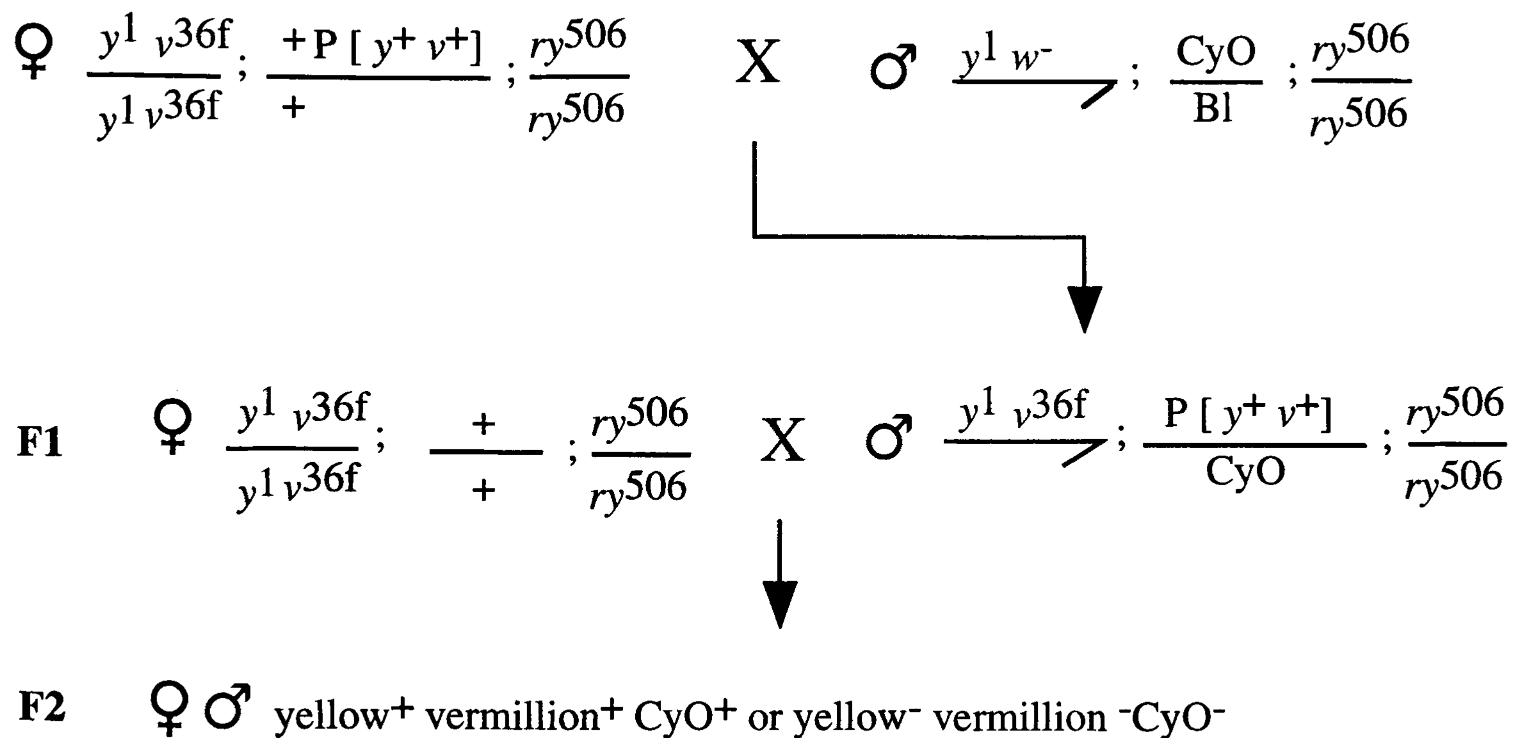
all females are yellow$^-$ with pale orange eyes   all males are yellow$^+$ with dark red eyes

Figure 7.5. Localising the chromosome of insertion for the *yellow*$^+$ marked *Tc*3 element, following *P* element mediated transformation.

Tc3y^+v^+ males, containing the *yellow* marked *Tc*3 element, were crossed with virgin females from strain y^1 v^{36f}; ry^{506}. This cross will produce female progeny which have dark red eyes and a yellow$^+$ phenotype, and males which are yellow$^-$ with pale orange eyes, if the insert is on the X chromosome.

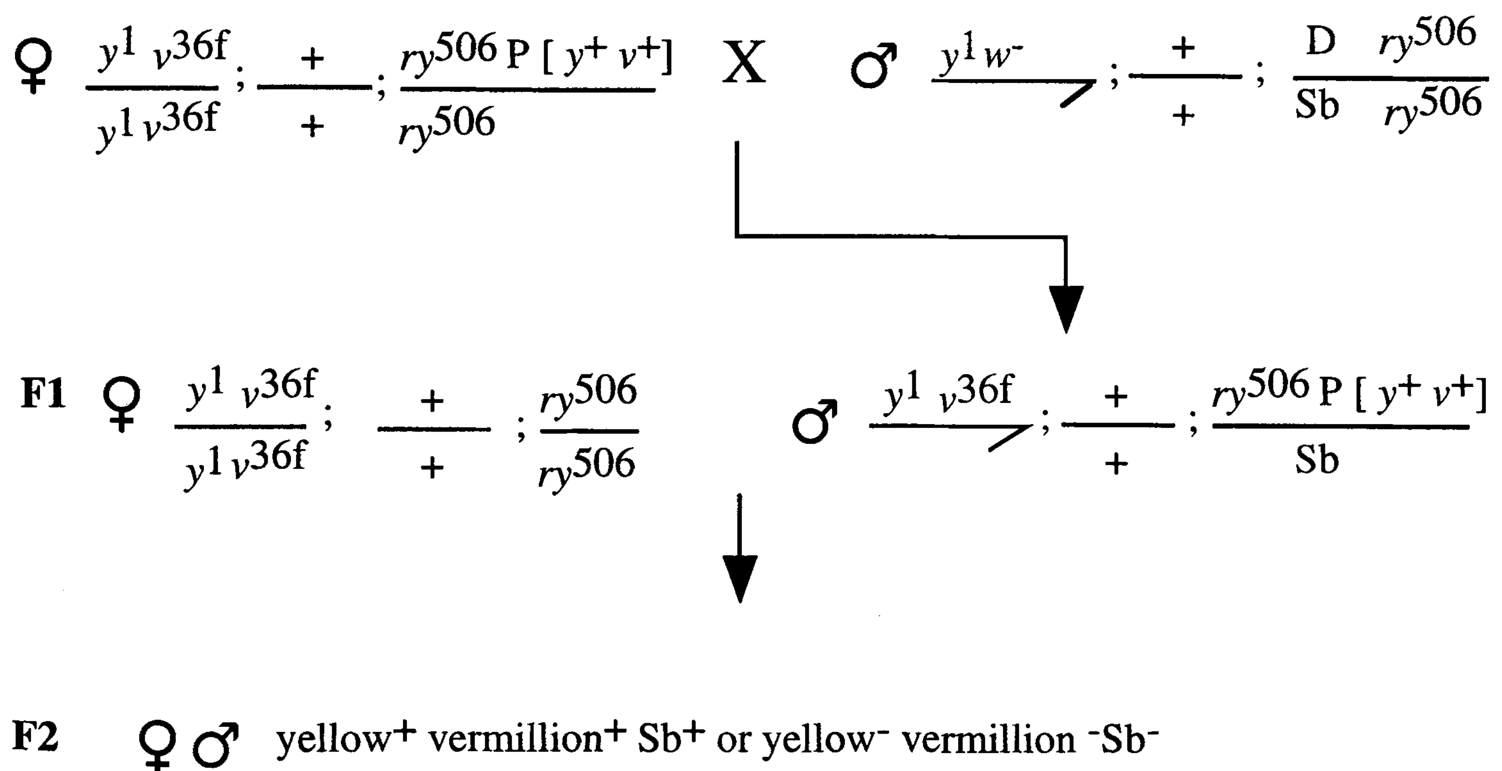
The same cross will produce females that are yellow$^-$ with pale orange eyes and males that are yellow$^+$ with dark red eyes, if the insert is on the Y chromosome.

III. On chromosome II



F2 ♀ ♂ yellow$^+$ vermillion$^+$ CyO$^+$ or yellow$^-$ vermillion $^-$CyO$^-$

IV. On chromosome III



F2 ♀ ♂ yellow$^+$ vermillion$^+$ Sb$^+$ or yellow$^-$ vermillion $^-$Sb$^-$

The Tc3y^+v^+ transformed lines were also crossed with *Drosophila* strains containing second or third balancer chromosomes with markers. The marker for the second chromosome balancer is *Curly of Oster* (*CyO*) and for the third is *Stubble* (*Sb*). Flies which are yellow$^+$vermillion$^+$ and contain the balancer were selected and crossed with the original y^1 v^{36f}; ry^{506} strain. If the insertion is on a chromosome homologous to the balancer, then the progeny will be either yellow$^+$vermillion$^+$marker$^+$ or yellow$^-$ vermillion$^-$ marker$^-$.

