THE UNIVERSITY OF CALGARY

Transposable Elements for Insect Transformation: The Mariner Element and the I-PpoI Intron-Encoded Endonuclease

by

Weimin Wang

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

CALGARY, ALBERTA

November, 1998

ABSTRACT

Mariner transposable elements and I-Ppo endonuclease, originally found in *Drosophila mauritiana* and *Physarum polycephalum* respectively, have shown the potential to conduct transposition in a wide range of host genomes. We established a Mariner gene transformation system in a *Bombyx mori* cell line using Mos1 transposable element. The system consists of a Mos1 expression vector to provide transposase and a Mos1 donor vector to provide the substrate DNA and desired genes, including the reporter gene CAT and the selection gene HmB, for transposition. Both the vectors were co-transfected into Bm5 cell line. Plasmid rescue assays identified Mos1-mediated genomic transpositions in 5 categories of insertion events within the transfected cells. Our work proves that Mariner transposable element can be used as a universal gene transformation vector and provides a convenient tool for developing transgenic silkworm. Using a similar concept, we also tested the feasibility of I-Ppo mediated gene transformation system in Bm5 cell line and showed that the I-Ppo expression vector we constructed can produces the endonucleases that cleave on target sites of both plasmid DNA and genomic DNA in Bm5 cells.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my gratitude for the endless support and enthusiastic supervision throughout my studies from my supervisor, Dr. Kostas Iatrou. I am in debt to him for his insights, perspective, knowledge and blunt criticism that not only have guided me through the course of my Master program, but also will benefit my future career. I also thank my committee members, Dr. Randy Johnson and Dr. James McGhee for their unreserved guidance and constructive comments. I am also grateful to Dr. K. Sanderson for consenting to act as my external examiner.

My appreciation is also extended to the members both past and present of Dr. Iatrou's lab for helping me out with problems. My warm appreciation is owed to Drs. Luc Severs and Maolong Lu for their experimental contributions in my study. Thanks are also due to Mr. Patrick Farrell for providing some valuable materials and Ms. Martine Lunke and Mr. Frederic Kendirgi for proof reading parts of this thesis. A special thank is reserved for Drs. D.L. Hartl and V.M. Vogt for the gifts of Mos1 and I-Ppo DNA, which made my research project possible.

I acknowledge gratefully the financial assistance that was provided to me by Dr. Iatrou (from his grant) and the University of Calgary (in form of GA). This work was supported through operating grants awarded to Dr. Iatrou by the National Sciences and

Engineering Research Council of Canada and the Medical Research Council of Canada.

Warm appreciation shall always be with me for my parents who gave me the best they could at the expense of personal sacrifices. I am forever indebted to my wife Shijing and my son Charles who endured the many months, weekends and evenings that were devoted to writing and for giving me the love, strength and encouragement whenever I needed them most.

DEDICATION

To my wife, Shijing Fan

and

my son, Charles Wang

and

my parents, Bijuen Chang and Kongjuen Wang

and

my brother, Weiren Wang

TABLE OF CONTENTS

Approval Page ii

Abstract..... iii

Acknowledgements iv

Dedication vi

Table of Contents vii

List of Tables xii

List of Figures xiii

List of Abbreviations xv

GENERAL INTRODUCTION 1

I. Classification of and Recent Advances in Transposable Elements.....1

II. Molecular Biology of Transposable Elements 3

1. Structure of Class II Transposable Elements.....3

2. The Mechanism of Transposition 4

3. Regulation of Transposable Elements 5

III. A Specific Case in Insects: Transposition of Drosophila P Elements 6

1. Structure 6

2. Regulation of Transposition..... 6

3. Maintenance in Genome.....7

IV. The Need to Develop non-Drosophila Gene Transformation Systems 8

V. The Need to Establish a Gene Transformation System in *Bombyx mori*....9

MATERIALS AND METHODS 10

I. Plasmid DNA Preparation 10

1. Quick miniprep 10

2. Mini-prep 11

3. Large-scale preparation 12

II. Cloning 13

1. Preparation of competent cells using calcium chloride 13

2. Preparation of competent cells for high-voltage electroporation 14

3. DNA Ligation 14

4. Transformation using chemical method 15

5. Transformation using high-voltage electroporation..... 15

6. Identification of recombinant clones..... 16

III. Purification of DNA Fragments..... 17

1. Electroelution 17

2. DEAE-cellulose membrane..... 18

IV. Maintenance of Bm5 cell line and Transfection Protocol 18

V. Nucleic Acid Isolation from Bm5 Cells 19

VI. Plasmid Rescue Assay..... 20

VII. Sequencing 20

1. "Classical" sequencing..... 20

2. Cycle sequencing.....21

VIII. Preparation of Radiolabelled ^{32}P Probes 22

1. Hexanucleotide primers 22

2. PCR labelling.....22

IX. Southern Blot Hybridization 23

X. Northern Blot Hybridization 24

XI. Protein Extraction 24

XII. Chloramphenicol Acetyltransferase (CAT) Assay 25

CHAPTER ONE: THE MARINER TRANSPOSABLE ELEMENT FOR GENETIC TRANSFORMATION IN *Bombyx mori* 26

I. Introduction.....27

II. Materials and Methods 32

1. Transformation Vectors 32

2. PCR reactions and sequencing 38

3. Southern blot analysis 43

III. Results 44

1. Donor Plasmid pMCH1 --- the Substrate DNA for Mos1 Transposase 44

2. The Post-Transfection Fate of Vectors within the Cells.....44

3. Plasmid Rescue Assays 50

Class I: Self-Ligated Vectors or Random Integrations.....53

Class II: Random Integration at the Mos1 3' IR Side.....59

Class III: Mos1 Mediated Inter-Plasmid Transposition 66

Class IV: Mos1 Mediated Intra-Plasmid Transposition.....66

Classs V: Mos1 Mediated Genomic Transposition 67

Class VI: Mos1 Excision Intermediate Products.....67

Class VII: Rescued Plasmids originated from complex Mos1 Recombination events 72

4. Confirmation of the Genomic Transposition.....73

IV. Discussion 79

CHAPTER TWO: THE I-Ppo ENDONUCLEASE AS TOOL FOR GENETIC TRANSFORMATION OF BOMBYX MORI 85

I. Introduction 86

1. GroupI Introns 86

2. Mechanism of Group I "homing" 87

3. Intron-Encoded Endonuclease I-Ppo 93

4. Regulation of Intron Mobility 101

5. Horizontal Transmission of I-Ppo Element 102

6. I-Ppo Endonuclease and Bombyx mori 104

II. Materials and Methods 108

1. Construction of the I-Ppo Expression Vectors with different promoters 108

2. Construction of the I-Ppo Donor Vectors 109

3. Construction of I-Ppo Target Test Vectors 114

4. Other Methods.....115

III. Results.....115

1. Cleavage of the I-Ppo recognition sequence on the plasmid DNA..115

2. I-Ppo Cleavage of Genomic rDNA in Bm5 Cells.....118

3. Cotransfection of the I-Ppo Expression Vectors and the Donor Vectors 123

IV. Discussion 123

PERSPECTIVE AND FUTURE DIRECTION 127

BIBLIOGRAPHY.....129

APPENDICES.....142

Appendix 1. The DNA sequence and the translated amino acid sequence of Mos1 mariner transposable element 143

Appendix 2. The DNA sequences of the rescued fragments from B. mori genome 146

Appendix 3. PCR Detection of I-Ppo Initiated Integration 154

LIST OF TABLES

Table 1-1. The Components of the Mos1 Transformation Vectors 39-40

Table 1-3. The Summary of the Plasmid Rescue Assays..... 54-55

Table 2-1. The Components of the I-Ppo Transformation Vectors 116-117

LIST OF FIGURES

Figure 1-1. The outline of the experimental processes.....30-31

Figure 1-2. The physical maps of the Mariner transformation vectors 33-34

Figure 1-3. Construction of the pMCH transformation vectors 35-37

Figure 1-4. The PCR primer locations on donor vector pMCH1 41-42

Figure 1-5. Southern blot of transfected plasmids in Bm5 cells 45-46

Figure 1-6. CAT assays using transfected Bm5 cell extracts 47-48

Figure 1-7. The illustration of Mos1 mediated pMCH1 transposition in B. mori genomic DNA 51-52

Figure 1-8. Class I: random integrations or self-ligated vectors 56-57

Figure 1-9. Class II: random integration 60-61

Figure 1-10. Class II: Mariner inter-plasmid integrations 62-63

Figure 1-11. Class IV: Mariner intra-plasmid integrations 64-65

Figure 1-12. Class V: Mos1 mediated genomic integrations 68-69

Figure 1-13. Class VI: Mariner self-excison 70-71

Figure 1-14. Southern blots of *Bombyx* genomic DNA hybridized with the rescued fragments 74-75

Figure 1-15. PCR analysis 77-78

Figure 1-16. Model for Mos1 mariner transposition 82-83

Figure 2-1. Intron structure and splicing pathway.....88-89

Figure 2-2. The double-strand-break repair pathway for intron mobility 91-92

Figure 2-3. The schematic representation of I-Ppo intron DNA constructs 95-96

Figure 2-4. The nucleotide and deduced amino acid sequences of PpLSU intron 97-98

Figure 2-5. Structural models of PpLSU intron 99-100

Figure 2-6. The I-Ppo recognition sequences in three isoforms of B. mori 28S rDNA 105-107

Figure 2-7. The physical maps of the plasmids constructed in I-Ppo project 110-111

Figure 2-8. I-Ppo donor vectors 112-113

Figure 2-9. The Southern blot showing I-Ppo cleavage on the transfected target plasmids 119-120

Figure 2-10. The Southern blot showing I-Ppo cleavage on the transfected Bm5 genomic DNA 121-122

Appendix 3. PCR detection of I-Ppo initiated integration 154-155

LIST OF ABBREVIATIONS

A	adenosine
acetyl CoA	acetyl coenzyme A
ASV	Avian Sarcoma Viruses
ATP	adenosine triphosphate
bp	base pair
BSA	bovine serum albumin
C	cytidine
°C	degree centrigrade
Ci	Curie, the unit for radioactivity
cpm	counts per minute
CRE	a 38 kDa phage P1 encoded recombinase protein (it is named for causing recombination)
ddNTP	dideoxyribonucleoside 5-triphosphate where N indicates adenosine, cytidine, guanosine, or thymidine
DEAE	diethylaminoethyl
DNA	deoxyribonucleic acid
DNase	dexyribonuclease

dNTP	deoxyribonucleoside 5-triphosphate where N represents adenosine, cytidine, guanosine or thymidine
DSB	double strand break
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetate
EtBr	ethidium bromide
EWB	electroshock washing buffer (10% redistilled glycerol in ultraopure water)
FLP	(pronounced "flip") a 43 kDa site-specific recombinase encoded by the 2 μ circular plasmid of budding yeast, *Saccharomyces cerevisiae*. This plasmid has two 34 bp inverted target sites. Recombination between these FRT sites result in inversion of one half of the plasmid with respect to the other, which provides a mechanism for producing multiple plasmid copies from a single replication initiation by flipping the direction of migration of one replication fork.
FRT	Flp recognition target
G	guanosine
Hepes	N-2-hydroxyethylpiperazine-N'-2thanesulfonic acid
HIV	human immunodeficiency virus
HmB	hygromycin B, an aminoglycosidic antibiotic that inhibits protein synthesis by interfering with translocation and causing misreading
IPS	internal processing site
kb	1,000 bases

L	liter
LB	Luria-Bertani medium
lox P	a 34 bp long Cre-specific recognition sequence site in phage P1 DNA
M	Molar
min	minute
NP-40	Nonidet P-40
OD	optical density
ORF	open reading frame
OLB	oligonucleotide-labeling buffer
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
Pfu	a proofreading DNA polymerase isolated from *Pyrococcus furiosus*
PMSF	phenylmethylsulfonyl fluoride
RBZ	ribozyme
RNA	ribonucleic acid
RNase A	Ribonuclease A
rpm	revolutions per minute
SDS	sodium dodecyl sulfate
Sevag	Chloroform : isoamyl alcohol (24:1)
T	thymidine
TE	Tris-EDTA buffer
TLC	thin-layer chromatography

Tris	tris-(hydroxylmethyl)-aminomethane
Tsg	a thermostable DNA polymerase isolated from a strain of *Thermus sp.*
U	uridine
UTR	untranslated region
UV	ultraviolet
V	volts
W/V	weight/volume
v/v	volume/volume
%	percent
μ	micro

GENERAL INTRODUCTION

Genomes are usually regarded as static, changing rarely on the evolutionary scale. Genetic recombination, although allowing exchange between homologous chromosomes, does not reorganize the genetic material within the species. In contrast to the overall stability, however, mobile elements, which can move directly from one site in the genome to another without using outside elements (such as phage or plasmid DNA), provide variations between individual genomes. Progress in the understanding of the mechanisms that control the mobility of mobile elements indicates the importance of transposable elements not only for their evolutionary consequences, but also for the experimental manipulation of DNA, and the introduction of foreign DNA into the genome [Boussy et al., 1988; Cohen et al., 1985].

I. Classification of and Recent Advances in Transposable Elements

Transposable elements represent probably the smallest entities with respect to autonomously replicating units. In their simplest forms, they consist of one gene, usually flanked by short repeated DNA sequences; the encoded protein (the transposase) can trigger the replicative spread of the element within or between genomes (Plasterk, 1993). Transposable elements were first identified as spontaneous insertions in bacterial operons. In recent years, transposable elements have been increasingly exploited as tools for

mutational analysis and *in vivo* genetic engineering, initially just in *E.coli*, but subsequently in a diverse array of microorganisms and eucaryotes. Among their major uses are insertion mutagenesis and the introduction of linked selectable markers, analysis of transcriptional and translational regulation and protein localization, DNA sequencing, and the introduction of engineered genes into the host organisms [Levis et al., 1984; Orr-Weaver et al., 1986; Fisher et al., 1986].

Mobile DNA elements fall into three general classes [Finnegan 1989, Kidwell 1993]. Class I encompasses the RNA-mediated transposons which are related to retroviruses. These RNA are converted into DNA copies, and then become integrated at new sites in the genome. Class II contains the DNA-mediated transposons which exist as sequences of DNA coding for proteins that are able to manipulate DNA directly so as to propagate themselves within the genome. The third class of the mobile elements comprises mobile introns which conduct genetic material exchanges in a site-specific fashion, a process that is termed as "intron homing".

The mobile DNA field is rooted in Barbara McClintock's seminal discovery of transposable elements in maize. Using genetic and cytogenetic analyses in the 1940's, she demonstrated that there are elements that can transpose to new chromosomal locations, alter the expression of nearby genes, and cause chromosome breakage, all in a developmentally regulated fashion. Her discoveries are more remarkable in that they were made without knowledge of DNA structure or the current recombinant DNA technologies. They preceded by several decades the discoveries of phage Mu, the bacterial

insertion sequence and drug resistance transposons and the mobile elements of *Drosophila melanogaster* and other eucaryotes.

The discovery of antibiotic resistance transposons led to a virtual explosion in the use of transposable elements for *in vivo* genetic engineering in diverse species. These have been successfully used to provide regions of homology and useful restriction sites for targeted *in vivo* and *in vitro* DNA rearrangement; to introduce specific genes into new hosts or into particular chromosomal sites; to introduce replication origins or the conjugative DNA. These elements can insert into many sites and transpose in many species, moreover, they are easy to select. Their overall small size facilitated physical mapping and cloning.

The current interest in mobile DNA also comes from many other research areas: basic DNA structure, the nature of repetitive DNA, the cause of mutation, the principle of genome evolution, the control of gene expression, the operation of developmental circuits, the virulence mechanisms of pathogenic microbes, the spread of antibiotic resistance of bacteria, and the emergence of drug resistance in human tumors.