In the transformed line *Tc3ry*$^+$, which contains the Tc3 ORF behind a *Drosophila* promoter, only females are observed in each generation. It was assumed that this is because the *Tc*3 construct has transposed into an essential gene on the X chromosome. As males contain a single copy of the X chromosome, only those that inherit a non-transformed chromosome will be viable. Females which contain the *Tc*3 construct are heterozygous for the insertion. They contain one non-transformed X chromosome and are therefore also viable.

7.2.6 Mobilisation of the marked *Tc*3 element

To study the mobility of the *yellow* marked *Tc*3 element in *Drosophila melanogaster* a series of crosses were performed such that the excision and transposition of *Tc*3 in the germline and soma could be scored.

The mating scheme shown in figure 7.6 was followed to examine the ability of *Tc*3 to excise in *Drosophila.* Virgin females containing the *rosy* marker and therefore the Tc3 ORF behind the *Drosophila* hsp26:Sgs-3 cassette, were initially mated with males containing the *yellow* marked *Tc*3 element. The males are shown to contain the insert on the 3rd chromosome (line 96) in figure 7.6, crosses were also performed with males containing the insert on the X Chromosome (line 18). *Tc*3 was followed in the soma and the germline through the presence of the dominant *yellow* marker gene, which gives a wild-type body colour. If somatic excision of *Tc*3 were to take place, then cells in which this had occurred would be *yellow*$^-$, resulting in the appearance of yellow spots on the abdomen of the F1 generation. As the hsp26:Sgs-3 cassette has previously been reported to drive transcription specifically in the female germline (Frank *et al*, 1992), somatic excision of *Tc*3 was unexpected, but nevertheless looked for. The results of somatic excision following the crossing of *rosy*$^+$ virgin females containing the Tc3 ORF, with males from lines *Tc*3*y*$^+$*v*$^+$18 and 96 are shown in tables 7.1 and 7.2.

If germline excision of *Tc*3 occurs then F2 flies shown in the scheme in figure 7.6, which contain the *vermillion*$^+$ allele will be *yellow*$^-$. This is manifested as flies with dark red eyes that have yellow-brown abdomen. The results of germline excision are shown in tables 7.3 and 7.4.

Transposition of *Tc*3 can be detected by looking for the presence of the *yellow*$^+$ allele in the absence of *vermillion*$^+$. Crosses were initially performed as for excision to introduce the marked *Tc*3 element and the source of Tc3 transposase into the same fly. Transposition of the marked *Tc*3 element to a new location, in the germline of these

Figure 7.6. Excision and transposition of *Tc3* in *Drosophila*.

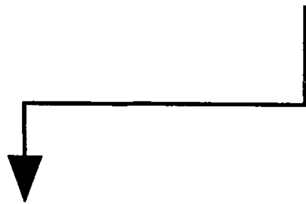
The *P* element insertions of the marked *Tc*3 element, and the Tc3 ORF are shown as *P*[Tc3y^+ v^+] and *P*[Tc3 ORF ry^+] respectively in the maring scheme. *P*[Δ v^+] refers to *P*[Tc3y^+ v^+] following excision of the *yellow* marked *Tc*3 element. All females used in the mating scheme are virgins. A cross is first performed to bring the source of Tc3 transposase into contact with the marked transposon.

F1 females were screened for somatic excision events. Cells from which somatic excision of *Tc*3 takes place will be *yellow*$^-$, resulting in yellow spots on the abdomen of the F1 generation.

The F2 generation were screened for germline excision/transposition events. If germline excision of *Tc*3 takes place then F2 flies which contain the *vermillion*$^+$ allele will be *yellow* $^-$. This is manifested as flies with dark red eyes that have yellow-brown abdomen. Females may also have inherited the ry^+ allele, and have a yellow$^-$ phenotype with wild-type eyes.

If transposition of *Tc*3 takes place in the germline of the F1 generation, then providing that the chromosome to which the *Tc*3 element has transposed does not co-segregate with the chromosome containing the *P*[Δ v^+] insert, then males and females which are *yellow*$^+$ *vermillion*$^-$ will be detectable. The genotypes shown in figure 7.6, demonstrating transposition of *Tc*3, are only two examples of a number of possible genotypes that could result. Females may also inherit the ry^+ allele.

♀ $\frac{y^{1}\ v^{36f}\ P\ [Tc3\ ORF\ ry^{+}]}{y^{1}\ v^{36f}}$; $\frac{+}{+}$; $\frac{ry^{506}}{ry^{506}}$ **X** ♂ $\frac{y^{1}\ v^{36f}}{\nearrow}$; $\frac{+}{+}$; $\frac{ry^{506}\ P\ [Tc3\ y^{+}\ v^{+}]}{ry^{506}\ P\ [Tc3\ y^{+}\ v^{+}]}$



F1

♀ $\frac{y^{1}\ v^{36f}\ P\ [Tc3\ ORF\ ry^{+}]}{y^{1}\ v^{36f}}$; $\frac{+}{+}$; $\frac{ry^{506}\ P\ [Tc3\ y^{+}\ v^{+}]}{ry^{506}}$ **X** ♂ $\frac{y^{1}\ v^{36f}}{\nearrow}$; $\frac{+}{+}$; $\frac{ry^{506}}{ry^{506}}$

(screened for somatic excision)

transposition examples

germline excision examples

F2

♂ $\frac{y^{1}\ v^{36f}}{\nearrow}$; $\frac{+}{+}$; $\frac{ry^{506}}{ry^{506}\ [Tc3\ y^{+}]}$

yellow+, pale orange eyes

♂ $\frac{y^{1}\ v^{36f}}{\nearrow}$; $\frac{+}{+}$; $\frac{ry^{506}\ P\ [\Delta\ v^{+}]}{ry^{506}}$

yellow-, dark red eyes

♀ $\frac{y^{1}\ v^{36f}}{y^{1}\ v^{36f}}$; $\frac{[Tc3\ y^{+}]}{+}$; $\frac{ry^{506}}{ry^{506}}$

yellow+, pale orange eyes

♀ $\frac{y^{1}\ v^{36f}\ P\ [Tc3\ ORF\ ry^{+}]}{y^{1}\ v^{36f}}$; $\frac{+}{+}$; $\frac{ry^{506}\ P\ [\ \Delta\ v^{+}]}{ry^{506}}$

yellow-, wild-type eyes

(screened for germline excision/transposition)

	Number of flies in the F1 generation	Number of y^+ v^+; ry^+ flies in the F1 generation	Number of somatic mosaic flies y^- v^+; ry^+, in the F1 generation
Females	520	227	4 (all with y^- wings but remainder of abdomen y^+)
Males	212	0	0

Table 7.1.

	Number of F1 flies, (from cross shown in figure 7.6)	Number of y^+ v^+; ry^+ flies in the F1 generation	Number of somatic mosaic flies y^- v^+; ry^+ in the F1 generation
Females	410	198	0
Males	179	0	0

Table 7.2.

Tables 7.1 and 7.2. Somatic excision of *yellow* marked *Tc*3

The first cross shown in figure 7.6 was followed to bring the marked *Tc*3 element and Tc3 transposase into the same background. Cells from which somatic excision of *Tc*3 takes place will have a $yellow^-$ phenotype.

Table 7.1, refers to the results obtained when line Tc3y^+v^+ 18 was used to provide the marked *Tc*3 element. Table 7.2, refers to the results obtained when line Tc3y^+v^+ 96 was used. The total number of F1 flies, the number of y^+ v^+; ry^+ flies and the number exhibiting somatic mosaicism are tabulated.

The number of males tabulated is ~ half the number of females, due to ~50% of males inheriting the reccesive lethal X chromosome containing the ry^+ insertion.

	No. of y^+ v^+ flies in the F2 generation	No. of y^- v^- flies in the F2 generation	No. of y^+ v^- flies in the F2 generation	No. of y^- v^+ flies in the F2 generation
Females	885	701	0	0
Males	472	469	0	0

Table 7.3.

	No. of y^+ v^+ flies in the F2 generation	No. of y^- v^- flies in the F2 generation	No. of y^+ v^- flies in the F2 generation	No. of y^- v^+ flies in the F2 generation
Females	460	425	0	0
Males	177	198	0	0

Table 7.4.

Table 7.3 and 7.4. Germline excision and transposition of *Tc3*

Virgin y^+ v^+; ry^+ females from the F1 generation in figure 7.4 were crossed with strain y^1 v^{36f}, ry^{506} and the progeny scored for eye and body phenotype. Flies that are y^- v^+ would represent germline excision events. Those that are y^+ v^- would represent transposition of the marked *Tc3* element.

Table 7.3, refers to the results obtained when line Tc3y^+v^+ 18 was used to provide the marked *Tc3* element. Table 7.4, refers to the results obtained when line Tc3y^+v^+ 96 was used.

flies will result in some chromosomes containing the *yellow* + allele, since it ispresent within the *Tc*3 element, but lacking *vermillion*+. Crossing these flies with strain $y^1 v^{36f}$; ry^{506} will allow the detection of flies which are yellow+ vermillion- in phenotype. This mating scheme is depicted in figure 7.6. The F2 generation resulting from these crosses were scored for eye and body colour and the phenotypes recorded in tables 7.3 and 7.4.

Tables 7.1 and 7.2 represent the data from studying somatic excision of the marked *Tc*3 element from two experiments performed in tandem. The results in table 7.1, were obtained when the marked *Tc*3 element was present on the X chromosome and in table 7.2, the third chromosome.

The total number of flies in the F1 generation have been scored. The number of F1 females receiving the *rosy*+ allele is approximately 50%. This is because the mother was heterozygous for the *rosy*+ insertion. No males were observed to contain this allele. This is as would be expected for a line containing an insertion into an essential gene on the X chromosome.

Of the 425 (227 + 198) females containing both the marked *Tc*3 element and the Tc3 transposase, four were observed with yellow- wings, that were yellow+ on the remainder of their abdomens. This presumably indicates that the *yellow* marked *Tc*3 element has excised in the wing cells of these *Drosophila*, and may have taken place at an early developmental stage, since the whole wing exhibits a yellow- phenotype. No evidence of yellow-brown mosaic spots on the abdomen of any of the F1 generation of flies was observed.

The yellow- wing phenotype was only observed when the father from line *Tc*3*y*+*v*+ 18 was employed to provide the marked *Tc*3 element. In a control experiment, females from strain Tc3*ry*+ which contain the Tc3 ORF, were substituted for females from strain $y^1 v^{36f}$; ry^{506}, which are absent for *Tc*3 transposase, in the first cross shown in figure 7.5. A total of 372 female progeny were recovered, all of which were yellow+ in phenotype. This suggests that the four offspring obtained with a yellow- wing phenotype, in table 7.1 are as a result of the activity of *Tc*3 transposase, as the only difference between the crosses is in the presence or absence of the *Tc*3 transposase gene.

Tables 7.3 and 7.4 document the phenotypes of the F2 generation shown in figure 7.5. Again no males were found to contain the *rosy*+ allele, as expected, resulting in roughly fifty percent less males than females. Germline excision was monitored by

screening for yellow$^-$ vermilion$^+$ flies. No progeny of this phenotype were observed. Transposition was analysed by looking for yellow$^+$ vermilion$^-$ flies, no flies of this phenotype either were observed.

7.2.7 Detecting Tc3 transposase by Western blot analysis

The presence of functional transposase is a prerequisite for both the excision and transposition of *Tc*3. One possible explanation for the lack of mobility of the marked *Tc*3 element in the germline, is that the *Tc*3 ORF is not being expressed in this tissue. This has been checked by performing a Western blot against protein extracts from strain Tc3*ry*$^+$ using anti Tc3 transposase anti-sera.

Protein extracts were prepared from whole *Drosophila* strains; Tc3*ry*$^+$and y^1v^{36f}; ry^{506} and from ovaries obtained from strain Tc3*ry*$^+$ as described in 2.3.3.10. The protein extracts plus recombinant Tc3 transposase overexpressed in *E. coli* were fractionated on a 10% polyacrylamide gel before transfer to PVDF membrane. Tc3 transposase was detected by incubating the blot with anti-Tc3 transposase antibody, followed by POD-conjugated goat anti-rabbit secondary antibody.

Figure 7.7, shows the result of the Western blot experiment. Lanes 1 and 2 demonstrate that the anti Tc3 transposase anti-sera recognises the product of the Tc3 ORF when expressed in *E. coli*. However, no cross reacting bands are detected in lanes 3-6 which contain protein extracts from the strain of *Drosophila* containing the Tc3 ORF behind the hsp26:Sgs-3 cassette. This cassette contains two copies of the female germ-line specific transcriptional regulator of the *Drosophila* hsp26 gene upstream of the *Drosophila* Sgs-3 gene promoter (see Frank *et al*, 1992) and would be expected to drive transcription of the Tc3 ORF in the *Drosophila* female germline. This result indicates that if Tc3 transposase is being expressed at all in the female germline, it is at levels too low to detect by Western blot analysis and may be at concentrations insufficient to support the mobilisation of the marked *Tc*3 element.

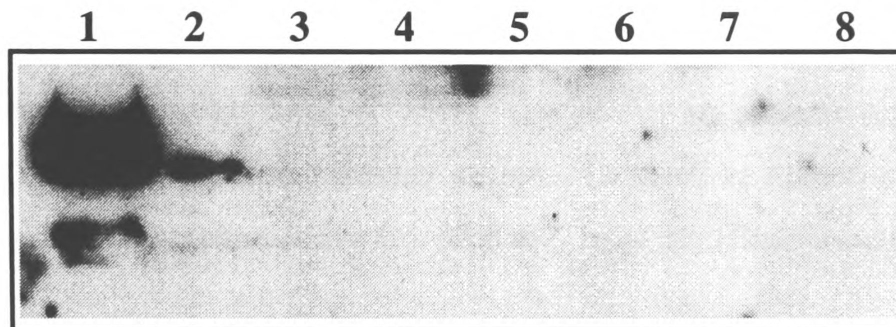


Figure 7.7. Reaction of anti- Tc3 transposase antiserum with extracts of Tc3*ry*$^{+}$ transformed flies.

Samples were fractionated on a 10% poly acrylamide gel prior to transfer to PVDF membrane. The anti-Tc3 transposase antibody was used at 1/1000 dilution.

Lanes 1 and 2 contain recombinant Tc3 transposase.

Lanes 3 and 4 contain whole *Drosophila* protein extract isolated from strain Tc3*ry*$^{+}$.

Lanes 5 and 6 contain whole ovary protein extract isolated from strain Tc3*ry*$^{+}$.

Lanes 7 and 8 contain whole *Drosophila* protein extract isolated from strain y^{1} v^{36f}, ry^{506}.

Lanes 3 and 7 contain protein extracted from 10 flies, and lane 5 from 16 ovaries. Lanes 2, 4, 6 and 8 are 10 fold dilutions of lanes 1, 3, 5 and 7 respectively.

7.3 Discussion

Plasmids pRP1406 and pRP1436 for studying the mobility of the *C. elegans* transposable element *Tc*3 in *Drosophila melanogaster* were constructed by Dr. Henri van Leunan.

Plasmid pRP1406 contains both the *Drosophila vermilion* allele, and a *Tc*3 element marked with the *yellow* gene from *Drosophila*, within *P* element terminal sequences. Plasmid pRP1436 contains the *Drosophila rosy* allele, plus the *Tc*3 ORF downstream of the *Drosophila* hsp26:Sgs-3 cassette within the terminal sequences of the *P* element. *Drosophila melanogaster* strain y^1v^{36f}; ry^{506}, in which any combination of functional copies of the three alleles; *yellow*, *vermilion* and *rosy* can be detected phenotypically, was constructed by Dr. Marie-Christine Chabboissier and myself.

The two *Tc*3 based constructs were successfully introduced into the y^1v^{36f}; ry^{506} strain of *Drosophila* by *P* element mediated germline transformation as determined by phenotype analysis and Southern blotting. Lines *Tc*3y^+v^+ 18, 25, 95 and 96 were obtained containing the *yellow* marked *Tc*3 element and line *Tc*3ry^+ comprising the *Tc*3 transposase gene. Crosses of the two different strains of flies were performed to bring the marked *Tc*3 element into the same genetic background as the *Tc*3 transposase gene, such that mobilisation of the marked *Tc*3 element could be examined.

Excision of *Tc*3 in the soma was scored in a total of 425 flies containing both the marked *Tc*3 element and the *Tc*3 ORF. Four flies were detected in which wings were yellow$^-$ in phenotype. In a control experiment in which the source of Tc3 transposase was absent from the flies, no yellow$^-$ phenotypes were observed. This may indicate that the loss of the *yellow* allele is coincident with the presence of the Tc3 ORF gene and is presumably due to the action of Tc3 transposase supplied *in trans*.

In the four flies exhibiting a *yellow*$^-$ wing phenotype, excision of *Tc*3 is from the X chromosome, since the male parent was from strain Tc3y^+v^+ 18. No excision events from chromosome III were detected. On both the X and the third chromosomes, the marked *Tc*3 elements are present in the same primary sequence environment since each is flanked by the *unc*22 gene sequence, within the P element termini. Excision of *Tc*3 from the *unc*22 gene has previously been documented (van Leunan *et al*, 1994). However, it is possible that the functional state of the DNA around the *Tc*3 locus is different on each of the two chromosomes. This may be reflected in the accessibility of the transposase to the element and result in the difference in excision frequency observed.

The recovery of *Drosophila* containing somatic excision events is somewhat surprising in view of the tissue specificity documented for the hsp26:Sgs-3 cassette. The hsp26:Sgs-3 cassette contains the Sgs3 promoter linked to two copies of the 171bp hsp26 transcriptional regulatory sequences and has previously been shown to result in expression of a reporter gene in cells of the female germline only (Frank *et al*, 1992). If Tc3 transposase is expressed somatically, however, as suggested by the somatic excision events observed, it is unusual that it can not be detected in Western blot experiments on whole *Drosophila* protein extracts using anti Tc3 anti-sera. This raises the possibility that the observed yellow$^-$ wing phenotype does not reflect Tc3 transposase induced excision of the marked Tc3 element, but is due to an unknown factor.