II. Molecular Biology of Transposable Elements

1. Structure of Class II Transposable Elements

Class II transposable elements are chosen as a prototype to be discussed here because they are directly relevant to my study and are well documented. These DNA transposons have been found in all phyla: they are the most common transposable

elements in bacteria, and numerous families have been described in fungi, ciliates, plants, worms, insects, fish, and mammals (Berg & Howe, 1989). This class of transposable elements is characterized by the presence of two inverted repeats flanking a DNA sequence encoding a protein named transposase. This protein is involved in the processing of DNA at the donor and target sites by a "cut-and-paste" process. The transposase, in most but not all instances, is the only requirement for transposition. For this reason, the control of transposition is often limited to the regulation of the expression of this protein.

2. The Mechanism of Transposition

The mechanism of transposition appears to be essentially the same for most transposases and integrases [Doak et al. 1994]. This similarity has been confirmed by the analysis of three crystallographic structures of transposases and integrases from bacteriophage Mu, HIV, and ASV retroviruses [Grindley & Leschziner, 1995]. Crystal structure comparisons show that the catalytic domain for all three proteins is very similar, bringing together two aspartate and glutamate residues separated by 50 to 70 (D-D) and 35 (D-E) residues in the primary structure. This catalytic domain is most likely responsible for a transphosphorylation reaction necessary for DNA cleavage and strand transfer during the transposition, and is highly conserved [Doak et al., 1994]. The transposition involves two sequential steps. Site-specific cleavage of the DNA takes place first at the ends of the transposable element, and the complex of transposase-element ends are then brought to a DNA target, where the strand transfer is carried out by covalently joining the 3' ends of the element to the target DNA. A strand gap is later filled, probably by host-dependent

repair mechanisms, and results in the characteristic direct duplications found at the ends of all transposable elements.

3. Regulation of Transposable Elements

Transposable elements propagate by inserting into new locations in the genomes of the host they inhabit. In fact, they are major constituents of eukaryotic genomes, making up to 10% of the *Drosophila* genome and 35% of the human genome. Their transposition might thus negatively affect the fitness of the host. Therefore, their mobilization must be under tight regulation, both in terms of its frequency and insertional specificity, to avoid the accumulation of mutations that could be deleterious to the host. But transposable elements must also maintain a certain level of activity to ensure their propagation and survival. This level of transposition is probably the result of a balance between the interests of the transposable elements and those of the hosts they inhabit.

The nature of this control depends on the structure of the transposable element. The transposase encoded by DNA elements is necessary, and in most cases sufficient, for mobilization. For this reason, the regulation of transposition must rely on cellular processes necessary for producing active transposase. This may include regulation at the level of transcription, differential splicing, translation, and protein-protein interactions. In general, however, the regulation of the activity of these elements depends on intrinsic factors with little direct input from the host.

The Tc1 element of *C. elegans* seems to represent the simplest case analyzed so far. The regulation takes place at the level of transcription [Vos et al., 1993]. Genetic

factors of the host can determine the transposition rate of Tc1, but these factors probably correspond to particular active copies of the Tc1 element. This simple mechanism might explain the relative success of this element and its widespread distribution, which has been found in many different species including ciliates, fungi, *Drosophila*, fish, and humans [Plasterk, 1996].

III. A Specific Case in Insects: Transposition of *Drosophila* P Elements

1. Structure

Drosophila P element is a 2.9 kb sequence flanked by 31 bp perfect inverted repeats and structured in 4 exons that encode a single 87 kDa transposase protein [Engels, 1989]. The transposase does not bind to the terminal repeat. Instead, two perfect 11 bp inverted subterminal repeats at the 5' and 3' ends are essential for transposition and are considered the binding sites for this protein. Transposase is the only requirement for P element transposition.

2. Regulation of Transposition

Mobilization of P element *in vivo* is a complex process controlled by interactions with the host at many different levels. For example, splicing of the third intron of the P element occurs only in the germ line but not in somatic cells, where a defective protein is produced. Therefore, somatic transposition is almost absent [Engels, 1989].

Copy number in the host also plays role in the regulation. Although the mobilization rate of P elements is close to 25×10^{-2} transpositions per element per generation under "dysgenic" conditions [Song et al., 1994], P strains that carry approximately 50 copies of the P element show very little or no transposition [Engels, 1989]. The evidence suggests that the P element-encoded 66 kDa protein is a repressor of transposition. Several additional mechanisms for cytotype repression have been suggested, including transcriptional regulation and multimer poisoning [Rio & Robin, 1988], antisense RNA [Rasmunson et al., 1993], and titration of transposase in the presence of many P element ends [Simmons & Bucholz, 1985].

3. Maintenance in Genome

Two different and general transposition strategies could be outlined for transposable elements: replicative and non-replicative transposition. The first is carried by retrotransposons and mobile introns. After transposition, the donor site is preserved intact and in the same position as the original copy. The second strategy is used by DNA transposons that move by a cut-and paste mechanism. Double-strand breaks (DSBs) are produced at the ends of the element during this process, and the donor copy physically moves to the new target site. This class of transposition raises an important question: how can non-replicative transposable elements, such as the P element and Mos1 transposable elements, increase their copy number so quickly after a short number of generations?

The explanation came from the gap repair process. After excision, the DSB ends start a search for homology of templates throughout the genome, and serve as primers that

bind to templates. Repair of the gap could be carried out very effectively by copying sequences up to 8000 bp into the gap [Nassif et al., 1994]. Interestingly, this search can even target sequences harbored in non-integrated plasmids [Keeler et al., 1996].

IV. The Need to Develop non-Drosophilid Gene Transformation Systems

Although an impressive list of organisms is now available for foreign gene transfer, insects, which account for over five sixths of all multicellular organisms on Earth, have so far proved stubbornly resistant to the overtures of the genetic engineer [Marshall, 1998]. Of the over 1 million species of insects identified so far, only the fruit, *D. melanogaster*, is transformable on a routine basis. Considering the huge impact of these insects both on global food production and as vectors for some of the deadliest diseases known to humans, broadening the range of species amenable to genetic manipulation will be important in designing novel pest control strategies and introducing new traits into beneficial insects.

With the recent reports of stable germline transformation of the Mediterranean fruit fly [Loukeris et al. 1995] and the dengue/yellow fever mosquito [Jasinskiene et al., 1998; Coates et al., 1998], the genetic engineering of other economically and medically important insect species now appears within reach.

The availability of transgenic insects will allow gene function to be analyzed systematically in relation to the ability of the vectors to transmit diseases; help clarify the types of genes that can be introduced into insects to break the cycle of transmission;

possibly alter behavioral genes to give pest an alternative predilection; or express antisense RNA to inhibit viral or parasite infectivities. An alternative possibility would be to use transgenes encoding single chain antibodies to neutralize and kill the parasites themselves.

V. The Need to Establish a Gene Transformation System in *Bombyx mori*

Of the few domesticated insects, silkworms have been drawing the geneticists' interest for economic and academic reasons. Particularly, the research interest in our lab is focused on the study of development of silkworm ovarian follicles during pharate adult development. So far, several genes with suggestive expression patterns have been identified, such as BHR3A, BmE75C, BmGATAβ, BmFTZ-F1, etc.. However, when the study reaches questions related to the function of these genes, constructs for over-expression of these genes under inducible promoter or ovarian-specific promoter control have to be introduced into the genome of *B. mori*. Another area of interest to our lab is the development of genetically engineered insects for insect pest control purposes. For example, a toxic gene driven by ovarian-specific promoters could be introduced into the population, resulting in the spreading of individuals with defective reproduction organs (ovaries). This could limit the growth of wild-type population. As with many other insects, transgenesis in silkworm has not been reproducible or of sufficient efficiency. Some newer attempts to generate transgenic silkworms have followed two different approaches: using the piggyBac transposon [Handler et al., 1998] and pantrophic retroviruses [Finnegan,

1997], or using homologous recombination to carry out gene targeting [Golic et al., 1997] or exploiting attenuated baculoriruses to ferry transgenes into insects [P.J. Farrell & K. Iatrou, unpublished data].

For my thesis research project, I have attempted to establish genetic transformation systems in a *B. mori* cell line using two mobile DNA elements. The first one, MosI Mariner element, is a DNA transposon; the second one, the I-PpoI endonuclease, is a group I mobile intron-encoded endonuclease responsible for initiating gene conversion events. Although both elements have never been tested in *Bombyx*, they have little apparent restriction on the selection of their hosts. Furthermore, a subfamily of Mariner-like elements and the I-Ppo recognition site have been found in the *Bombyx* genome and the genomes of a wide variety of other organisms [Robertson, 1993]. Hence it is likely that both elements could be developed as universal gene transformation vectors.

GENERAL MATERIALS AND METHODS

I. Plasmid DNA Preparation

1. Quick miniprep

A single colony from *E.coli* HB101 transformed with pBluescript SK (pBS, Stratagene) or recombinant plasmid was inoculated in 2 ml LB medium containing 100 μg/ml of ampicillin and cultured at 37°C overnight. 100 μl of the culture was pelleted at 3,000 rpm for 1 min in a benchtop microcentrifuge. The pellet was resuspended in 50 μl

TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and vortexed vigorously with an equal volume of phenol:chloroform (1:1). After centrifuging for 2 min, 15 µl of supernatant was mixed with 2.5 µl of 6× DNA dye [0.25% bromophenol blue, 0.25% xylene cyanol FF and 40% (w/v) sucrose] and 0.5 µl of 10 µg/µl RNase A. The mixture was kept at room temperature for 2 min and resolved on a 1% agarose gel for supercoiled plasmid DNA. Supercoiled recombinant plasmid DNA with inserts migrates slower than that from pBS.

2. Mini-prep

One and one half ml of an overnight bacterial culture was pelleted at 14,000 rpm for 30 seconds in a benchtop microcentrifuge and resuspended by vortexing in 100 µl of ice-cold solution I (50 mM glucose, 25 mM Tris-HCl pH 8.0, and 10 mM EDTA). Then, 200 µl of freshly prepared solution II (0.2 N NaOH and 1% SDS) were added and mixed. After a 5 min incubation on ice, 150 µl of cold solution III (90 µl of 3 M potassium acetate, 17.25 µl of glacial acetic acid, and 42.75 µl H_2O) was added, mixed well, and stored on ice for 5 min. After spinning for 10 min at 14,000 rpm to pellet debris, the supernatant was transferred to a fresh tube and extracted with 500 µl of phenol:chloroform (1:1). In a new tube, the aqueous phase was mixed with 1 ml of 95% ethanol. Nucleic acid, including plasmid DNA and bacterial RNA, was pelleted by spinning at 14,000 rpm in a benchtop microcentrifuge. The pellet was washed with 75% ethanol and dissolved in 50 µl of TE containing DNase-free RNase A (20 µg/ml).

3. Large-scale preparation

A single colony was vigorously shaken in 2 ml LB overnight. The culture was inoculated into 200 ml LB and further incubated overnight. Cells were cooled on ice for 5 min and pelleted by centrifugation at 4,500 rpm for 10 min in a Sorvall GS3 rotor. The pellet was resuspended in 100 ml of ice-cold STE (100 mM NaCl, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA). After centrifugation, the pellet was resuspended in 10 ml of solution I (same as in minipreps) and lysed by storing on ice for 15 min following addition of 1 ml lysozyme (10 mg/ml in 10 mM Tris-HCl pH 8.0). The bacterial and plasmid DNA were denatured for 15 min by adding 20 ml of freshly prepared solution II, and the plasmid DNA was renatured for 10 min by adding 15 ml of solution III. After spinning at 5,000 rpm for 10 min, the supernatant was mixed well with 0.6 volume of isopropanol and stored at room temperature for 10 min. The nucleic acid from the supernatant was precipitated by centrifugating at 5,000 rpm for 10 min and dissolved in 3 ml of TE.

To purify the plasmid DNA, 3.4 g of cesium chloride and 200µl of ethidium bromide were added (EtBr, 10 mg/ml). The tube was spun at 10,000 rpm for 10 min at room temperature in a SS34 rotor. The cleared supernatant was loaded into a 3.9 ml Quick-seal™ tube (Beckman) and centrifuged at 100,000 rpm for at least 10 hours or overnight at 20°C in a TL-100 benchtop ultracentrifuge (Beckman) equipped with a TLN-100 rotor.

After centrifugation, the banded plasmid DNA was recovered by using a 1 ml syringe and a 21-gauge needle. Typically, 0.5 ml of the solution containing plasmid DNA was collected. The EtBr in the solution was removed by extraction at least 5 times with an

equal volume of n-butanol saturated with H_2O until the pinkish colour disappeared. The solution was diluted with 3-4 volumes of distilled water and precipitated with 2.5 volumes of 95% ethanol. After centrifugation at 10,000 rpm for 20 min in the SS34 roter (Sorvall), plasmid DNA was dissolved in distilled water and precipitated three times from 0.25 M ammonium acetate with 2.5 volumes of 95% ethanol. Finally, the pellet was dissolved in distilled water, and the DNA concentration was determined on a Beckman spectrophotometer.

II. Cloning

1. Preparation of competent cells using calcium chloride

E. coli strain HB101 was streaked on a LB plate and incubated at 37°C for 20 hours. A single colony was inoculated into 2 ml of LB and cultured at 37°C for 5 hours. The culture was then inoculated into 200 ml LB and shaken vigorously until the OD_{600} reached 0.3-0.5. The cells were recovered by centrifugation at 4,000 rpm for 10 min in a GS3 rotor (Sorvall). The pellet was resuspended in 40 ml of ice-cold 0.1 M $CaCl_2$ and stored on ice for 30 min. The cells were pelleted again and gently dissolved into 8 ml of ice-cold 0.1 M $CaCl_2$. Competent cells were used for transformation within two days by storing at 4°C. Otherwise, 0.2 ml of the cells were mixed with 0.1 ml of 50% glycerol and stored at -70°C for later use.

2. Preparation of competent cells for high-voltage electroporation

E. coli strain DH10B was streaked on a freshly-prepared SOB plate and incubated at 37°C overnight (approximately 24 hours). A single colony was inoculated in 2 ml SOB (20 g Bacto-tryptone, 5 g Bacto yeast extract, 0.58 g NaCl, 0.19 g KCl, 10 ml 2 M $MgCl_2$ or $MgSO_4$, dH_2O to 1 L) and cultured for about 2 hours till it became slightly cloudy. One half ml of the culture was mixed with 0.5 ml of 60% SOB-40% glycerol in 1.5 ml tubes, chilled on ice, then stored at -80°C for further use. The DH10B cells in frozen glycerol stock were streaked on SOB-Mg plates and incubated at 37°C overnight. A single 2-3 mm colony was dispersed into 50 ml SOB-Mg (SOB media without Mg^{2+}) and incubated in a rotary shaker (at 275 rpm) at 37°C overnight. Five ml of the culture was then inoculated into 500 ml of SOB-Mg and incubated at 37°C until OD_{500} reached 1 (about 2-3 hours). The cell suspension was pelleted at 4,000 rpm in Sorvall GSA rotor at 4°C and resuspended in equal volume of EWB (40 ml glycerol, 360 ml dH_2O) to wash cells twice. The pellet was resuspended in 100 μl of EWB. Then 110 (or 25) μl aliquots were dispensed into 1.5 ml tubes and frozen at -80 °C. To measure the concentration of the cells, 10 μl of cells was diluted to 3,000 μl and their OD measured in the spectrophotometer. The OD_{500} was usually around 0.75 to 1.4.

3. DNA Ligation

Every 10 μl of ligation mixture for cohesive termini contained 0.1-0.5 μg of vector DNA, a 3-fold molar excess of insert DNA, 20 mM Tris-HCl pH 7.6, 5 mM $MgCl_2$, 5 mM DTT, 5 μg/ml of BSA, 1 mM ATP and 1 unit of T_4 DNA ligase (Gibco/BRL). The

mixture was incubated at 14°C overnight. The condition for blunt-end ligation was the same as that for cohesive termini except that the concentration of vector DNA was 1 μg, ATP was 0.5 mM, and the incubation temperature was 16°C.

4. Transformation using chemical method

Three μl of ligation solution was gently mixed with 100 μl of competent cells in a 1.5 ml tube and stored on ice for 30 min. The tube was then incubated at 42°C for 2 min and transferred back to the ice immediately. The contents of the tube were added into 800 μl of LB and gently mixed at 37°C for 1 hour. The cells were recovered by centrifugation at 3,000 rpm for 5 min in a microcentrifuge and resuspended into 200 μl LB. One hundred μl of cell suspension was spread on a LB plate containing ampicillin (100 μg/ml). The LB plate was incubated at 37°C overnight.

5. Transformation using high-voltage electroporation

Twenty μl of electrocompetent cells was mixed gently with 1 μl of previously precipitated ligation mixture that had been dissolved in 0.5×TE. The DNA-competent cell mixture was stored on ice for 30 min. The mixture was then transferred to ice-cold electroporation cuvette. The electroporation device (BRL) was set at the following condition: voltage = 2,500 V, Cap = 2.5μS. The cuvette was placed in the chamber and the electric pulse was discharged. The sample was immediately mixed with 1 ml of SOC recovery buffer (100 ml of SOB-Mg medium, 1 ml of 2M $MgCl_2$-$MgSO_4$, 1 ml of 2M glucose) and transferred to a 15 ml tube, which was incubated with shaking (225 rpm) at

37°C for 45 min. Twenty five μl and 100 μl aliquots of the transformation mixture were plated on the LB-Amp (100 μg/ml) plates and incubated at 37°C overnight.

6. Identification of recombinant clones

Supercoiled plasmid DNA was extracted from transformed single colonies by the quick miniprep method and analysed by a PCR method on a 1% agarose gel. There is a linear relationship between the logarithm of the electrophoretic mobility of DNA (μ) and the gel concentration (τ), which is described by the equation [Sambrook et al., 1989]:

$$\log \mu = \log \mu_0 - K_r\tau$$

where μ_0 is the free electrophoretic mobility of DNA and K_r is the retardation coefficient. Thus, by using gels of different concentrations, it is possible to resolve a wide size range of DNA molecules. On the 1% agarose gels, given the insertion length (< 1.8kb) and expected length of the linearized recombined plasmid, an empirical formula can be applied to predict the migrating position of supercoiled DNA of recombinant clones, which run the same distance on the gel as a linearized plasmid with M_s kb (calculated as follows):

$$M_S = M_L - 1.8 + M_I$$

Where M_S represents the length (in kb) of a plasmid which the supercoiled recombinant plasmid DNA would co-migrate with, M_L the length of linearized vector DNA (kb) and M_I the length of insert DNA (kb). This formula allows an easier identification of a circularized recombined plasmid by comparing it with a DNA marker on the gel.

This is a pre-screening step which saves time and enzymes. Plasmid DNA from clones calculated to have the expected size was extracted by the miniprep method as described in Section I(2). The identity of the plasmids as recombinants was further confirmed by a restriction enzyme digestion or sequencing.

III. Purification of DNA Fragments

Two methods have been used to isolated DNA fragments. All of the procedures were modified from methods described by Sambrook *et al.* (1989).

1. Electroelution

Digested DNA was run on an agarose gel until the desired DNA bands were completely separated. A gel slice containing the DNA band was cut out, and sealed in a piece of dialysis tubing with a minimal volume of 1×TAE electrophoresis buffer (40 mM Tris-acetate, 10 mM EDTA; adjusted to pH 7.9 with glacial acetic acid). The tubing was put back to the electrophoresis tank and electrophoresis continued for around 30 min to run the DNA out of the gel and onto the inner wall of the tubing. To release the DNA from the tubing wall, a reversed electrophoresis was made for 5 seconds. The TAE solution containing the DNA fragment was collected , extracted with phenol:chloroform (1:1) twice, precipitated with 1/10 volume of 3 M NaAc pH 5.2 and 2.5 volumes of 95% ethanol overnight. The DNA fragment was pelleted by centrifugation at 14,000 rpm for 15 min and washed by 75% ethanol.

2. DEAE-cellulose membrane

Digested DNA was completely separated on an agarose gel. A piece of DEAE-cellulose membrane was inserted into the gel immediately in front of the desired DNA band. The electrophoresis was continued until the DNA of interest totally migrated onto the membrane. The membrane was rinsed at room temperature in 10 ml of low-salt wash buffer (50 mM Tris-HCl pH 8.0, 0.15 M NaCl, 10 mM EDTA), and immersed at 65°C for 30 min in a high-salt elution buffer (50 mM Tris-HCl pH 8.0, 1 M NaCl, 10 mM EDTA pH 8.0). The eluate containing the DNA of interest was extracted with phenol:chloroform (1:1) twice, and precipitated by centrifugation in the presence of 0.2 volume of 10 M ammonium acetate and 2.5 volumes of 95% ethanol.

IV. Maintenance of Bm5 cell line and Transfection Protocol

Bm5 cells [Grace, 1967], cultured in IPL-41 media (Life Technologies, Weiss *et al.* 1981) supplemented with 10% BSA at 28°C [Iatrou *et al.*, 1985], were seeded into 6-well culture plates (35 mm diameter) at a density of $1x10^6$ cells/well with 2 ml of medium one day prior to transfection [Zhang *et al.* 1994]. Plasmid DNA was purified by CsCl density gradient centrifugation. The helper and donor vectors were co-transfected into the cells using Lipofectin™ (BRL) according to the procedures described by the manufacturer. A transfection solution (500 µl/well) containing 0.5 µg/ml of pMCH1 (see below), along with 2.0 µg/ml of pBmA.MosI (see below) in basal IPL-41 medium containing 20 µg/ml Lipofectin™ (BRL/Gibco), was applied to the culture cell for 6

hours. Control cells were transfected with either pBmA.Mos1 or pMCH1 complemented with Bluescript DNA to total 5.0 μg/ml. After washing, the cells were left to recover in 2 ml of complete IPL-41 media for 3 days and then transferred to a 25 ml flask with 5 ml medium. To observe early transposition events (fig.1-1), the cells were collected one week after transfection. To observe long term transposition events, the cells were subcultured for 3 weeks to dilute the unintegrated plasmids, and then Hygromycin (HmB, Boehringer-Mannheim) (0.25 mg/ml) selection was applied for 6 weeks.

V. Nucleic Acid Isolation from Bm5 Cells

Bm5 cells transfected with relevant plasmids were collected by centrifuging at 3,000 rpm for 5 min in a benchtop microcentrifuge. The cells were washed three times in PBS (10 mM KH_2PO_4, 140 mM NaCl, 40 mM KCl) and stored at -70°C until further use.

Two million cells were resuspended in 400 μl of lysis buffer (7 M Urea, 135 mM NaCl, 1 mM EDTA, 2% SDS, and 10 mM Tris-HCl, pH 7.5), and extracted with phenol:chloroform (10:1 once, 1:1 twice) and chloroform (once). The extracted total nucleic acid was precipitated from 0.25 M ammonium acetate with 2.5 volumes of 95% ethanol. The integrity of the RNA was monitored by electrophoresis in an agarose gel. Total nucleic acid preparations usually contained about 90% RNA and 10% DNA from transfected cells.

VI. Plasmid Rescue Assay

Eight hundred ng of genomic DNA, extracted from the transfected cells and digested with EcoRI, was ligated using T4 ligase (BRL) in a 10 μl reaction. The ligation mix was purified with phenol/chloroform, and 1/10 aliquot of the ligated DNA was used to transform *E. coli* DH10B strain by electroporation. Transformation efficiency was approximately 10^8 colonies/μg DNA.

VII. Sequencing

1. "Classical" sequencing

The protocol for double-stranded plasmid DNA sequencing was modified from United States Biochemical. To prepare the template, 10 μg of purified plasmid DNA was denatured in 40 μl of 0.2 N NaOH and 0.2 mM EDTA at 37°C for 30 min. Then, 4 μl of 2.5 M ammonium acetate and 100 μl of ethanol were added and the solution was kept on dry-ice for 20 min. The DNA was precipitated at 14,000 rpm for 20 min in a microcentrifuge. The DNA pellet was dissolved in 7 μl of ice-cold sterile water for 15 min and 2 μl of 5× Annealing buffer (200 mM Tris-HCl pH 7.5, 100 mM $MgCl_2$, 250 mM NaCl) and 1 μl of sequencing primer (0.5 pmol) was further added. Template-primer annealing was achieved by heating the tube at 68°C for 2 min, and cooling it slowly to room temperature for about 30 min. To the annealed template mixture, the following were added for sequencing: 1 μl of 0.1 M DTT, 2 μl of 1× labelling buffer, 1.5 μM each of

dGTP, dCTP and dTTP, 5 μCi of ^{35}S-dATP and 2 μl of diluted Sequenase (1:8 dilution in 10 mM Tris-HCl, pH 7.5, 5 mM DTT, 0.5 mg/ml BSA). The labelling mixture was incubated for 3-5 min at room temperature. To set up the termination reaction, 3.5 μl aliquots of the labelling reaction were transferred to four tubes labelled G, A, C, T, each containing 2.5 μl of the termination solutions (8 μM of one ddNTP, 80 μM each of the three other dNTPs and 50 mM NaCl). After incubation for 15 min at 37°C, the reactions were stopped by adding 4 μl of stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol FF). One quarter of the 10 μl reaction was analysed on a 6% polyacrylamide sequencing gel.

2. "Cycle" sequencing

Cycle sequencing [Fajas *et al.*, 1997; Yanakawa *et al.*, 1997] of the rescued plasmids was performed on the ABI Prism 310/377 DNA sequencer (Applied Biosystems) using the Terminator Ready Reaction Kit (TRRM; PE Applied Biosystems) according to the manufacturer's protocol. The same primers were used as the ones used in the PCR screening. The reaction mixture contained 4 μl of TRRM, 0.5-1 μg of template DNA, 50 ng of primer DNA, and distilled water to a total volume of 10 μl. The condition was set at 96°C for 30 seconds, 50°C for 15 seconds, 60°C for 4 min, and total 30 cycles. The reaction mixture was added to 20 μl of dH_2O, 3 μl of 3 M NaAC, 75 μl of 95% ethanol, and stored on ice for 15 min. The DNA was spun down at 14,000 rpm for 10 min on a microcentrifuge and resolved on the sequencing device.

VIII. Preparation of Radio-labelled ^{32}P Probes

1. Hexanucleotide primers

This process was performed as described by Feinberg and Vogelstein (1983, 1984). Thirty two μl of solution containing 0.2-0.5 μg of DNA was boiled for 6 min, and cooled rapidly on ice. The following components were then added: 10 μl of 5× OLB buffer (see below), 2 μl of 10 mg/ml BSA, 5 μl of α-^{32}P-dCTP (3,000 Ci/mmole, 10 mCi/ml; Dupont) and 1 μl (2-3 units) of large fragment (Klenow) of DNA polymerase I. The mixture was incubated overnight at room temperature. The probe was purified from free nucleotides by Sephadex G75 spin-column centrifugation. The probe specific activity was usually about 5×10^8 cpm/μg.

Five times OLB buffer is Solution A: Solution B: Solution C (100:250:150). Solution A contains 1 ml solution O (1.25 M Tris-HCl pH 8.0 and 0.125 M $MgCl_2$), 18 μl β-mercaptoethanol, 5 μl each of 0.1 M dATP, dTTP and dGTP (previously dissolved in 3 mM Tris-HCl pH 7.0 and 0.2 mM EDTA). Solution B contains 2 M Hepes pH 6.6 adjusted with 4 M NaOH. Solution C contains random hexanucleotides (Gibco/BRL) in 3 mM Tris-HCl pH 7.0 and 0.2 mM EDTA at 90 OD_{260}/ml.

2. PCR labelling

This process was performed as described by Schowalter & Sommer (1989). The reaction mixture contained 0.1-10 ng DNA template, 100 ng of forward and reverse primers, 25 μM of dATP, dTTP and dGTP, 1μl of 10×PCR buffer (500 mM KCL, 100

mM Tris HCl, pH9.0 at 25 °C, 1% Triton X-100, 15 mM $MgCl_2$), 5 μl of α-^{32}P-dCTP (3,000 Ci/mmole, 10 mCi/ml; Dupont) and 0.2 μl Tsg/pfu DNA polymerase mix (Tsg:pfu=10:1, totally 1-2 Units). The PCR cycles were carried out as follows: 94°C preheating for 4 min, then 94°C 1 min, 60°C 1 min, and 74°C 1 min with 30 cycles, finished by extension at 72°C for 8 min. The probe was purified from free nucleotides by Sephadex G75 spin-column centrifugation.

IX. Southern Blot Hybridization

Digested DNA fragments were resolved on a 1% (W/V) agarose gel with EtBr. Following photography, the gel was briefly depurinated for 10 min in 0.4 N HCl, neutralized twice in distilled water, and denatured twice for 20 min each in 0.4 N NaOH and 1 M NaCl. The DNA fragments were transferred by capillary action to a Hybond™ N^+ nylon membrane (Amersham) with 0.4 N NaOH and 1 M NaCl. After transfer overnight, the membrane was rinsed in 6× SSC and baked under vacuum at 80°C for 1 hour.

The membrane was prehybridized for at least 3 hours in a hybridization mixture containing 0.3 M NaCl, 50 mM sodium phosphate pH 7.0, 5× Denhardt's (Denhardt, 1966), 10% dextran sulphate, 1% SDS, 5 mM EDTA and 2.5 mg/ml total yeast RNA. Hybridization was carried out at 68°C for 16 hours with about 5×10^5 cpm of labelled probes per ml of hybridization solution. Following hybridization, the membrane was washed at 68°C with 2×, 0.5× and 0.1× SSC containing 0.1% SDS twice for 20 min each

time. Finally, the membrane was autoradiographed at -70°C using a Kodak X-ray film with intensifying screens (Dupont Cronex lightning-Plus).

X. Northern Blot Hybridization

Total RNA was extracted from transfected cells as described previously (Section V). Five μg aliquots of total RNA were resolved on a 1% agarose gel containing 10 mM methylmercuric hydroxide [Bailey and Davidson, 1976] and immobilized onto a Hybond-N^+ membrane by overnight capillary transfer with 3 mM NaOH. Hybridization to relevant radio-labelled probes was carried out as described in Section 9 for Southern hybridization except that 1×10^6 cpm of each probe was added per ml of hybridization solution.

XI. Protein Extraction

Transfected cells were harvested , pelleted at 3,000 rpm for 2 min in a microcentrifuge and washed three times with 1 ml PBS (8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4 and 0.24 g KH_2PO_4 in 1 L H_2O, pH 8.0). The pellet was resuspended in 100μl of 0.25 M Tris-HCl pH 7.8 and the cells were lysed by freeze-thawing three times. The debris was pelleted by centrifugation for 10 min at 14,000 rpm in a benchtop microcentrifuge. The supernatant containing soluble protein were transferred to a fresh tube. Protein concentration was quantified by Bradford (1976) using the BioRad protein assay reagent and bovine serum albumin (BSA) as standard.

XII. Chloramphenicol Acetyltransferase (CAT) Assay

CAT assays were carried out as previously described [Iatrou *et al.*, 1989]. Ten µg of soluble protein was brought to 100 µl with 0.25 M Tris-HCl pH7.8 and mixed with 20 µl of 8 mM acetyl CoA (Pharmacia) and 20 µl of 1,2 ^{14}C-chloramphenicol (0.08 µCi, 108 mCi/mmol; ICN). The reaction was incubated at 37°C for 2 hours and stopped by extraction in 500 µl of ethyl acetate. The organic phase was removed to a new tube, lyophilized and dissolved in 15 µl of ethyl acetate. The solution was spotted on a TLC plate (20×20 cm, Silica gel 60; E Merck) and developed with chloroform:methanol (95:5). The TLC plate was exposed to a Kodak X-film at room temperature overnight.