The marked *Tc*3 element was found to be immobile in the germline. Expression of the Tc3 ORF was predicted to take place within the germline, due to the tissue specificity of the hsp26:Sgs-3 cassette as discussed previously. However, the failure to detect transposase by Western blot analysis using anti Tc3 antisera, suggests that expression of the transposase has not taken place in the germline in this instance. It is therefore difficult to draw any conclusions regarding the mobility of *Tc*3 in *Drosophila* in the germline.

To draw a satisfactory conclusion on the activity of *Tc*3 in *Drosophila* it will be neccesary to construct more lines of the Tc3*ry*$^+$ strain, in which flies contain the Tc3 ORF at different chromosomal locations. It will also be important to demonstrate the presence of expressed transposase in these lines, since transposase is a prerequisite for both excision and transpositional activity.

Chapter 8

Summary and Future Work

8.1 Summary and Future Work

Mariner elements have been assumed to transpose via a 'cut and paste' mechanism, in which the element is physically removed from the chromosome before insertion elsewhere. Several other short inverted repeat type elements, such as the *Tc*1 and *Tc*3 elements from *C. elegans* and the *P element* from *D. melanogaster* transpose by this mechanism.

Excised forms of *mariner* have been detected *in vivo* from *Drosophila* DNA preparations, as discussed in Chapter 5, lending support to the excision/insertion mechanism of transposition for these elements. Moreover, as described in Chapter 4, addition of recombinant Mos1 transposase to plasmids containing the *mos*1 DNA sequence results in the appearance of linear, excised copies of *mos*1.

The sequences of circular, extrachromosomal copies of *mariner* detected *in vivo*, together with 'footprint' data, previously reported (Bryan *et al*, 1990; Coates *et al*, 1995), suggest that the active intermediate in transposition of *mos*1 and other *mariner* elements in *Drosophila* contains a three base pair overhang of *mariner* sequence at each transposon end. The transpositional intermediates of related transposons from the hornfly and from nematodes also contain staggered ends (van Leunan *et al*, 1994; Vos *et al*, 1996; Lampe *et al*, 1996).

Linear copies of *mos*1 obtained *in vitro* are unlikely to be competent to reintegrate. However, they are identical in sequence to the predicted *mos*1 intermediate, on one strand at each end of the element and are therefore probably not entirely artifactual in structure. The cleavage of *mos*1 *in vitro* can be checked by incubating Mos1 transposase with short, radiolabelled fragments of DNA containing the *mos*1 inverted repeat, as described for *Himar*1, in Lampe *et al*, 1996. This method also provides a convenient means of determining the order of strand cleavage.

Mariner elements are capable of horizontal transfer between genomes. This has been taken to indicate that *mariner* transposition does not require species specific host factors beyond those necessary to transcribe and translate the transposase and repair the ends of the element after transposition. Indeed, the detection of linear copies of *mos*1 *in vitro* demonstrates that excision at least, in which four strand cleavage events are required to release the element from flanking DNA, can occur independently of *Drosophila* factors.

Mariner elements are attractive candidates for the development of gene delivery vectors. Recombinant Mos1 transposase alone is sufficient to transform the germline of *D. melanogaster*, as demonstrated in Chapter 6, following co-injection with a marked *mos*1 element. This method eliminates the need to transiently express the transposase gene in embryos and could circumvent potential problems if *mariner* elements were used as a transformation system in organisms where molecular biology is in its infancy. For the development of *mariner* as a universal germline transformation vector, the green fluorescent protein may prove to be a useful marker gene.

Future work to be continued on this project includes analysing the mechanism of transposition in further detail. The presence of 3' hydroxyl groups on the transposition intermediates of the elements *Tc*1 and *Tc*3 have been implicated to be essential for the strand transfer events involved in element insertion. The presence of these groups has not yet been looked for in *mariner* intermediates.

Little is known of the stoichiometry of transposase to transposon ends in *mariner* transposition. This issue may be approached by a combination of glutaraldehyde and UV cross linking techniques, and may also yield information regarding the regulation of *mariner* activity.

A number of essential residues have previously been identified in Mos1 transposase using a genetical method. A further task is to determine at which point in transposition these residues are vital, for example, in DNA binding, in excision, or in strand transfer activity. Ultimately, the resolution of the crystal structure of the protein, will allow the mechanism to be studied in the greatest detail.

Bibliography

Adams, M.D., Tang, R.S. and Rio, D.C. (1997). The alternative splicing factor PSI regulates *P* element third intron splicing in vivo. Gen. and Dev. **11**, 129-138.

Alam, J., Vrba, J.M., Cai, Y., Martin, J.A., Weislo, L.J. and Curtis, S.E. (1991). Characterisation of the IS895 family of Insertion Sequences from the Cyanobacterium, Anabaena sp. Strain PCC7120J. Bac. **173**, 5778-5783.

Arca, B., Zabalou, S., Loukeris, T.G. and Savakis, C. (1997). Mobilisation of a *minos* transposon in *Drosophila melanogaster* chromosomes and chromatid repair by heteroduplex formation. Genetics. **145**, 267-279.

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**, 587-594.

Ashburner, M. (1989). *Drosophila*: A laboratory handbook. Cold Spring Harbour Laboratory Press. Cold Spring Harbour, N.Y. pp549-574.

Ashburner, M. (1995). Through the glass lightly. Science. **267**, 1609-1610.

Atkinson, P.W., Warren, W.D and O'Brochta, D.O. (1993). The *hobo* transposable element of *Drosophila* can be cross-mobilised in house flies and excises like the Ac element of maize. Proc. Natl. Acad. Sci. U.S.A. **90**, 9693-9697

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 Letts. **368**, 541-546

Bainton, R., Gamas, P. and Craig, N. (1991). Tn7 transposition *in vitro* proceeds through an excised transposon intermediate generated by staggered breaks in DNA. Cell. **65**, 815-816.

Bainton, R., Kubo, K.M., Feng, J.N. and Craig, N. (1993). Tn7 transposition: target DNA recognition is mediated by multiple Tn7 encoded proteins in a purified *in vitro* system. Cell. **72**, 931-943.

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**, 6654-6658.

Baker, T.A. and Mizuuchi, K. (1992). DNA-promoted assenmbly of the active tetramer of the Mu transposase. Genes Dev. **6**, 2221-2232.

Beall, E.L., Adman, 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**, 12681-12685.

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

Bender, J. and Kleckner, N. (1992). Tn10 insertion specificity is strongly dependent upon sequences immediately adjacant to the target site consensus sequence. Proc. Natl. Acad. Sci. U.S.A. **89**, 7996-8000.

Benjamin, H.W. and Kleckner, N. (1989). Intramolecular transposition by Tn10. Cell. **59**, 373-383.

Benjamin, H.W. and Kleckner, N. (1992). Excision of Tn*10* from the donor site during transposition occurs by flush double-strand cleavages at the transposon termini. Proc. Natl. Acad. Sci. U.S.A. **89**, 4648-4652.

Benton, W. and Davis, R. (1977). Screening λgt recombinant clones by hybridisation to single plaques *in situ*. Science. **196**, 180-182.

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

Berg, D.E., Schmandt, M.A. and Lowe, J.B. (1983). Specificity of Transposon Tn5 Insertion. Genetics. **105,** 813-828.

Bingham, P.M., Levis, R. and Rubin, G.M. (1981). Cloning of DNA sequences from the *white* locus of *Drosophila melanogaster* by a novel and general method. Cell. **25**, 693-704.

Birnhoim, H.C. and Doly, J. (1979). A rapid method for screening recombinant plasmid DNA. Nucl. Acids. Res. **7**, 1513-1523.

Blackman, R.K. and Gelbart, W.M. (1989). The Transposable Element *hobo* of *Drosophila melanogaster*. In Mobile DNA, ed. by D.E. Berg and M. Howe. American society for Microbiology, Washington D.C. p 523-529.

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

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**, 7814-7818.

Bolland, S. and Kleckner, N. (1996). The three chemical steps of Tn10/IS10 transposition involve repeated utilisation of a single active site. Cell. **84**, 223-233.

Brezinsky, L., Wang, G.V.L., Humphreys, T. and Hunt, J. (1990). The transposable element *Uhu* from Hawaiian *Drosophila*, a member of the widely dispersed class of *Tc*1-like transposons. Nucl. Acids. Res. **18**, 2053-2059.

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**, 669-675.

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

Bryan, G. J. and Hartl, D.L. (1988). Maternally inherited transposon excision in *Drosophila simulans*. Science. **240**, 215-217.

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

Bujacz, G., Jaskolski, M., Alwxandratos, J., Wlodover, 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. **235**, 333.

Calvi, B.R., Hong, T.J., Findlay, S.D. and Gelbert, W.M. (1991). Evidence for a Common Evolutionary Origin of Inverted Repeat Transposons in *Drosophila* and Plants: hobo, Activator, and Tam3. Cell. **66**, 465-471.

Calvi, B.R. and Gelbart, W.M. (1994). The basis for germline specificity of the *hobo* transposable element in *Drosophila melanogaster.* EMBO J. **13**, 1636-1644.