CHAPTER ONE

THE MARINER TRANSPOSABLE ELEMENT FOR GENETIC TRANSFORMATION OF *Bombyx mori*

I. Introduction

The functional analysis of developmental processes in non-drosophilid insects has been hampered so far by the lack of techniques for genetic transformation. Recently two major strategies are being followed to introduce foreign gene constructs into insect cells; one involving genetically engineered viruses [Vile & Russell, 1995; Iatrou *et al.*, 1995; Welsh *et al.* 1997], the other one employing transposable elements [O'Brochta *et al.*, 1996]. Transposable elements can move from one position on the host DNA to another, and therefore other genes that are incorporated into these elements can be moved together into new locations. This feature makes mobile DNA element a very powerful tool in both genetic study and gene engineering.

Among mobile DNA, Mariner elements seem to have great potential for being developed as major tools for genetic transformation of many insect species. First of all, Mariner-like elements are widespread in the genomes of many insects [Robertson, 1993], including *B. mori* [Robertson *et al.*, 1997, Tomita *et al.*, 1997]. Secondly, functional Mariner elements encode a protein --- a transposase, which is capable of introducing foreign DNA into target sequences in the absence of any other host-specific factors [Vos *et al.*, 1996, Lampe *et al.*, 1996]. This indicates that Mariner-like elements could be functional in any host cell, as long as inhibitory factors are absent. Lastly, genetic transformation using Mariner elements has been achieved in dipteran insects such as *D. melanogaster* [Garza *et al*, 1991] and the medfly *Ceratitis*

capitata [Loukeris *et al.*, 1995]. Moreover, Mos1 Mariner, identified and isolated from *D. mauritiana*, has been shown to successfully transform the protozoan *Leishmania* [Gueiros-Filho *et al.*, 1997].

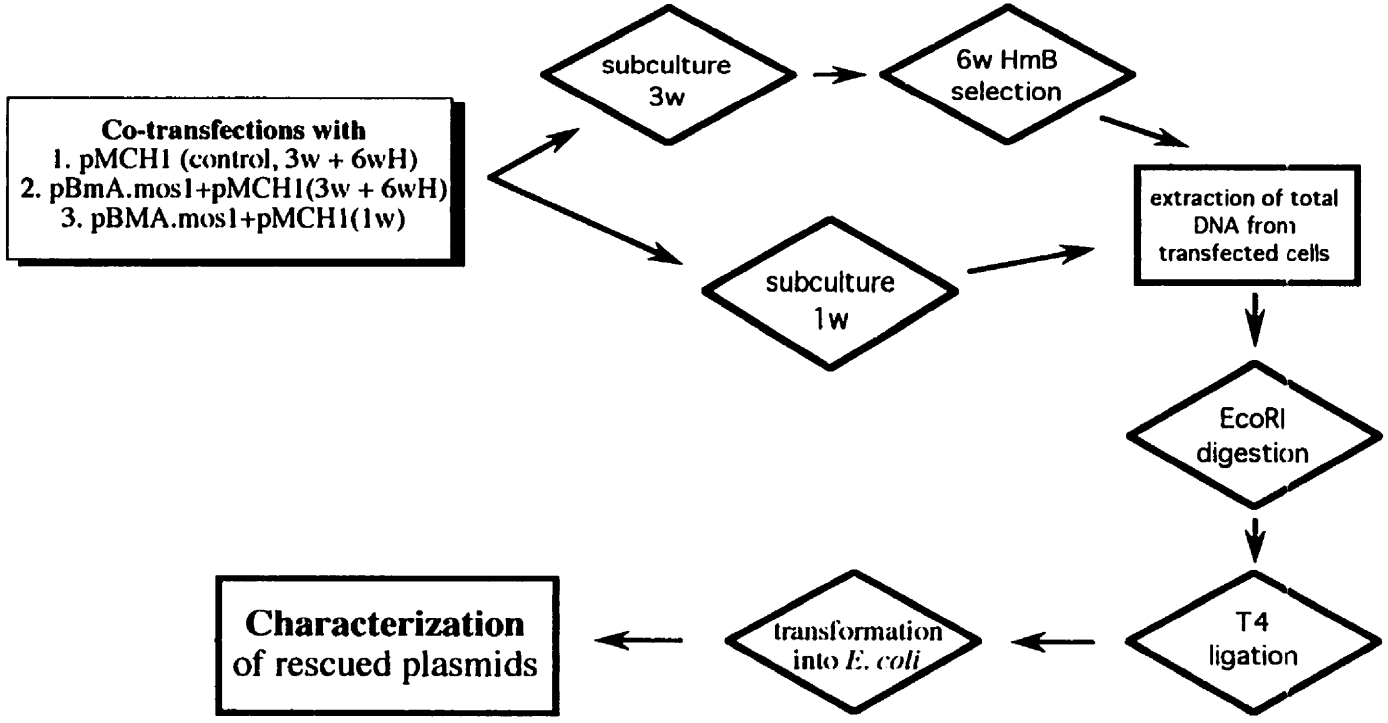
Mariner elements belong to the DNA-mediated class of transposable elements [Kidwell, 1993]. Most members of the Mariner superfamily characterized so far carry mutated transposase genes and hence remain in a non-mobile status [Lohe *et al.*, 1995]. However, in the presence of an active complementing transposase, inactive mariner elements can move from one location to another [Lidholm *et al.*, 1993]. Therefore, such a system is capable of introducing exogenous DNA into host genomes as a versatile genetic vector.

Among the Mariner element superfamily, Mos1 is one of the few members capable of performing autonomous germline transformation at a TA dinucleotide target site in the host genome [Medhora *et al.*, 1988, 1991, Maruyama *et al.*, 1991]. The element is a 1286 bp of DNA containing a pair of 28 bp terminal inverted repeats (IRs) and a single open reading frame (ORF) encoding the transposase [Jacobson *et al.*, 1986]. The DNA and amino acid sequences are shown in Appendix 1. It transposes through a "cut-and-paste" mechanism [Plasterk, 1996; Hartl *et al.*, 1997]. The element-encoded transposase excises the transposon from the flanking DNA. The resulting 3' hydroxyl at the free ends of the element attack a TA dinucleotide target site which is bound by the transposase, and create a covalent bond between the element and the target site in a strand transfer reaction [Mizuuchi, 1992]. On the other hand, it has been found that non-functional Mariner-like sequences exist in the

Bombyx genome [Robertson et al., 1996]. This evidence strongly suggests that the active Mos1 transposase should be able to function properly in silkworm.

Based on the above studies, I employed the Mos1 Mariner element to establish a genetic transformation system in Bm5 culture cells, a *Bombyx* cell line of ovarian origin. In this system, a helper vector was used to express an active Mos1 transposase, and a donor vector carried a pair of Mos1 IRs with an antibiotic selection and a reporter gene introduced between them. After both plasmids were co-transfected into cultured cells, the expressed Mos1 transposase recognized the Mos1 terminal inverted repeats, excised the defective Mos1 element along with the foreign genes, and transferred the fragment into available TA dinucleotides sites in the host genome. Either before or after antibiotic selection, total DNA was collected from the surviving cells. EcoRI fragments containing pBS-SK, 3' IR and the flanking sequences were subsequently rescued and sequenced (see fig.1-1 for the outline of the experimental design). My results demonstrate that the Mos1 element is capable of conducting transposition in the *Bombyx* genome, and such transposition does not require any specific host elements. In other words, the presence of a high copy number of endogenous Mariner-like element in the host genome does not prevent *D. mauritiana* Mos1-mediated transposition.

Figure 1-1. The outline of the experimental approach. (the numbers 1, 2, 3 represent the control, the long term, and the short term transfection groups respectively.).


Co-transfections with
1. pMCH1 (control, 3w + 6wH)
2. pBmA.mos1+pMCH1(3w + 6wH)
3. pBMA.mos1+pMCH1(1w)
subculture 3w
6w HmB selection
subculture 1w
extraction of total DNA from transfected cells
EcoRI digestion
T4 ligation
transformation into E. coli
Characterization of rescued plasmids

II. Materials and Specific Methods

1. Transformation Vectors

The Mos1 expression vector (pBmA.Mos1, Fig.1-2A) consisted of the Mos1 ORF inserted into the HindIII site of the *Bombyx* actin expression cassette (pBmA plasmid from Johnson and Iatrou, 1992), thus placing the Mos1 transposase gene under the transcriptional control of the constitutive actin promoter of *B. mori*. The Mos1 ORF was obtained as a 1.5 kb SspI/HindIII fragment from pBS.Mos1 [Lidholm *et al.*, 1993; Appendix 1]. The junction of the cloning site was sequenced to confirm the correct orientation of the insertion.

The Mos1 donor vector pMCH1 was constructed by replacing the *Drosophila* reporter gene in vector pMIwB [Lidholm *et al.*, 1993] with the hygromycin B [HmB; Gritz & Davies, 1983] selective gene and chloramphenicol acetyl transferase [CAT; Crabb et al., 1989] reporter gene, both of them driven by the silkmoth actin promoter (Fig. 1-2B). To obtain the tandemly arranged genes, a 3.6 kb SacI fragment containing an actin-HmB expression cassette [Farrell et al., *in press*] from pBmA.hmB was inserted into the Sal I site in pBmA.cat [Johnson *et al.*, 1992], the resulting in pBmA.CH. A 6.5 kb SacI fragment of the Cat-HmB expression cassette from pBmA.CH (Fig.1-3A) was then fused with the 7.4 kb SacI fragment of pMIwB (Fig. 1-3B), resulting in pMCH donor vector (Fig. 1-3B). Hence, in the pMCH1 vector (Fig.1-3B, see also Fig. 1-7), the *cat* and *hmB* genes, as well as Bluescript KS$^-$, the P transposase promoter and the 5' portion of *E. coli* LacZ gene inherited from

Figure 1-2. The physical maps of the Mariner transformation vectors. **A:** pBmA.Mos1 plasmid --- Mos1 helper vector (Mos1 expression vector). The Mos1 ORF (the thick arrow) was inserted into HindIII site in the pBmA expression cassette (the open box with the arrow inside to show the expression orientation), and was brought under the transcriptional control of the actin promoter. **B:** The physical map of pMCH1--- donor vector (transformation vector). It contains a pair of mariner IRs sandwiching 500 bp of *D. mauritiana* flanking sequences, Bluescript KS, HmB and CAT gene expression cassettes, followed by residues of LacZ and a truncated P-element. The *cat* and *hmB* genes are shown with the arrows, while the Mos1 element is split into 5' and 3' parts and reversed in orientation (shown with thick arrows).

Mariner Transformation vectors

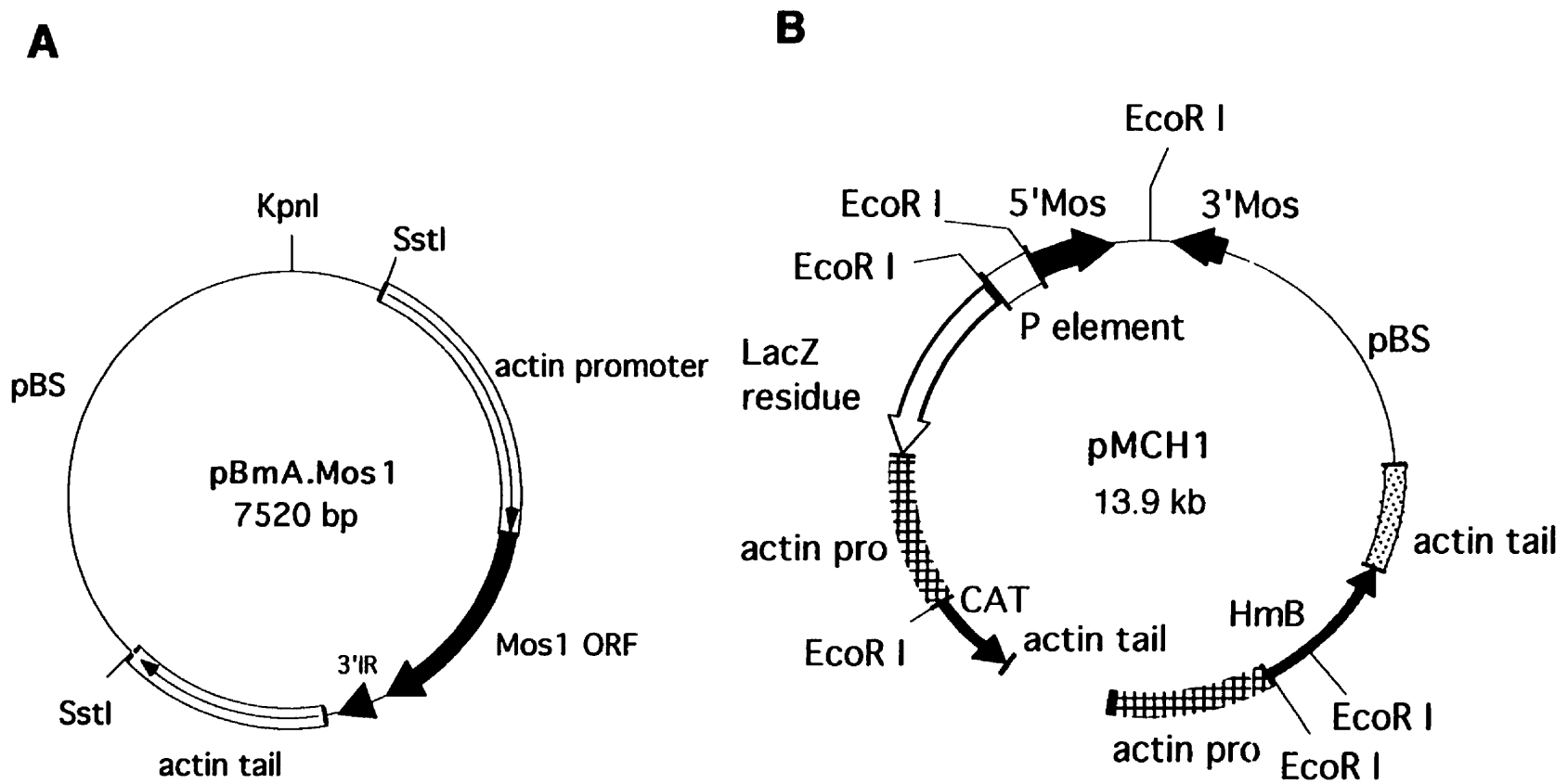
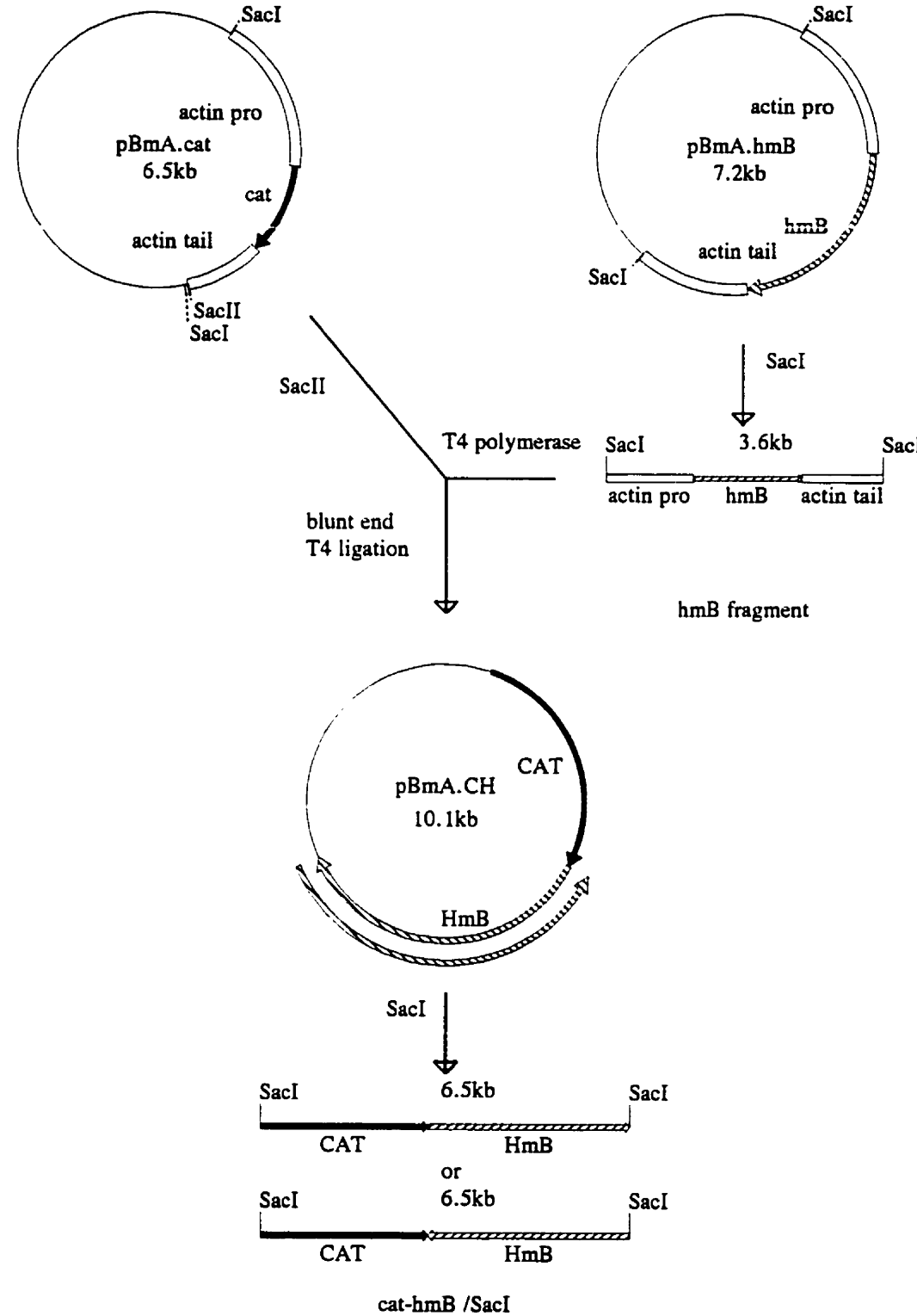
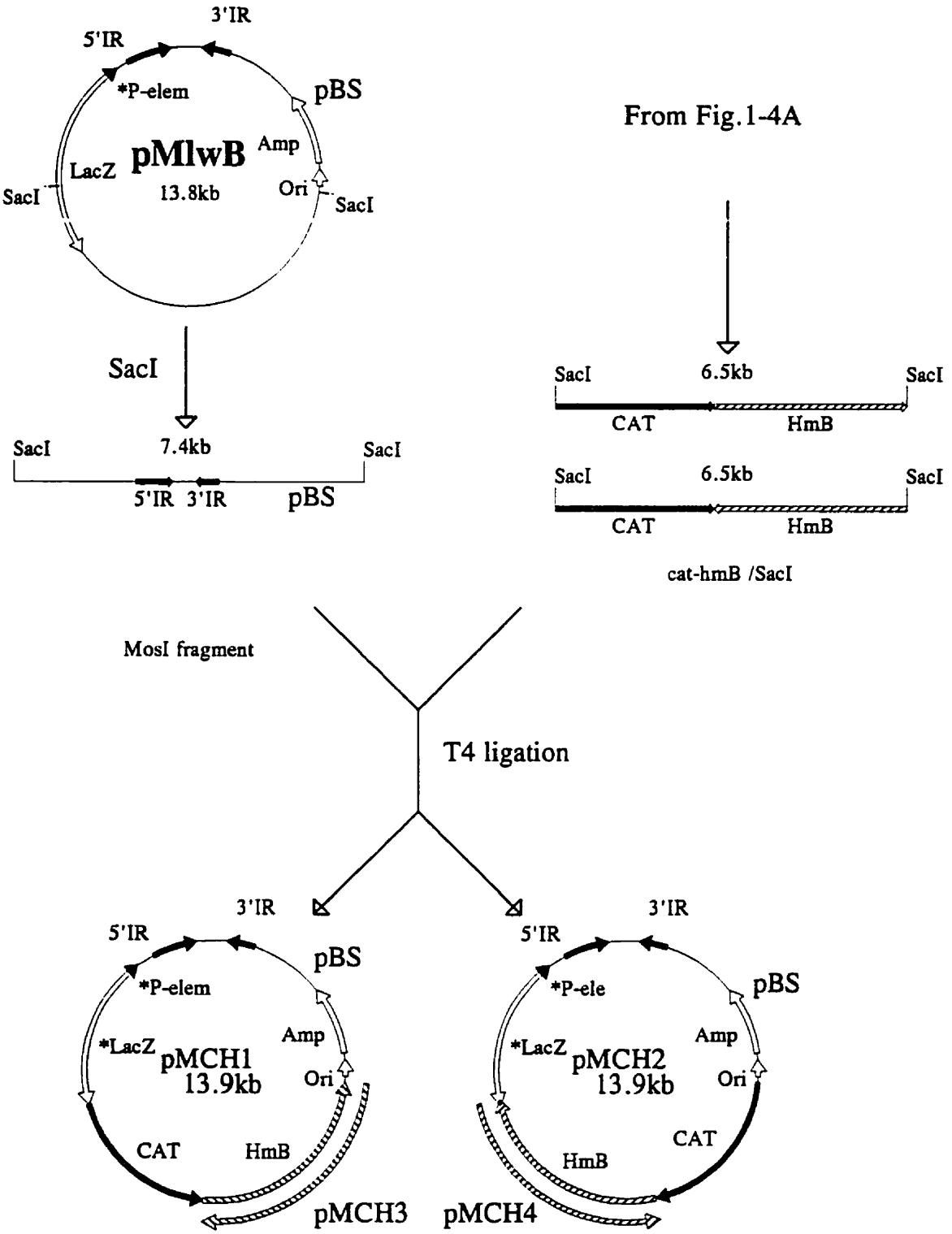


Figure 1-3. Construction of the pMCH transformation vectors. The plasmids were created from the *Drosophila* mariner transformation vector pMlwB by replacing a SacI fragment containing *Drosophila* reporter gene (eye color *white* and LacZ) with an array of reporter (*cat*) and selection (*hmB*) genes for expression in *Bombyx*. A: construction of the plasmid pBmA.CH containing the array of cat and hmB genes. B: replacement of the reporter gene in pMlwB with *Bombyx* gene expression cassettes. Amp gene and Ori site in pBS: open arrows; Actin promoter and terminator: open boxes; *cat* gene: shaded arrows; *hmB* gene: slant arrows; CAT and HmB expression cassettes are simplified as single arrows for clarity in the subsequent steps. Lac Z: open arrow, residue of the P-element (P-ele): shaded arrowhead. "*" indicates the residue of the gene or the sequence. The inverted repeats of the Mariner element (5' IR and 3' IR): solid arrows. See Materials and Methods for a detailed description of the cloning procedure.

A:


SacI
actin pro
pBmA.cat
6.5kb
cat
actin tail
SacII
SacI
SacI
actin pro
pBmA.hmB
7.2kb
hmB
actin tail
SacI
SacII
T4 polymerase
blunt end
T4 ligation
SacI
3.6kb
SacI
actin pro
hmB
actin tail
hmB fragment
CAT
pBmA.CH
10.1kb
HmB
SacI
SacI
6.5kb
SacI
CAT
HmB
or
SacI
6.5kb
SacI
CAT
HmB
cat-hmB /SacI

B:


3'IR
5'IR
*P-elem
pBS
Amp
LacZ
pMlwB
13.8kb
Ori
SacI
SacI
SacI
7.4kb
SacI
5'IR 3'IR
pBS
From Fig.1-4A
SacI
6.5kb
SacI
CAT
HmB
SacI
6.5kb
SacI
CAT
HmB
cat-hmB /SacI
MosI fragment
T4 ligation
3'IR
5'IR
pBS
*P-elem
*LacZ
pMCH1
13.9kb
Amp
Ori
CAT
HmB
pMCH3
3'IR
5'IR
pBS
*P-ele
*LacZ
pMCH2
13.9kb
Amp
Ori
CAT
HmB
pMCH4

pMlwB vector, were bracketed by a pair of Mos1 IRs, the target sites for the transposase. In addition, both the *cat* and *hmB* genes were driven by the *Bombyx* actin promoters. Several sets of different donor vectors were also constructed but with a different selective gene (Neo) at different orientations in the vectors (data not shown but see table 1-1 for details).

2. PCR Reactions and Sequencing

To distinguish between plasmids rescued from genomic DNA and unintegrated plasmids, a PCR screening assay was conducted among the isolated transformants. The forward primer was derived from the 3' end of Mos1 element (nucleotides 1176-1194 bp in Mos1 element [Jacobson *et al.*, 1986], 5' ATGTGTAGCTAGCGACGGC 3'). Two reverse primers were used. One of them, T7, was derived from the pBS.KS$^-$ portion of the rescued plasmids. The other one was located at the EcoRI site in the middle of the Hmb gene sequence [Gritz & Davies, 1983; 5' AGGCTCTCGCTGAATTC 3']. The primer locations are indicated in Fig. 1-4.

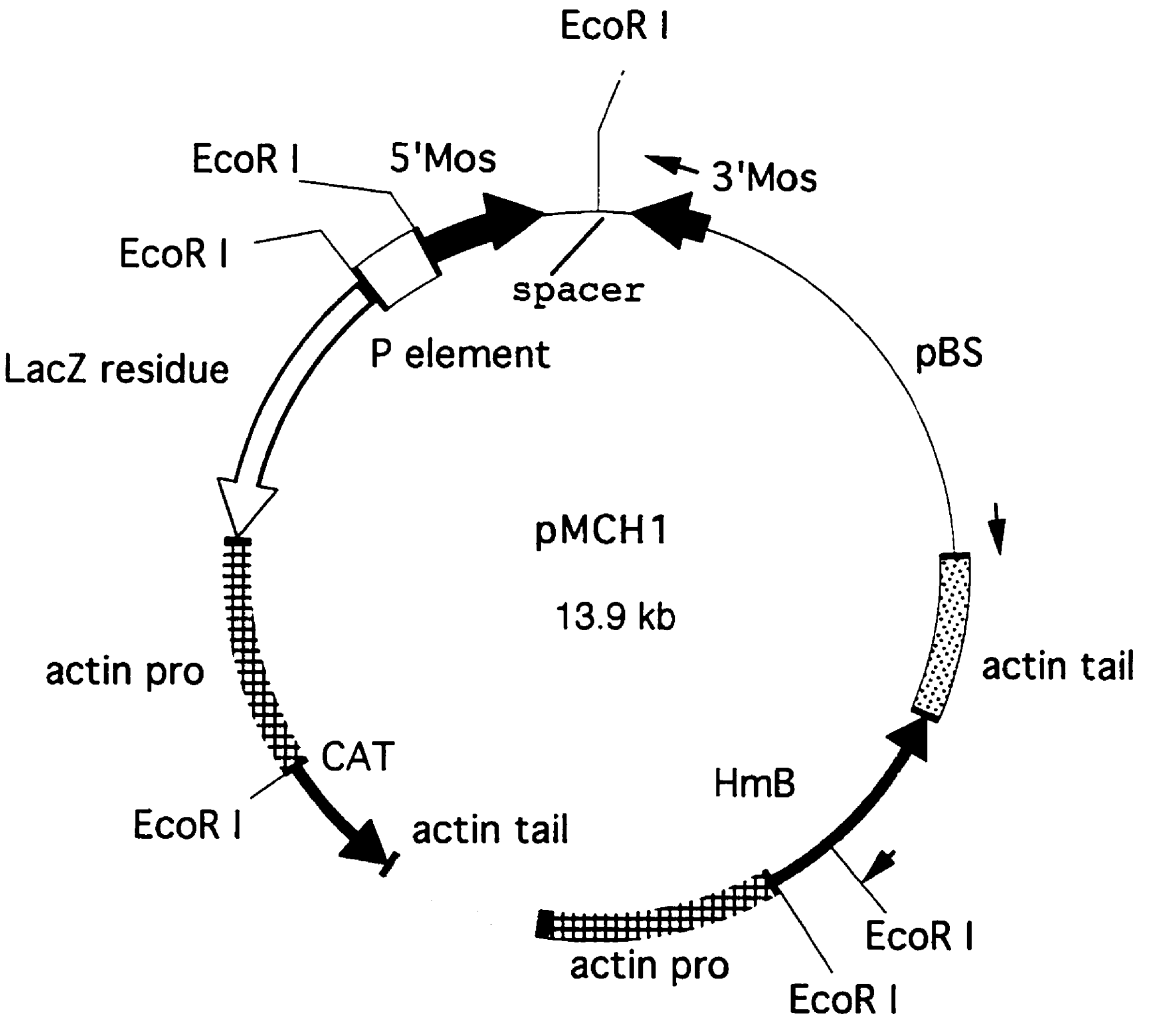
For the plasmids yielding PCR products different than those obtained from the intact pMCH1, their insertion fragments were sequenced from 3' IR (Fig. 1-4). For those plasmids whose sequences were unknown or showed homology to unique *Bombyx* genes, a genomic PCR reaction was performed to examine if the sequences obtained from *Bombyx* genomic DNA were the same as those from the rescued plasmids. Each primer pair was designed to specifically amplify each insertion fragment (Appendix 2). The PCR conditions were 94 °C, 1 min; 58 °C, 1 min;

Table 1-1. The Components of the Mos1 Transformation Vectors

A set of 7 Mos1 donor vectors was constructed according to their conponents and orientations. The three columns in middle show what genes the plasmid contained, the column at the right indicates the relative positions and orientations of these genes, referring the plasmid map shown in Fig. 1-3B.

No.	plasmid	Reporter actin-CAT (⬅)	Selective Genes		gene arrangement
			ie1-Neo (⇦)	actin-HmB (←)	
1	pMCN2	✓	✓		⬅⇦
2	pMCN8	✓	✓		⬅⇨
3	pMCN9	✓	✓		⇦ ⬅
4	pMCH1	✓		✓	➡ →
5	pMCH3	✓		✓	➡ ←
6	pMCH2	✓		✓	← ⬅
7	pMCH4	✓		✓	→⬅

Figure 1-4. The PCR primer locations on donor vector pMCH1. These primer pairs were used in the PCR screening to distinguish the rescued plasmids from the unintegrated donor vector pMCH1. The arrows show the primers in 5'-3' direction. The 500bp "Spacer" region between 5' and 3' IRs was *D. mauritiana* genomic DNA and inherited from plasmid pMlwB (see Fig. 1-3B).


EcoR I
EcoR I
5'Mos
3'Mos
EcoR I
spacer
LacZ residue
P element
pBS
pMCH1
13.9 kb
actin pro
actin tail
CAT
HmB
EcoR I
actin tail
EcoR I
actin pro
EcoR I

74°C, 1 min. The PCR products were then separated by electrophoresis and each of them sequenced by cycle sequencing [Fajas *et al.*, 1997; Yamakawa *et al.*, 1997] according to the procedure recommended by the manufacturer, using the corresponding primer pairs.

3. Southern Blot Analysis

Southern hybridizations of genomic DNA were carried out as previously described [Fotaki and Iatrou 1988]. In the analysis to determine the post-transfection fate of the plasmid within cell lines, the genomic DNA from the transfected cells was probed with either a 0.8 kb BamHI fragment which derived from *cat* gene ORF in pBmA.cat [Johnson & Iatrou, 1992], or with a 0.7 kb SphI/SspI fragment from pBS.mos1 [Lidholm et al., 1993]. Labelling of the fragments was carried out by the random oligo-labelling technique [Feinberg & Vogelstein, 1984].