Caparon, M.G. and Scott, J.R. (1989). Excision and insertion of the conjugative transposon Tn916 involves a novel recombination mechanism. Cell. **59**, 1027-1034.

Capy, P., Chakrani, F., Lemeunier, F., Hartl, D.L. and David, J.R. (1990). Active *mariner* transposable elements are widespread in natural populations of *Drosophila simulans*. Proc. Royal. Soc. Lond. B. **242**, 57-60.

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**, 450-456.

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.

Casadesus, J. and Roth, J.R. (1989). Transcriptional occlusion of transposon targets. Mol. Gen. Genet. **216**, 204-209.

Chalmers, R.M. and Kleckner, N. (1994). Tn10/IS10 transposase purification, activation and *in vitro* reaction. J. Biol. Chem. **269**, 8029-8035.

Clark, J.B., Madison, W.P. and Kidwell, M.G. (1994). Phylogenetic analysis supports horizontal transfer of *P* transposable elements. Mol. Biol. Evol. **11**, 40-50.

Clark, J.M. (1988). Novel non-templated nucleotide addition catalysed by procaryotic and eucaryotic DNA polymerases. Nucl. Acids. Res. **16**, 9677-9686.

Coates, C.J., Turney, C.L., Frommer, M., O'Brochta, D.A. and Atkinson, P. (1995). The transposable element *mariner* can excise in non-drosophilid insects. Mol. Gen. Genet. **249**, 246- 252.

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

Collins, F.H. (1994). Prospects for malaria control through the genetic manipulation of its vector. Parasitol. Today. **10,** 370.

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

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

Craig, N.L. (1991). Tn7- a target site specific transposon. Mol. Microbiol. **5**, 2569-2573.

Craig, N.L. (1995). Unity in transposition reactions. Science. **270**, 253-254.

Craigie, R., Mizuuchi, M. and Mizuuchi, K. (1984). Site Specific Recognition of the Bacteriophage Mu Ends by the MuA Protein. Cell. **39**, 387-394.

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

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.

Craigie, R. and Mizuuchi, K. (1995). Mechanism of transposition of bacteriophage Mu: structure of a transposition intermediate. Cell. **41**, 867-876.

Curtis, C.F. (1994). The case for malaria control by genetic manipulation of its vectors. Parasitol. Today. **10,** 371-374.

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

Daniels, S.B. and Strausbough, L.D. (1986). The distribution of *P* element sequences in *Drosophila* : the *willistoni* and *saltans* species groups. J.Mol. Evol. **23**, 138-148.

Daniels, S.B., Peterson, K.R, Strausbough, L.D., Kidwell, M.G. and Chovnick, A.C. (1990)a. Evidence for horizontal transmission of the *P* transposable element between *Drosophila* species. Genetics. **124**, 339-355.

Daniels, S.B., Chovnik, A. and Boussey, I.A. (1990)b. Distribution of *hobo* transposable elements in the genus *Drosophila.* Mol. Biol. Evol. **7**, 589-606.

Datta, A.R. and Rosner, J.L. (1987). Reduced Transposition in *rho* mutants of *Escherichia coli* K-12. J. Bacteriol. **169**, 888-890

de la Cruz, N.B., Weinrich, M.D., Wiegand, T.W., Krebs, M.P. and Reznikoff, W.S. (1993). Characterisation of the Tn5 transposase and inhibitor proteins: a model for the inhibition of transposition. J. Bacteriol. **175**, 6932-6938.

Deninger, P..L. (1989). Short Interspersed Repeated DNA Elements in Higher Eukaryotes. In Mobile DNA, ed. by D.E. Berg and M. Howe. American society for Microbiology, Washington D.C. p 619-636.

Derbyshire, K.M. and Grindley, N.D.F. (1992). Binding of the *IS903* transposase to its inverted repeat *in vitro.* **11**, 3449-3455.

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**, 942-946.

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. Science. **266**, 1981-1985.

Eichinger, D.J. and Boeke, J.D. (1990). A specific terminal structure is required for Ty1 transposition. Genes. Dev. **4,** 324-330.

Eide, D. and Anderson, P. (1988). Insertion and excision of the *C. elegans* transposable element *Tc1*. Mol. Cell. Biol. **8**, 737-746.

Ellinger, J.S., Mach, M., Korm, K. and Jahn, G. (1991). Cleavage and purification of procaryotically expressed HIV gag and env fusion proteins for detection of HIV antibodies in the ELISA. Virology. **180**, 811-813.

Emmons, S.W. and Yesner, L. (1984). High frequency excision of transposable element *Tc1* in the nematode *Caenorhabditis elegans* is limited to somatic cells. Cell. **36**, 599-605.

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

Engelman, A. and Craigie, R, (1992). Identification of Conserved Amino Acid Residues Critical for Human Immnuodeficiency Virus Type I Integrase Function *In Vitro*. J. Virol. **66**, 6361-6369.

Engels, W.R. (1984). A *trans* acting product needed for *P* factor transposition in *Drosophila*. Science. **226**, 1194-1196.

Engels, W.R. (1989). *P* elements in *Drosophila melanogaster.* In Mobile DNA, ed. by D.E. Berg and M. Howe. American society for Microbiology, Washington D.C. p 437-484.

Engels. W.R., Johnson-Schlitz, D.M., Eggleston, W.B. and Sved, J. (1990). High frequency *P* element loss in *Drosophila* is homologue dependent. Cell. **62**, 515-525.

Evans, J.P. and Palmiter, R.D. (1991). Retrotransposition of a mouse *L1* element. Proc. Natl. Acad. Sci. U.S.A. **88**, 8792-8795.

Federoff, N.V., Furtek, D.B. and Nelson, O.E. (1984). Cloning of the *bronze* locus in maize by a simple and generalizable procedure using the transposable controlling element *Activator* (*Ac*). Proc. Natl. Acad. Sci. U.S.A. **81**, 3825-3829.

Federoff, N.V. (1989). Maize transposable elements. In Mobile DNA, ed. by D.E. Berg and M. Howe. American society for Microbiology, Washington D.C. p 375-412.

Feng, Q., Moran, J.V., Kazazian, H,.H. and Boeke, J.D. (1996). Human L1 Retrotransposon Encodes a Conserved Endonuclease Required for Retrotransposition. Cell. **87**, 917-927.

Finnegan, D.J. (1992). Transposable elements. Curr. Opin. Genet. Dev. **2**, 861-867.

Frank, L.H., Cheung, H-K. and Cohen, R.S. (1992). Identification and characterisation of *Drosophila* female germline transcriptional control elements. Development. **114**, 481-491.

Franz, G. and Savakis, C. (1991). *Minos*, a new transposable element from *Drosophila hydei*, is a member of the *Tc*1-like family of transposons. Nucl. Acids. Res. **19**, 6646.

Franz, G., Loukeris, T.G., Dialektaki, G., Thompson, C.R.L. and Savakis, S. (1994). Mobile *minos* elements from *Drosophila hydei* encode a two-exon transposase with similarity to the paired DNA binding domain. Proc. Natl. Acad. Sci. U.S.A. **91**, 4746-4750.

Friedel and Searles (1991). *Vermillion* as a small selectable marker gene for *Drosophila* transformation. Nucl. Acids. Res. **19**, 5082-

Garcia-Ferandez, J., Baguna, J. and Salo, E. (1991). Planarian homeobox genes: Cloning, sequence analysis, and expression. Proc. Natl. Acad. Sci. U.S.A. **88**, 7338-7342.

Garcia-Ferandez, J., Marfany, G., Baguna, J. and Salo, E. (1993). Infiltration of *mariner* elements. Nature. **364**, 109.

Garcia-Ferandez, J., Bayascas-Ramirez, J.R., Marfany, G., Munaz-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**, 421-431.

Garza, D.M., Medhora, M., Koga, A. and Hartl, D.L. (1991). Introduction of the transposable element *mariner* into the germline of *Drosophila melanogaster.* Genetics. **128**, 303-310.

Gellert, M. (1996). A new view of V(D)J recombination. Genes to Cells. **1**, 269-276.

Geyer, P.K. and Corces, V.G. (1987). Separate regulatory elements are responsible for the complex pattern of tissue specific and developmental transcription of the *yellow* locus in *Drosophila melanogaster.* Genes and Dev. **1**, 996-1004.

Gierl, A., Lutticke, S. and Saedler, H. (1988). *TnpA* product encoded by the transposable element En-1 of *Zea mays* is a DNA binding protein. EMBO. **7**, 4045-4053.

Gough, J.A. and Murray, N.M. (1983). Sequence diversity among related genes for recognition of specific targets in DNA molecules. J. Mol. Biol. **166**, 1-19.

Grindley, N.D.F. and Leschziner, A.E. (1995). DNA transposition: From a Black Box to a Colour Moniter. Cell. **83**, 1063-1066.

Hagemann, S., Miller, W.J. and Pinster, W. (1992). Identification of a complete *P* element in the genome of *Drosophila bisfasciata*. Nucl. Acids Res. **20**, 409-413.

Halling, S.M. and Kleckner, N. (1982). A Symmetrical Six-Base-Pair Target Site Sequence Determines Tn10 Insertion Specificity. Cell. **28**, 155-163

Haniford, D. B., Benjamin, H.W. and Kleckner, N. (1991). Kinetic and structural analysis of a cleaved donor intermediate and strand transfer product in Tn10 transposition. Cell. **64**, 171-179.

Haniford, D. B. and Kleckner, N. (1994). Tn10 transposition *in vivo*: temporal separation of cleavages at the two transposon ends and roles of terminal base pairs subsequent to interaction of ends. EMBO J. **13**, 3401-3411.

Harris, L.J., Baillie, D.L. and Rose, A.M. (1988). Sequence identity between an inverted repeat family of transposable elements in *Drosophila* and *Caenorhabditis*. Nucleic. Acids. Res. **16**, 5991-5998.

Hartl, D. (1989). The Transposable Element *mariner* in *Drosophila* species. In Mobile DNA, ed. by D.E. Berg and M. Howe. American society for Microbiology, Washington D.C. pp 531-536.

Hartl, D, Lozovskaya, E.R., Nurminsky, D.I. and Lohe, A.R. (1997). What restricts the activity of *mariner* like transposable elements. TIGS. **13**, 197-201.

Haymer, D.S. and Marsh, J.L. (1986). Germline and somatic instability of a *white* mutation in *Drosophila mauritiana* due to a transposable element. Dev. Genet. **6**, 281-291.

Hazelrigg, T. Levis, R. and Rubin, G. (1984). Transformation of *white* locus DNA in *Drosophila*: dosage compensation, *zeste* interaction, and position effects. Cell **36**, 469-481.

Henikoff, S. and Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. U.S.A. **89**, 10915-10919.

Huisman, O., Errada, P.E., Signon, L. and Kleckner, N. (1989). Mutational analysis of *IS10* 's outside end. EMBO J. **8**, 2101-2109.

Isberg, R.R. and Syvenan, M. (1982). DNA Gyrase is a Host Factor Required for Transposition of Tn5. Cell. **30**, 9-18.

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**, 8684-8688.

Jarakzawski, J.W. and Jahn, C.L. (1993). Elimination of Tec elements involves a novel excision process. Genes and Dev. **7,** 95-105.

Jensen, S. and Heidmann, T. (1991). An indicator gene for detection of germline retrotransposition in transgenic *Drosophila* demonstrates RNA mediated transposition of the LINE *I* element. EMBO. **10**, 1927-1991.

Ji, H., Moore, D.P., Blomberg, M.A., Braiterman, L.T., Voytas, D.F., Natsoulis, G. and Boeke, J.D. (1983). Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences. Cell. **73**, 1007-1018.

Karess, R. E. and Rubin, G. M. (1984). Analysis of *P* transposable element functions in *Drosophila.* Cell. **38**, 135-146.

Katz, R.A. and Skalka, A.M. (1994). The retroviral enzymes. Annu. Rev. Biochem. **63**, 133-173.

Kaufman, P.D., Doll, R.F. and Rio, D.C. (1989). *Drosophila P* element transposase recognises internal *P* element DNA sequences. Cell. **59**, 359-371.

Kaufman, P.D. and Rio, D.C. (1991). Germline transformation of *Drosophila melanogaster* by purified *P* element transposase. Nucl. Acids. Res. **19**, 6336.

Kelley, M.R., Kidd, S., Berg, R.L. and Young, M.W. (1987). Restriction of *P* Element Insertions at the *Notch* Locus of *Drosophila melanogaster.* Mol. Cell. Biol. **7**, 1545-1548.

Kidwell, M. (1983). Evolution of Hybrid Dysgenesis determinants in *Drosophila melanogaster.* Proc. Natl. Acad. Sci. U.S.A. **80**, 1655-1659.

Kidwell, M. (1982). Horizontal transfer. Curr. Op. in Genetics and Dev. **2**, 868-873.

Kim, A., Terzian, C., Santamaria, P., Pelisson, A., Prud'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**, 1285-1289.

Kim, K., Namgoong, S.Y., Jayaram, A. and Harshey, R.M. (1995). Step-arrest Mutants of Phage Mu Transposase. J. Biol. Chem. **270**, 1472-1479.

Kleckner, N. (1989) Transposon Tn10. In Mobile DNA, ed. by D.E. Berg and M. Howe. American society for Microbiology, Washington D.C. p 335-374.

Kleckner, N., Chalmers, R., Kwon, D., Sakai, J. and Bolland, S. (1995). Tn10 and IS10 transposition and chromosome rearrangements: mechanism and regulation *in vivo* and *in vitro*. Curr. Top. Microbiol. Immunol. **204**, 49-82.

Klemenz, R., Weber, U. and Gehring, W.J. (1987). The *white* gene as a marker in a new *P* element vector for gene transfer in *Drosophila*. Nucl. Acids. Res. **15**, 3947-3957.

Kulkosky, J., Jones, K.S., Katz, R.A., Mack, J.P.G. 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 Transposons. Mol. Cell. Biol. **12**, 2331-2338.

Kunze, R. and Starlinger, P. (1989). The putative transposase of transposable element *Ac* from *Zea mays* interacts with subterminal sequences of *Ac*. EMBO. **8**, 3177-3185.

Kunze, R., Behrens, U., Courage-Franzkowiak, U., Felmar, S. and Kuhn, S. (1993). Dominant transposition-deficient mutants of maize Activator (Ac) transposase. Proc. Natl. Acad. Sci. U.S.A. **90**, 7094-7098.

Lachaise, D., Kariou, M., David, J.R., Lemuenier, F., Tsacas, L. and Ashburner, M. (1988). Historical biogeography of the *Drosophila melanogaster* species subgroup. Evol. Biol. **22**, 159-227.

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

Langin, T., Capy, P. and Dabouss, M.J. (1995). The transposable element *impala*, a fungal member of the *Tc*1-*mariner* superfamily. Mol. Gen. Genet. **46**, 19-28.

Lansman, R.A., Stacey, S.N., Grigliatti, T.A. and Brock, H.W. (1985). Sequences homologous to the *P* mobile element of *Drosophila melanogaster* are widely distributed in the subgenus Sophophora. Nature. **318**, 561-563.

Laski, F., Rio, D.C. and Rubin, G. (1986). Tissue specifity of *Drosophila P* element transposition is regulated at the level of mRNA splicing. Cell, **44**, 7-19.

Lavoie, B.D., Chan, B.S., Allison, R.G. and Chaconas, G. (1991). Structural Aspects of a Higher-Order Nucleoprotein Complex- Induction of an Altered DNA-Structure at the Mu-Host Junction of the Mu-type I Transposasome. EMBO. J. **10**, 3051-3059.

Lawrence, J.G. and Hartl, D.L. (1992). Inference of horizontal genetic transfer from molecular data: an approach using the bootstrap. Genetics. **131**, 753-760.

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. and Cell. Biol. **16**, 5616-5622.

Lemaitre, B. and Coen, D. (1991). *P* regulatory products repress *in vivo* the *P* promoter activity in *P*-lacZ fusuion genes. Proc. Natl. Acad. Sci. U.S.A. **88**, 4419-4423.

Lemaitre, B., Ronsseray, S. and Coen, D. (1993). Maternal repression of the *P* element promoter in the germline of *Drosophila melanogaster*: a model for *P* cytotype. Genetics. **135**, 149-160.

Leung, P.C., Teplow, D.B. and Harshey, R.M. (1989). Interaction of distinct domains in Mu transposase with Mu DNA ends and an internal transpositional enhancer. Nature. **338**, 656-658.

Lidholm, D., Gudmundsson, G.H. and Boman, H.G. (1991). A highly repetitive, *mariner* like element in the genome of *Hyalophora cecropia.* J. Biol. Chem. **226,** 11518-11521.

Lidholm, D., Lohe, A.R. and Hartl, D.L. (1993). The Transposable Element *mariner* mediates Germline Transfomation in *Drosophila melanogaster.* Genetics. **134**, 859-868.

Lindsley, D. and Grell, E. (1968). Genetic variations of *Drosophila melanogaster.* Carnegie institue of Washington, Publication no. 627.

Lindsley, D. and Zimm, G.G. (1992). The genome of *Drosophila melanogaster.* Academic press, inc. Harcourt Brace Jovanovich, Publishers.

Lohe, A.R., Moriyama, E.N. Lidholm, D. and Hartl, D. (1995). Horizontal Transmission, Vertical Inactivation, and Stochastic loss of *Mariner*-Like Transposable Elements. Mol. Biol. Evol. **12**, 62-72.

Lohe, A.R and Hartl, D.L. (1996). Autoregulation of *mariner* Transposase Activity by Overproduction and Dominant-Negative Complementation. Mol. Biol. Evol. **13**, 549-555.

Lohe, A.R and Hartl, D.L. (1996). Reduced germline Mobility of a *mariner* Vector containing Exogenous DNA: Effect of size or site? Genetics. **143**, 1299-1306.