In the analysis to confirm that the rescued fragments were indeed coming from the Bombyx genome, 3 μg of the control Bm5 genomic DNA digested with EcoRI was probed with the fragments from the rescued plasmids. The radioactive probes were synthesized by PCR [Schowalter and Sommer, 1989]. The primer pairs were the same as the ones used in the genomic PCR.

III. Results

1. Donor Plasmid pMCH1 — the Substrate DNA for Mos1 Transposase

Our Mos1 mediated gene transformation system consisted of a helper and a donor vector. The helper vector (pBmA.Mos1; Fig.1-2A) provided active Mariner transposase to mobilize the defective Mos1 element in the donor vector (pMCH1; Fig. 1-2B) to the host genome. Thus the *cat* and *hmB* genes that were inserted into the defective element in pMCH1 would be co-transposed into the same sites.

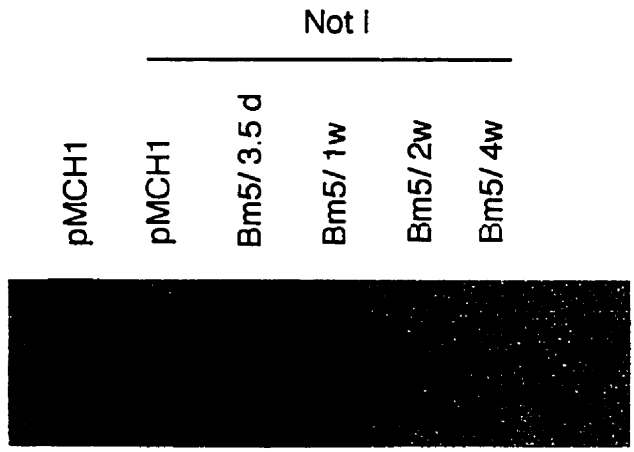
The donor vector pMCH1 (Fig.1-2B) was derived from pMlwB, and contained *cat* and *hmB* genes, Bluescript and residues of the P and LacZ gene from pMlwB; all of them bracketed by a pair of Mos1 inverted repeat termini. The upstream Mariner sequences were retained, since they may be required for transposition [Lohe, *et al.*, 1996]. The Bluescript component was located at the 3' end of the donor vector and was essential for direct cloning of adjacent genomic sequences during the plasmid rescue process. The vector pMCH1 also carried approximately 600 bp of *D. mauritiana* DNA between the IRs (Fig. 1-2B and 1-4).

2. The Post-transfection Fate of Vectors Within the Cells

Prior to the transformation experiment with pMCH1 (Fig. 1-2B and 1-4), we tested the stability of the transfected plasmid in the cells in absence of selective pressure. Plasmid pMCH1 was transfected into Bm5 cells and cell samples were collected on a weekly base for up to 5 weeks. Southern blots were conducted on total

Figure 1-5. Southern blots of total DNA from transfected Bm5 cells collected at 3.5 days, 1 week, 2 weeks and 4 weeks post-transfection, and probed with the 0.8 kb *cat* gene fragment from the plasmid pMCH1. The genomic DNA was digested with NotI. **Panel A**: the autoradiography of the Southern blot. The left two lanes: pMCH1 DNA control and the same DNA treated with NotI used as controls. The right four lanes: genomic DNA samples from the transfected Bm5 cells collected at different time points and digested with NotI; d: days post-transfection; W, weeks post-transfection. **Panel B**: The relative intensity of the hybridization signal from the transfected genomes vs. time post-transfection.

A:


Not I
pMCH1
pMCH1
Bm5/ 3.5 d
Bm5/ 1w
Bm5/ 2w
Bm5/ 4w

B:

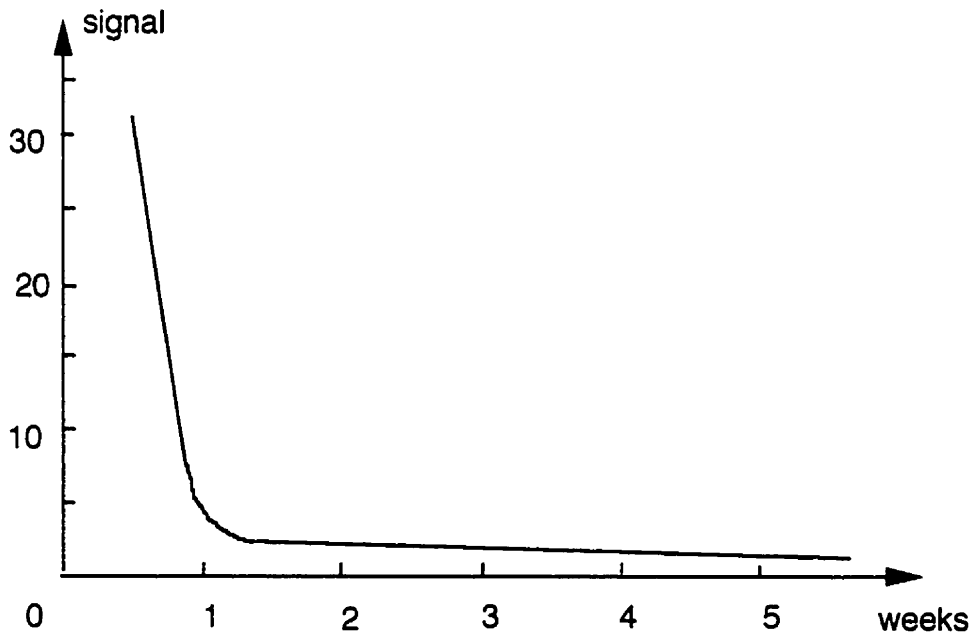
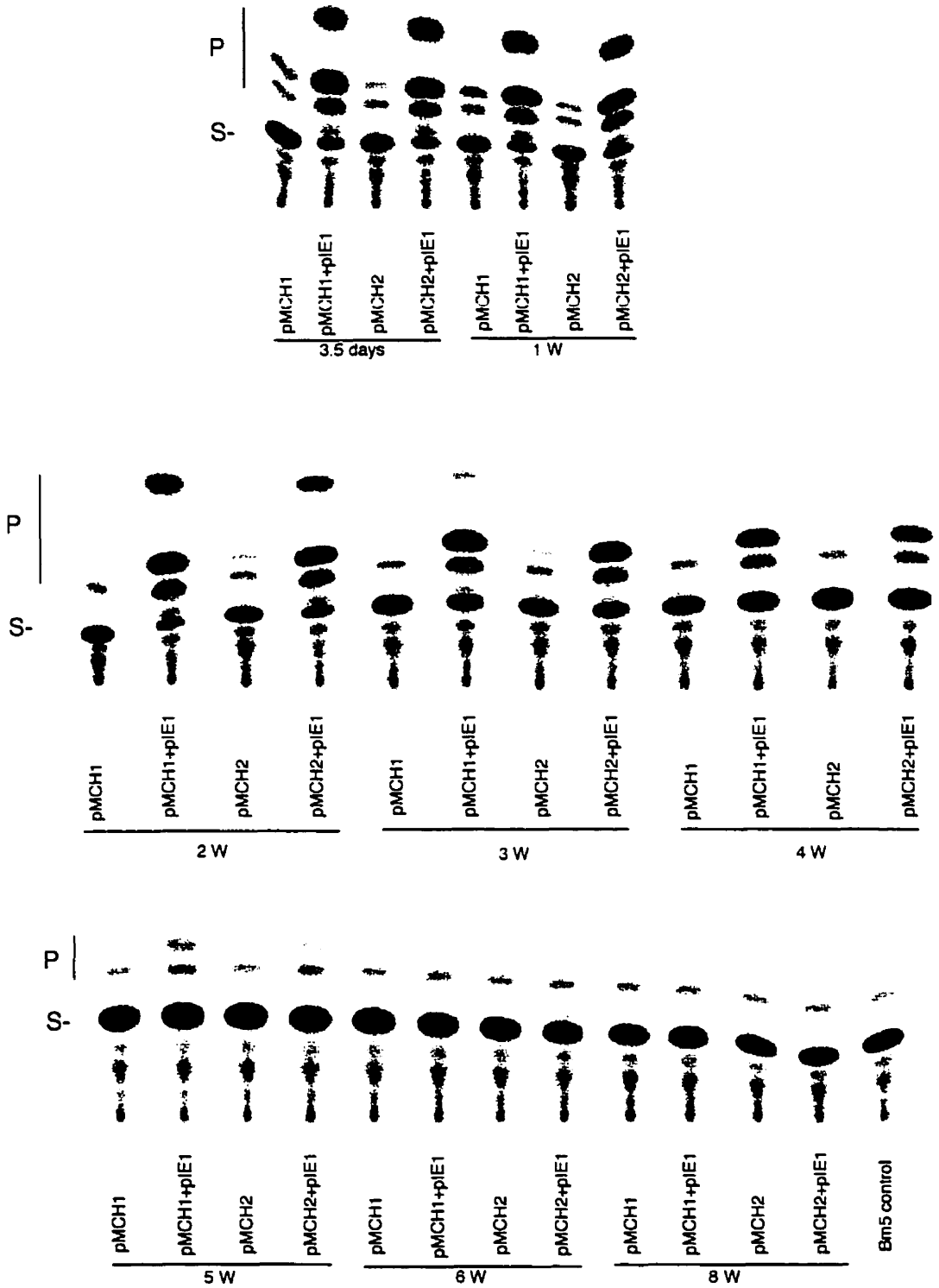

signal
30
20
10
0
1
2
3
4
5
weeks

Figure 1-6. CAT assays using extracts from transfected Bm5 cells. The cells were transfected with pMCH1 alone or pIE1 + pMCH1, respectively, and were collected through a time course post-transfection (weeks). In a parallel group, plasmid pMCH1 was replaced by pMCH2, which shares the same structural components with pMCH1, but with opposite direction of the *cat* and *hmB* genes. Plasmid pIE1 contains an IE1 ORF driven by the actin promoter, providing an activating function to the expression of pMCH in Bm5 cells. Each sample contained 10 μg of soluble protein from the transfected cells. The CAT activity declined alone with the time course and no difference was observed between pMCH1 and pMCH2. S: substrates; P: acetylated products.


P
S-
pMCH1
pMCH1+pIE1
pMCH2
pMCH2+pIE1
3.5 days
1 W
2 W
3 W
4 W
5 W
6 W
8 W
Bm5 control

DNA isolated from these cells and linearized by digestion with NotI (Fig.1-5). The gradually fading band representing the pMCH/NotI 14.9 kb plasmid indicated that the intact pMCH1 plasmids were diluted during a time course, but they were still detectable 5 weeks post-transfection, which was the end of our cell collection.

In addition, CAT assays were carried out to determine whether the reporter gene in the donor vector can function properly in the *Bombyx* host cell line (Fig. 1-6). Bm5 cells were transfected with pMCH1 or with pMCH and pIE1. No antibiotic selection was applied on the cells. Plasmid pIE1 expresses the IE1 gene in Bombyx cells, which can enhance actin promoter activity 100-fold [Lu & Iatrou, 1997]. Since plasmid pMCH1 contained *cat* gene driven by actin promoter, our CAT assays showed that the presence of IE1 element increased the CAT activity in pMCH1. As expected, CAT activity declined along the time course, but was still detectable even 6 weeks post-transfection. The same results were achieved from the parallel transfections with pMCH2 or with pMCH2 and pIE1 (Fig. 1-6; pMCH2 differed from pMCH1 only in the orientations of *hmB* and *cat* genes, see Fig. 1-3B).

The CAT assays proved that the genes in our constructed plasmids were functional in Bm5 cells and that the copy numbers of the transfected plasmids in the host declined in absence of selective pressure, which was consistent with the Southern results (Fig. 1-5). Although some plasmids may have integrated into the host genome, given the fact that the transfected DNA persists as an extrachromosomal entity even after 6 weeks post-transfection, CAT assay cannot constitute a reliable assay for plasmid integration.

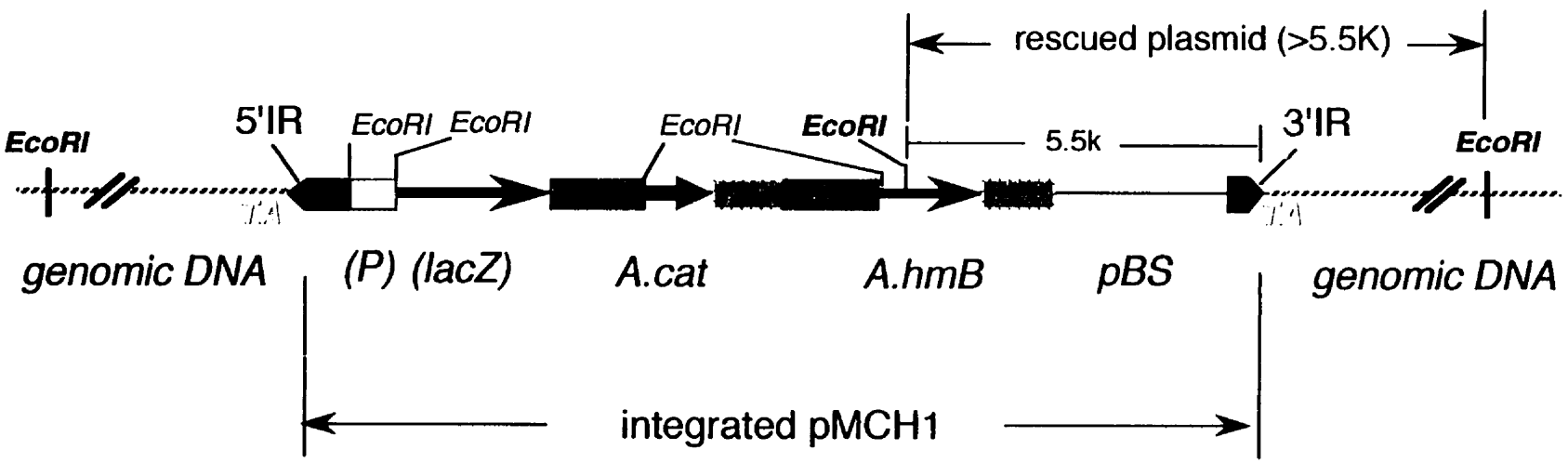
3. Plasmid Rescue Assays

Because of the presence of high copy numbers of endogenous Mariner-like elements [Robertson *et al.*, 1996] in the *Bombyx* genome, and the long-lasting unintegrated plasmids in Bm5 cells, it would be very difficult to detect *in vivo* Mos1 transposase activity (from the integration of the donor vector) on the basis of CAT assays. For this reason, we decided to undertake plasmid rescue experiments. The predicted structure of an integrated copy of the pMCH1 vector is shown in Fig. 1-7. Upon cleavage of the Mos1 transposase, pMCH1 vector should yield two fragments --- a major fragment bracketed by the two IRs, and a small "spacer" fragment flanking the IRs. The linearized defective Mos1 element should then be inserted into a TA site in the *Bombyx* genome. Therefore, the sequences in the insertion site from 5'-3' should contain the *Bombyx* genomic DNA, a TA dinucleotide, the 5' IR of Mos1, the residues of P element and LacZ genes which were inherited from pMlwB, the CAT and HmB expression cassettes, pBS, the 3' IR of Mos1, a TA dinucleotide and the Bombyx genomic DNA (Fig. 1-7). Most importantly, the integrated plasmid sequences should lack the *D. mauritiana* spacer DNA and the EcoRI site [EcoRI(a), Fig.1-8] located in it.

Genomic DNA of transfected cells was treated with EcoRI, and this was followed by ligation, transformation into *E.coli* and plating on ampicillin plates (the "plasmid rescue" process). In this process, only EcoRI fragments containing a pBS structure would be rescued. If the EcoRI site in genomic DNA was not located

Figure 1-7. The illustration of Mos1 mediated pMCH1 transposition in *B. mori* genomic DNA. The arrow line beneath the map shows the integrated pMCH in Bm genome. The arrow line above the map shows the basic structure of the rescued plasmid in conjunction with the genomic insertion site and the integrated donor vector sequences. The red arrow-heads, Mariner inverted terminal repeats; the blue boxes, Mariner element; pink arrow, hmB gene ORF; red arrow, cat gene ORF. The brown boxes represent actin promoter areas, while the brown dot boxes represent actin terminator regions. The genes in braces are non-functional: the yellow box represents p-element residue; the green arrow represents the Lac Z residue. Both of the genes are inherited from pMlwB.

Mariner-Specific Integration of pMCH1



immediately adjacent to the TA dinucleotide at the 3' IR end, the rescued plasmid should be larger than 5.5 kb (Fig. 1-7). Sequencing of the rescued plasmids would automatically distinguish the integrated Mos1 from extrachromosomal donor plasmid, because their DNA sequences downstream of the 3'IR would be different. In the latter case, the 3'IR sequences would be followed by the *D. mauritiana* spacer sequences present in the donor vector (Fig. 1-4). Thus, the "signature" of a Mos1 mediated transposition event would be a junction between the 3' IR and the genomic sequence that would contain a TA dinucleotide immediately downstream of an intact 3' IR terminal sequence [Lohe et al., 1996a, 1996b]. In fact, we did observe such structures by sequencing the rescued plasmids. After determining the sequences of the DNA found between 3' IR end and the insertion sites, all of the analyzed rescued plasmids were categorized into seven classes that are summarized in Table 1-2 and described below.

Class I: Self- Ligated Vectors or Random Integrations (Fig. 1-8)

As the Southern and CAT assays (Fig.1-5 and 1-6) have suggested, the transfected plasmids are capable of maintaining themselves inside the cells for more than two months. Because 6 EcoRI sites were present in the donor vector (Fig. 1-8A), upon EcoRI digestion followed by ligation during the rescue assay, there were many possible ways to re-circularize the digestion products of the pMCH1 plasmid. However, only the digestion products containing the Bluescript component would survive ampicillin selection. Therefore, such plasmids had to possess the first EcoRI

Table 1-2. Summary of the Plasmid Rescue Assays

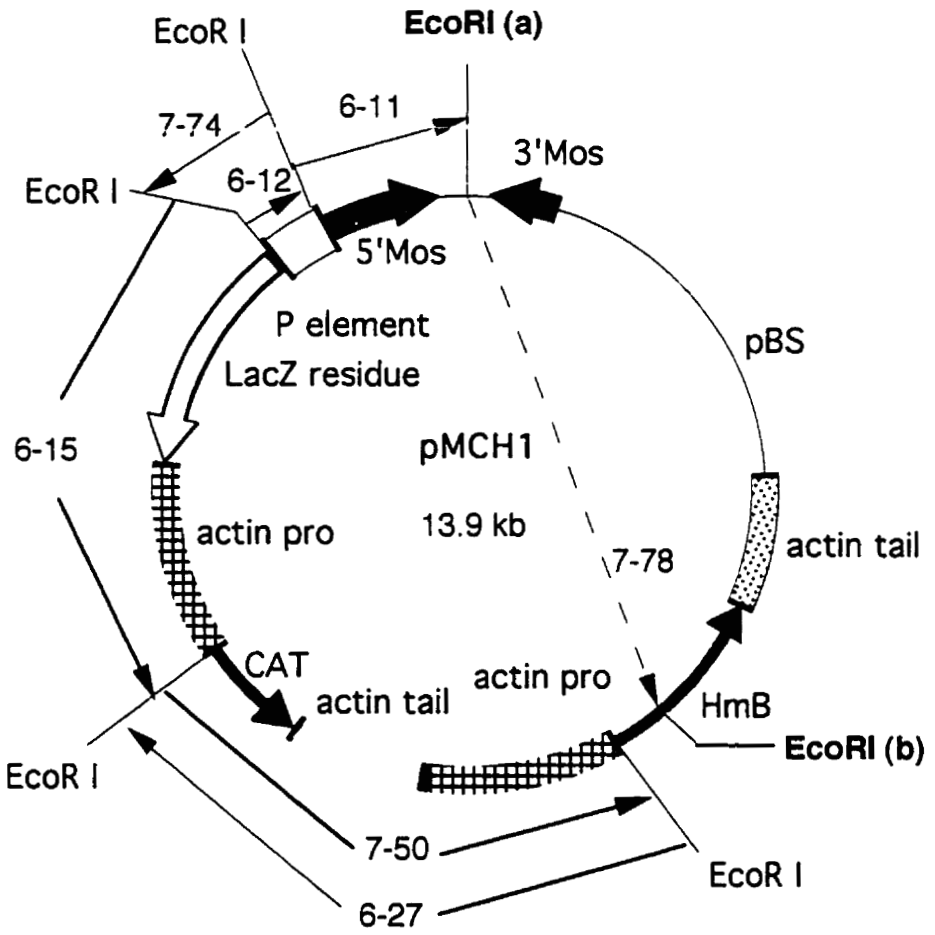
Bm5 cells were transfected by Mariner donor-helper vectors and cultured for 1 week (without antibiotic selection), or 9 weeks (with antibiotic selection). The transfected DNA was subjected to plasmid rescue assays. The 7 categories of rescued plasmids obtained from these experiments are listed.

transfection	rescued clones	self-ligation or random integration	random integration	mos1 mediated mobility					
				plasmid		excision	genome	complex	sub-total
				inter-	intra-				
pBmAMos1+ pMCH1(1w)	11	7	0	1	0	0	2	1	**4**
pBmAMos1+ pMCH1 (3w+6wH)	43	20	1	9	2	4	6	1	**22**
pMCH1 (3w+6wH)	23	23	0	0	0	0	0	0	**0**

Figure 1-8. Illustration of the partial pMCH1 rescued vectors generated from random integrations or self-ligations during the rescue process. **A:** The break-ligation sites on the plasmid pMCH1 for the self-ligated or random integration rescued plamids. The basic component was located between the EcoRI (a) and (b) sites shown in bold, which contained pBS. A total of 7 different rescued clones were characterized, in which different combinations of EcoRI fragments were found. **B:** The sequence details of the rescued vectors. The TA dinucleotide and the EcoRI sites are shown in bold. Plasmid 7-78 contained the basic structure, while the others represented various combinations with different EcoRI fragments of pMCH1. The arrows indicate the EcoRI fragments that were ligated into the basic rescued EcoRI(a)/(b) structure.

Class I
Self-Ligation or Random Integration

A



B

3'IR	*D. mauritiana* Spacer seq.	EcoRI	frag. of pMCH1

```
6-11: ......CCTGATATAGTTTCT....0.25k spacer....GAATTCCTGCAGCCCTAG...Mol
6-12: ......CCTGATATAGTTTCT.....................GAATTCGATACAAAGCTT...P-element
6-15: ......CCTGATATAGTTTCT.....................GAATTCACTGGCCGTCGT...LacZ
6-27: ......CCTGATATAGTTTCT.....................GAATTCGATATCAAGCTT...actin
7-50: ......CCTGATATAGTTTCT.....................GAATTCCGTATGGCAATG...CAT
7-74: ......CCTGATATAGTTTCT.....................GAATTCACTGGCCGGTCG...P-element
7-78: ......CCTGATATAGTTTCT.....................GAATTCAGCGAGAGCCTG...HmB
```

sites outside the Mos1 3' IR and the part of the HmB gene which flanked the two ends of pBS (Fig.1-4). Other EcoRI fragments from the same vector could also be ligated into these basic components. Starting from the 3' IR, the rescued plasmids could contain the 3' IR, 250 bp of *D. mauritiana* 3' flanking sequence (IR spacer), other EcoRI fragments from the donor vector (optional), an EcoRI site, the HmB cassette residue, and pBS (Fig. 1-8A, B).

On the other hand, if pMCH1 plasmid was linearized and integrated randomly into the *Bombyx* genome or other co-existing plasmids, as long as the integration sites were not located between the EcoRI (a) and (b) sites (Fig. 1-8A), the structures of the rescued plasmids would be the same as those of the self-ligated unintegrated plasmids. Because EcoRI digestion would destroy the sequential structure on the insertion site and make the rescued plasmids undistinguishable from those generated from self-ligation, the identity of the rescued plasmids from random integration or unintegrated donor vector could not be deduced.

Seven different rescued plasmids were identified among a total of 50 clones in all three groups (long-term, short-term and control groups). As predicted, the majority of the 23 rescued plasmids from the control group (pMCH1, 3w + 6wH) and 27 from experimental groups (pBmA.Mos1 + pMCH1, 1w, or 3w + 6wH) had the basic structures, *i.e.* they had the same structures as plasmid 7-78 and did not contain any other EcoRI fragments from pMCH1(Fig. 1-8; see also Table 1-2).

Class II: Random Integration at the Mos1 3' IR side (Fig.1-9)

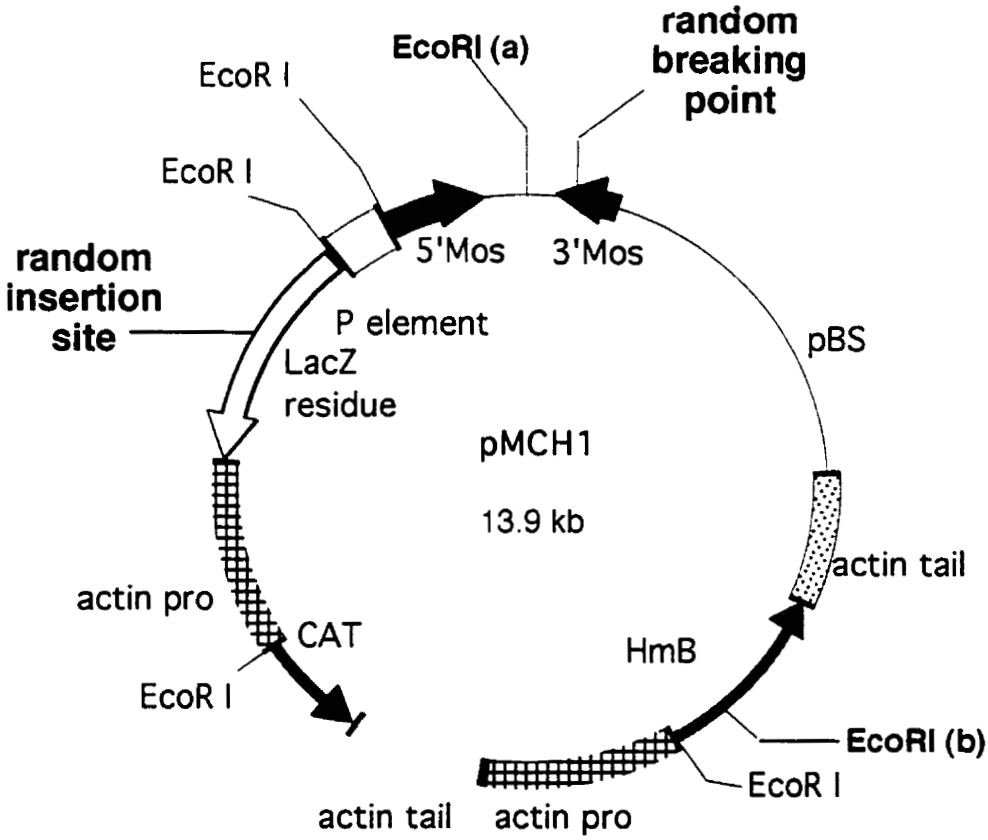
For vectors randomly integrated into either another plasmid molecule or into genomic DNA, the rescued plasmid should contain Bluescript, and maybe a variable length of the flanking components in the vector (Fig.1-4). Unlike the Mos1 mediated transposition, the random integration site would not necessarily be at a TA dinucleotide, and the junction between the vector and the insertion site would not be at the exact 3' IR end. Any site in the donor vector can be broken and integrated into the target DNA. In the cases of random integration that occurred outside the EcoRI (a)/EcoRI (b) region (Fig.1-9), only an "intact" Bluescript moiety would be guaranteed to be presented in the rescued plasmid because this moiety is required for surviving the ampicillin selection and DNA replication. In contrast, integrants for which integration occurred in the region bracketed by the two relevant EcoRI sites (a and b) would not be rescued. A rescued plasmid from a random integration event that involved sequences located immediately downstream of EcoRI (a) or immmediately upstream of EcoRI (b) (but leaving the pBS moiety intact; Fig 1-9) would most likely possess a variable length of foreign DNA at either side of the flanking sequence around the pBS moiety, *i.e.* an incomplete 3' IR or part of the IR spacer, or an incomplete *hmB* gene.

The structure of D7 fits these criteria (Fig 1-9). Apparently, this event involved random recombination between two plasmids, because the insertion site was a bacterial LacZ sequence, which could only come from the donor vector pMCH1.

Figure 1-9. **A:** The physical map of plasmid pMCH1 indicating the random recombination event that resulted in the rescue of the plasmid D7. The random breaking point was located 15 bp upstream of the Mos1 3'IR end. The integration site was at the middle of the LacZ gene of another copy of the same plasmid. Both pieces of evidence indicate that rescued plasmid D7 was not generated from a Mariner-specific integration event. **B:** Sequence details for the integration junction.

Class II
Inter-Plasmid Random Integration

A



B

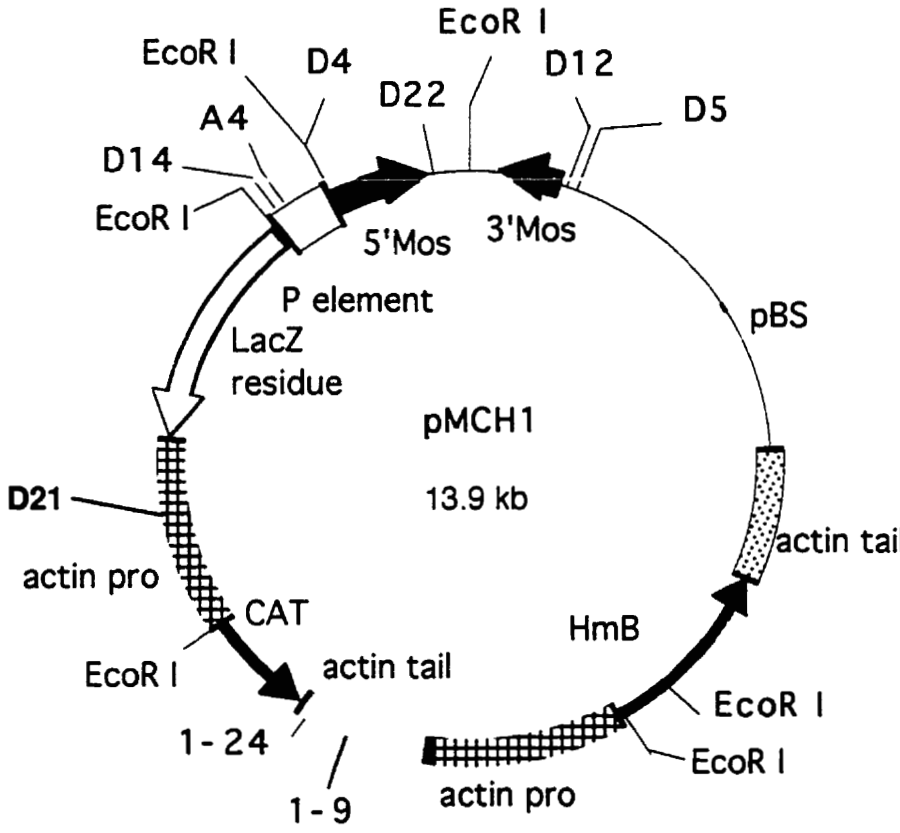
```
incomplete 3'IR (lacking    | non-TA insertion site
 15 bp out of 28 bp IR)     |

D7 : .........AAACGACATTTCA ACAGATAAAAAATG...... LacZ gene
```

Figure 1-10. The diagrammatic representation for Mos1 mediated inter-plasmid transposition. **A:** The insertion map of the donor vector pMCH1, showing the inter-plasmid insertion sites. The bold letters indicate the TA insertion sites of the rescued plasmids, which were adjacent to the Mos1 3' IR. **B:** The sequences of the rescued plasmid insertion site junctions which were identified as inter-plasmid mariner insertions. The TA dinucleotides at the insertion sites and the EcoRI are shown in bold.