Lohe, A.R., Sullivan, D.T. and Hartl, D.L. (1996). Subunit Interactions in the *mariner* Transposase. Genetics. **144**, 1087-1095

Lohe, A.R., de Aguiar, D. and Hartl, D.L. (1997). Mutations in the *mariner* transposase: D,D(35)E sonsensus sequence are nonfunctional. Proc. Natl. Acad. Sci. U.S.A. **94**, 1293-1297.

Loukeris, T.G., Arca, B., Lividaris, I., Dialektaki, G. and Savakis, C. (1995)a. Introduction of the transposable element *Minos* into the germline of *Drosophila melanogaster.* Genetics. **134**, 859-868.

Loukeris, T.G., Lividaris, I., Arca, B., Zabalou, S. and Savakis, C. (1995)b. Gene transfer into the Medfly, *Ceratitis capitata*, using a *Drosophila hydei* transposable element. Science. **270**, 2002-2005.

Maruyama, K. and Hartl, D.L. (1991a). Evolution of the Transposable Element *mariner* in *Drosophila* species. Genetics. **128**, 319-329.

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

Maruyama, K., Schoor, K.D. and Hartl, D.L. (1991). Identification of Nucleotide Substitutions Neccessary for *Trans*-activation of *mariner* transposable elements in *Drosophila*: Analysis of Naturally Occuring Elements. Genetics. **128**, 777-784.

McBlane, J.F., van Gent, D.C., Ramsden, D.A., Romeo, C., Cuomo, C.A., Gellert, M. and Oettinger, M.A. (1995). Cleavage at a V(D)J recombination signal requires only Rag1 and Rag1 proteins and occurs in two steps. Cell. **83**, 387-395.

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

McKinnon, R.D., Waye, J.S., Bautista, D.S. and Graham, F.L. (1985). Non-random insertion of Tn5 into cloned human adenovirus DNA. Gene. **40**, 31-38.

Medhora, M.M., Macpeek, A.H. and Hartl, D.L. (1988). Excision of the *Drosophila* transposable element *mariner*: Identification and characterisation of the *mos* factor. EMBO.**7**, 2185-2189.

Medhora, M.M., Maruyama, K. and Hartl, D.L. (1991). Molecular and Functional analysis of the *mariner* Mutator Element *Mos1* in *Drosophila.* Genetics. **128**, 311-318.

Misra, S. and Rio, D.C. (1990). Cytotype Control of *Drosophila P* Element Transposition: The 66kD Protein is a Repressor of Transposase Activity. Cell. **62**, 269-284.

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

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

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

Moermann, D.G and Waterston, R.H. (1989). Mobile Elements in *C. elegans* and Other Nematodes. In Mobile DNA, ed. by D.E. Berg and M. Howe. American society for Microbiology, Washington D.C. p 537-556.

Moore, S.P. and Garfinkel, D.J. (1994). Expression and partial purification of enzymatically active recombinant integrase in *S. cerevisiae*. Proc. Natl. Acad. Sci. U.S.A. **91**, 1843-1847.

Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardiris, R.J., Boeke, J.D. and Kazazian, H.H. (1996). High frequency Retrotransposition in Cultured Mammalian Cells. Cell. **87**, 917-927.

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

Mori, I., Moermann, D.G. and Waterston, R.H. (1988). Analysis of a mutator activity necessary for germline transposition and excision of *Tc1* transposable elements in *Caenorhabditis elegans*. Genetics. **120**, 397-407.

Morisato, D. and Kleckner, N. (1984). Transposase promotes double strand breaks amd single strand joints at Tn10 termini *in vivo*. Cell. **39**, 181-190.

Mullins, M.C., Rio, D.C. and Rubin, G.M. (1989). Cis-acting DNA sequence requirements for *P* element transposition. Genes Dev. **3**, 729-738.

Nyyssonen, E., Amutan, M., Enfield, L., Stubbs, J. and Dunn-Coleman, N.S. (1996). The transposable element *Tan1* of *Aspergillus niger* var. *awamori*, a new member of the *Fot1* family. Mol. Gen. Genet. **253**, 50-56.

O'Brochta, D.A., Warren, W.D., Saville, K.J. and Atkinson, P.W. (1994). Interplasmid transposition of *Drosophila hobo* elements in non-drosophilid insects. Mol. Gen. Genet. **244**, 9-14.

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

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

O'Kane, C.J. and Gehring, W.J. (1987). Detection *in situ* of genomic regulatory elements in *Drosophila*. Proc. Natl. Acad. Sci. U.S.A. **84**, 9123-9127.

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

Pabo, C.O. and Sauer, R.T. (1992). Transcription factors: structural families and principles of DNA recognition. Annu. Rev. Biochem. **61**, 1053-1095.

Pato, M.L. (1989). in Mobile DNA, eds. Berg, D.E. and Howe,M.M. American society for Microbiology, Washington, D.C., p 23-25.

Pelisson, A., Finnegan, D.J. and Bucheton, A. (1991). Evidence for Retrotransposition of the *I* factor, a LINE like element of *D. melanogaster* . Proc. Natl. Acad. Sci. U.S.A. **88**, 4907-4910.

Pietrovski, S. and Hennikoff, S. (1997). A helix-turn-helix DNA binding motif predicted for Tc1, mariner and pogo family transposases. in press.

Plasterk, R.H. (1987). Differences between *Tc1* elements from the *C. elegans* strain Bergerac. Nucl. Acids. Res. **15**, 10050.

Plasterk, R.H. (1991). The origin of footprints of the *Tc1* transposon of *Caenorhabditis elegans*. EMBO. **10**, 1919-1925.

Plasterk, R.H. and Groenen, J.T.M. (1992). Targeted Alterations of the *Caenorhabditis elegans* genome by transgene instructed DNA double strand break repair following Tc1 excision. EMBO J. **11**, 287-290.

Polard, P., Prere, M.F., Fayet, O. and Chandler, M. (1992). Transposase induced excision and circularisation of the bacterial insertion sequence IS911. EMBO. J. **11**, 5079-5090.

Presnail, J.K. and Hoy, M.A. (1992). Stable genetic transformation of a beneficial arthropod: Metaseiulus occidentalis, (Acai:Phytoseiidee), by a microinjection technique. Proc. Natl. Acad. Sci. U.S.A. **89**, 7732-7736.

Radice, A.D. and Emmons, S.W. (1993). Extrachromosomal circular copies of the transposon *Tc*1. Nucl. Acids. Res. **21**, 2663-2667.

Reznikoff, W.S. (1993). The Tn5 transposon. Annu. Rev. Microbiol. **47**, 945-963.

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**, 209-220.

Rio, D.C., Laski, F.A. and Rubin, G.M. (1986). Identification and Immunochemical analysis of biologically active P element transposase. Cell. **44,** 21-32.

Roberts, D. and Kleckner, N. (1988). Tn10 transposition promotes RecA-dependent induction of a lambda prophage. Proc. Natl. Acad. Sci. USA. **85**, 6037-6041.

Robertson, H., Lampe, D.J. and MacLeod, E.G. (1992). A *mariner* transposable element from lacewing. Nucl. Acid. Res. **20**, 6409.

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

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

Robertson, H. and Lampe, D.J. (1995). Distribution of transposable elements in arthropods. Annu. Rev. Entomol. **41**, 333-357.

Robertson, H. and Asplund (1996). Bmmar1: a basal lineage of the *mariner* family of transposable elements in the silkworm moth, *Bombyx mori*. Insect Biochem. mol. biol. **26**, 945-954.

Robinson, P.A., Anderson, P.H. and Loving, T.L.F (1988). Nitrocellulose-bound antigen repeatedly used for the affinity purification of specific ploycional antibodies for screening DNA expression libraries. J. Immunol. Methods. **108**, 115-122.

Ronsseray, S. Lemaitre, B. and Coen, D. (1993). Maternal inheritance of P cytotype in *Drosophila melanogaster:* a 'pre-P cytotype' is strictly extra-chromosomally transmitted. Mol. Gen. Genet. **241**, 115-123.

Rose, A.M. and Snutch, T.P. (1984). Isolation of the closed circular form of the transpoable element *Tc*1 in *Caenorhabditis elegans*. Nature. **311**, 485-486.

Rosenzweig, B., Liao, L.W. and Hirsch, D. (1983). Sequence of the *C. elegans* transposable element *Tc*1. Nucl. Acid. Res. **12**, 4201-4209.

Ruan, K.S. and Emmons, S.W. (1984). Extrachromosomal copies of the transposon Tc*1* in the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. U.S.A. **81**, 4018-4022.

Ruan, K.S. and Emmons, S.W. (1987). Precise and imprecise somatic excision of the transposon *Tc1* in the nematode *C. elegans.* Nucl. Acids. Res. **15**, 6875-6881.

Rubin, G. M., Kidwell, M.G. and Bingham, P.M. (1982). The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell. **29**, 987-994.

Rubin, G. M. and Spradling, A.C. (1982). Genetic transformation of *Drosophila* with transposable element vectors. Science. **218**, 348-353.

Rubin, G. M. and Spradling, A.C. (1983). Vectors for *P* element mediated gene transfer in *Drosophila*. Nucl. Acids. Res. **11**, 6341-6351.

Sakai, J., Chalmers, R.H. and Kleckner, N. (1995). Identification and characterisation of a pre-cleavage synaptic complex that is an early intermediate in Tn10 transposition. EMBO. J. **14**, 4374-4383.

Sandmeyer , S.B., Hansen, L.J. and Chalker, D.L. (1990). Integration Specificity of Retrotransposons and Retroviruses. Annu. Rev. Genet. **24**, 491-518.

Sanger, F. Nicklen, S. and Coulson, A. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. **74**, 5463-5467.

Sarnovsky, R.J., May, E.A. and Craig, N.L. (1997). The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. EMBO J. **15**, 6348-6361.

Schatz, D.G., Oettinger, M.A. and Baltimore, D. (1989). The V(D)J recombination activating gene (RAG-1). Cell. **59**, 1035-1048.

Scottoline, B.P., Chow, S., Ellison, V. and Brown, P.O. (1997). Disruption of the terminal base pairs of retroviral DNA during integration. Genes. Dev. **11**, 731-382.

Searles, L.L., Jokerst, R.S., Bingham, P.M., Voelker, R.A. and Greenleak, A.L. (1982). Molecular cloning of sequences from a *Drosophila* RNA polymerase II locus by *P* element transposon tagging. Cell. **31**, 585-592.

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

Siegal, M.L. and Hartl, D.L. (1996). Transgene coplacement and High Efficiency Site-Specific Recombination With the *Cre/loxP* System in *Drosophila*. Genetics. **144**, 715-726.

Simmons, G.M. (1992). Horizontal transfer of *hobo* transposable elements within the *Drososphila melanogaster* species complex: evidence from DNA sequencing. Mol. Biol. Evol. **9**, 1050-1060.

Simonelig, M. and Anxolabetere, D. (1991). A *P* Element of *Scaptomyza pallida* is active in *Drosophila meanogaster*. Proc. Natl. Acad. Sci. U.S.A. **88**, 61026106.

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

Smith , D.B. and Johnson, K.S. (1988). Single step purification of polypeptides in *Escherichia coli* as fusions with glutathione S-transferase. Gene. **67**, 31-40.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**, 503-517.

Song, S.U., Gerasiova, 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**, 2046-2057.

Spielman, (1994). Why entomological malaria research should not focus on transgenic mosquitos. Parasitol. Today. **10**, 374-376.

Spradling, A.C. and Rubin, G. M. (1982). Transposition of cloned *P* elements into *Drosophila* germ line chromosomes. Science. **218**, 341-347.

Spradling, A.C. (1986). *P* element-mediated transformation. In *Drosophila* a practical approach, ed. by D.B. Roberts. IRL Press. p 175-197.

Spradling, A.C. Karpen, G., Glaser, R. and Zhang, P. (1992). Evolutionary Conservation of Developmental mechanisms: DNA Elimination in *Drosophila*. In *Evolutionary Conservation of Developmental mechanisms*, ed. by Spradling, A. New York: Wiley-Liss; p 39-53.

Stalder, R., Caspers, P., Olasz, F. and Arber, W. (1990). The N-terminal Domain of the Insertion Sequence IS30 Transposase Interacts Specifically with the Terminal Inverted Repeats of the Element. J. Biol. Chem. **265**, 3757-3762.

Steller, H and Pirotta, V. (1986). *P* transposons controlled by the heat shock promoter. Mol. Cell. Biol. **6**, 1640-1649.

Streck, R.D., MacGeffey, J.E. and Beckendorf, S.K. (1986). The structure of *hobo* transposable elements and their insertion sites. EMBO J. **5**, 3615-3623.

Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. Methods in Enzymology, **185**, 60-89.

Sundarasen, V, and Freeling, M. (1987). An extrachromosomal form of the Mu transposons of maize. Proc. Natl. Acad. Sci. U.S.A. **84**, 4924-4928.

Surette, M.G., Buch, S.J. and Chaconas, G. (1987). Transpososomes: Stable Protein-DNA Complexes Involved in the *In Vitro* Transposition of Bacteriophage Mu. Cell. **49**, 253-262.

Tausta, S.L. and Klobutcher, L.A. (1989). Detection of circular forms of eliminated DNA during macronuclear development in *E. crassus*. Cell. **59**, 1019-1026.

Tenzen, T., Matsutani, S. and Ontsubo, E. (1990). Site Specific Transposition of IS630. J. Bac. **172**, 3830-3836.

Ton-Hoang, B., Betermier, M., Polard, P. and Chandler, M. (1997). Assembly of a strong promoter following IS911 circularisation and the role of circles in transposition. EMBO.J. **16**, 3357-3371.

Tsubota, S., Ashburner, M. and Schedl, P. (1985). *P* Element Induced Control Mutations at the *r* gene of *Drosophila melanogaster*. Mol. Cell. Biol.. **5**, 2567-2574.

Tudor, M., Lobocka, M., Goodell, M., Peltitt, J. and O'Hare, K. (1992). The *pogo* transposable element family of *Drosophila melanogaster*. Mol. Gen. Genet. 232, 126-134.

Turlan, C. and Chandler, M. (1995). IS-1 medited intramolecular rearrangements: formation of excised transposon circles and replicative deletions. EMBO.J. **14**, 5410-5421.

van Gent, D.C., Oude Groeneger, A.A. M. and Plasterk, R.H.A. (1992). Mutational analysis of the integrase protein of human immunodeficiency virus type two. Proc. Natl. Acad. Sci. U.S.A. **89**, 9598-9602.

van Gent, D.C., Mizuuchi, K. and Gellert, M. (1996). Similarities between Initiation of V(D)J recombination and Retroviral Integration. Science. **271**, 1592-1594.

van Luenen, H.G.A.M., Colloms, S.D. and Plasterk, R.H.A. (1993). Mobilisation of quiet, endogenous *Tc*3 transposons of *Caenorhabditis elegans* by forced expression of Tc3 transposase. EMBO. J. **12**, 2513-2520.

van Luenen, H.G.A.M., Colloms, S.D. and Plasterk, R.H.A. (1994). The mechanism of transposition of Tc3 in *C. elegans*. Cell. **79**, 293-301.

van Luenen, H.G.A.M. and Plasterk, R.H.A. (1994). Target site choice of the related transposable elements *Tc*1 and *Tc*3 of *Caenorhabditis elegans.* Nucl. Acid. Res. **22**, 262-271.

Varmus, H. and Brown, P. (1989). Retroviruses. In Mobile DNA, ed. by D.E. Berg and M. Howe. American society for Microbiology, Washington D.C. p 53-108.

Velterop, J.S., Dijkhuizen, M.A., van't Hof, R. and Postma, P.W. (1996). A versatile vector for controlled expression of genes in *Escherichia coli* and *Salmonella typhimurium.* Gene.**153,** 63-65.

Vink, C., Yeheskiely, E., van der Marel, G., Boom, J.H. and Plasterk, R.H.A. (1991). Site-specific hydrolysis and alcoholysis of human immunodeficiency virus DNA termini mediated by the viral integrase protein. 19, 6691-6698.

Vos, J.C., van Luenen, H. G. A. M. and Plasterk, R.H.A. (1993). Characterisation of the *Caenorhabditis elegans Tc1* transposase *in vivo* and *in vitro*. Genes. Dev. **7,** 1244-1253.

Vos, J.C and Plasterk, R.H.A. (1994). Tc1 transposase of *Caenorhabditis elegans* is an endonuclease with a bipartite DNA binding domain. EMBO J. **13**, 6125-6132.

Vos, J.C., De Baere, I. and Plasterk, R.H.A. (1996). Transposase is the only nematode protein required for *in vitro* transposition of *Tc1*. Genes. Dev. **10**, 755-761.

Wang, H-M (1997). The sequence specific DNA binding activity of *hobo* transposase. PhD Thesis.

Williams, K., Doak, T.G. and Herrick, G. (1993). Developmentally precise excision of *Oxytricha trifallax* telomere bearing elements and formation of circles closed by a copy of the flanking target DNA duplication. EMBO. J. **12**, 4593-4601.

Wilson, C., Pearson, R.K., Bellen, H.J., O'Kane, C.J., Grossniklaus, U. and Gehring, W.J. (1989). *P* element mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in *Drosophila*. Genes Dev. **3**, 1301-1313.

Wishart, W.L., Broach, J.R. and Ohtsubo, E. (1985). ATP dependent specific binding of *Tn3* transposase *Tn3* inverted repeats. Nature, **314**, 556-558.

Yoshioki, K., Kanda, H., Akiba, H., Enoki, M. and Shiba, M. (1991). Identification of an unusual structure in *D. melanogaster* transposable element copia: Evidence for *copia* transposition through an RNA intermediate. Gene. **103**, 179-184.

Zerbib, D., Polard, P., Escoubas, J.M., Galas, D. and Chandler, M. (1990)a. The regulatory role of the IS1- encoded InsA protein in transposition. Mol. Microbiol. **4**, 471-477.

Zerbib, D., Prentki, P., Gamas, P., Freund, E., Galas, D.J. and Chandler, M. (1990)b. Functional organisation of the ends of IS1: specific binding site for an IS1 encoded protein. Mol. Microbiol. **4**, 1477-1486.