Class III
Inter-Plasmid Mariner Insertions

A



B

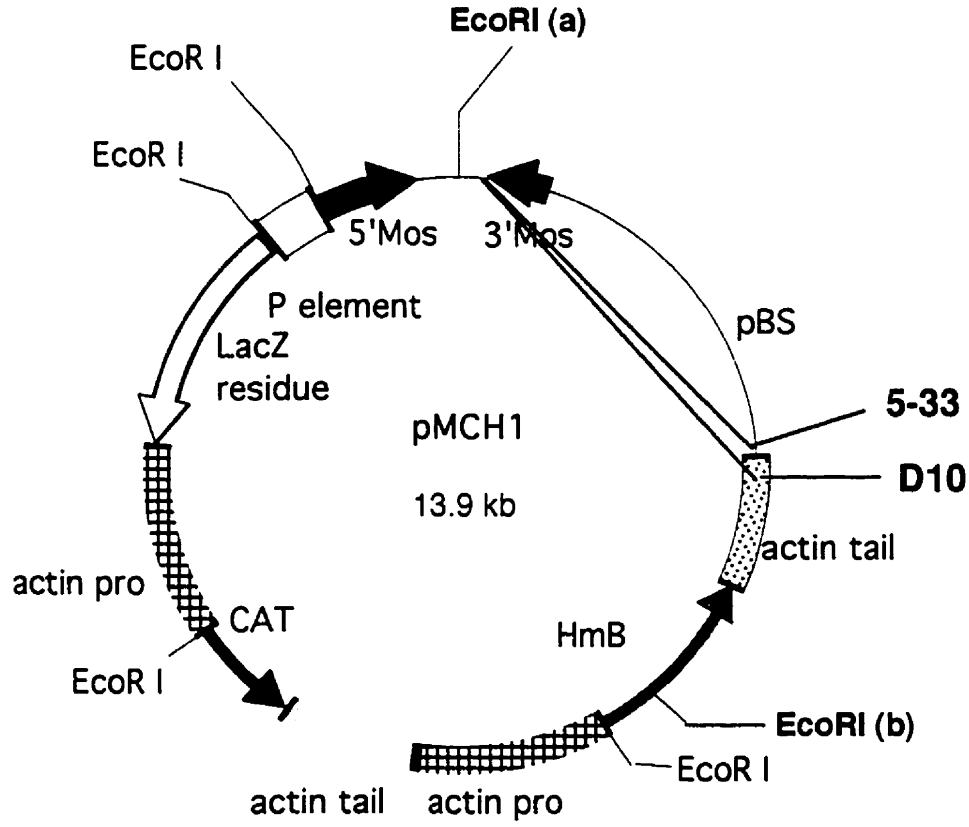
3'IR	Insertion sites in another pMCH1	hmB gene

```
2-28: ....TACTCCCACTGG...1.5K P element.......GAATTC.....GAATTCAGCG.....
1-9 : ....TAAATAAATTCAG...4.5k ActinA3(3730).............GAATTCAGCG.....
1-24: ....TAAATAAAGAAAT...1.0k ActinA3 (2211)............GAATTCAGCG.....
A4  : ....TATTTCATTTTTT...0.5k P-element(61).............GAATTCAGCG.....
D4  : ....TAAATGACACCGC...6.6k MosI(782) .....GAATTC.....GAATTCAGCG.....
D5  : ....TAATCGCCTTGCA...1.0k LacZ  (182)...............GAATTCAGCG.....
D12 : ....TAGTTCGCCAGTT...1.9k Amp(270) ......GAATTC.....GAATTCAGCG.....
D14 : ....TAACCCTTACGTG...1.2k P-ele (458) ..............GAATTCAGCG.....
D21 : ....TACAGTTGTTTTT...0.7k ActinA3(1202).............GAATTCAGCG.....
D22 : ....TAGGTCCCCGTTT...0.5k 5'Mos1(319) ..............GAATTCAGCG.....
```

Figure 1-11. The physical map for Mos1 mediated intra-plasmid transpositions. **A:** The insertion map of the donor vector pMCH1 showes the intra-plasmid excision and integration. Upon the attack of the Mos1 transposase, the 3'IR end was first excised. In the cases of D10 and 5-33, Mos1 transposase attacked a TA site within the same defective Mariner element as an insertion target. Thus, the 3' IR was attached to the TA site on its own plasmid and was repaired by the host enzyme system. Both recircularized plasmids were about 3 kb, and must have existed in the cells before the rescue process was applied. **B:** The sequencing results of the rescued plasmids which were identified as the intra-plasmid mariner insertions, showing the sequences at the junctions of 3'IR and the insertion sites. The TA dinucleotides at the insertion junctions are shown in bold.

Class IV
Intra-Plasmid Mariner Insertions

A



B

3'IR | Insertion Sites within the Basic Structures |

D10 : ...**TA**GTTGTGGTTTG...Actin-tail(SV40 polyA 298bp+Bluescript (No EcoRI)...
5-33: ...**TA**TTTGAATGTAT...62bp before Bluescript Ori (No EcoRI)...

Class III: Mos1 mediated inter-plasmid transposition (Fig. 1-10)

Mos1 transposition activity is not affected by the type of the target DNA, i.e. Mos1 could direct transposition of the donor vector equally well in chromosomal DNA, or in other copies of the transfected plasmids. For the latter, the transposition could occur in an inter or intra-plasmid fashion. In Mos1 mediated inter/intra plasmid transposition, the transposase would excise the reporter element exactly at the two ends of the inverted repeats, make a cut at an available TA sites in another copy of the vector, and insert the free fragment into the TA sites. The host enzyme system would repair the nicks left on the insertion sites. Therefore, the rescued plasmid would contain the 3' IR, a TA dinucleotides, the insertion sequences at the point of integration (which was at another vector), an EcoRI site, a partial HmB, and pBS. Ten plasmids were identified that fall into this category (Fig. 1-10).

Class IV: Mos1 mediated intra-plasmid transposition (Fig. 1-11)

In a specific situation, should the transposase recognize the TA sites in the actin-terminator region of the HmB cassette, and bring the 3' IR of the same vector at this point, the circularized vector should form a survivable plasmid of a size smaller than 5.5 kb (Fig. 1-7), lacking the relevant EcoRI site. Such a plasmid could have existed even before the rescue process. The structure should contain the 3' IR, a TA dinucleotide, and a partial (but functional) pBS, which would contain no EcoRI site. Rescued plasmids D10 and 5-33 fit this formula (Fig. 1-11).

Class V: Mos1 mediated genomic transposition (Fig. 1-7, 1-12)

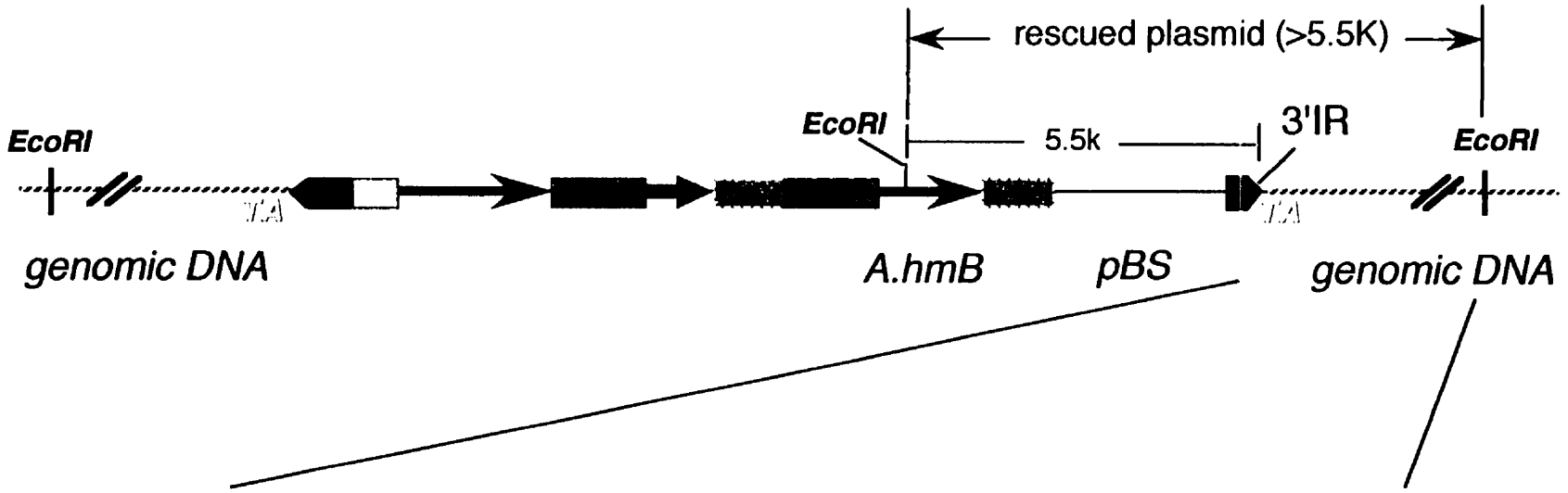
In Mos1 mediated genomic transposition, a situation similar to inter-transposition can occur except that the insertion site is in *Bombyx* genomic DNA. In this case, rescued plasmids should contain the 3' IR, a TA dinucleotide, the *Bombyx* genomic DNA, an EcoRI site [EcoRI (b)], the HmB residue, and pBS (Fig 1-8, 12). A total of 8 plasmids were identified that fit this formula. BLAST searches via GenBank revealed that among the unknown sequences at the insertion sites, some scored high in homology to *B. mori* genes, including those for the anti-bacterial peptide Enbocin [Furukawa *et al.*, 1997], Xanthine dehydrogenase, B1Rm retrotransposon [Ichimura *et al.*, 1997], ErA chorion and silk fibroin genes. Other short segments of the genomic DNA inserts shared high similarity with some *Drosophila* sequences including the Ecdysone inducible genes E75B and E75A, and other *Drosophila* genomic DNA sequences of unknown function (see Fig.1-12 and Appendix 2 for sequence details).

Class VI: Mos1 Excision intermediate products (Fig. 1-13)

Other researchers in this field have observed the presence of intermediate Mariner excision products [van Luenen et al., 1994, Goats et al., 1997]. The majority of the empty Mos1 excision sites contain 3 bp from one of the Mariner inverted terminal repeats [Bryan et at.1990]. Here, we also rescued such excision plasmids (Fig. 1-13). It appears that a linear excision intermediate was formed first, then this was circularized in a 5'-3' end-to-end manner. Coincidently, in each

Figure 1-12. Illustration of Mos1 mediated genomic integration.The upper part of the figure shows the sequences of plasmid pMCH1 between the 5' and 3' IRs that are integrated into the *Bombyx* genome. The lower part of the figure shows the insertion junction sequences of the rescued plasmids that were identified as derived from Mariner mediated integrations. The sequences start 100 bp upstream of the 3' IR (not shown), and are followed by the TA sites (bold) and the genomic insertion sequences, then linked back to a EcoRI site [EcoRI(b)] at the hmB gene. The segments of the genomic insertions showing high similarity with some known *Bombyx* and *Drosophila* genomic sequences are indicated. D.m.: *Drosophila melanogaster*; B.m.: *Bombyx mori*; L.t.: *Leishmania tareitolae.*

Class V
Mariner-Specific Integration of pMCH1



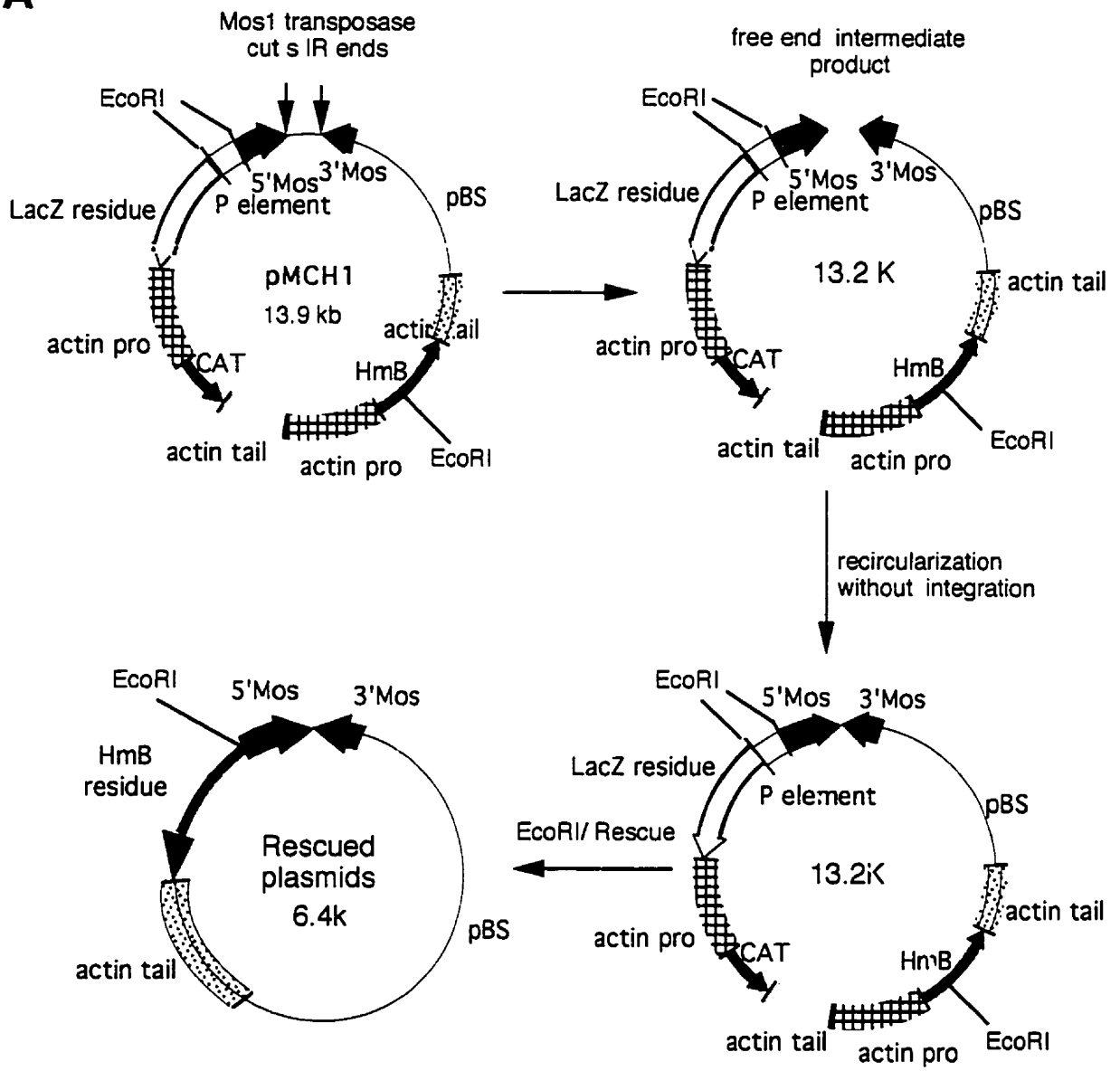
3'IR	Genomic Insertion Sequences	hmB Gene

```
2-18: ......TATACGGCCCGG......0.5k unknown seq.................TTTGCCTCTAAAACGAATTC.....
2-24: ......TATACTTCGGTA......1.1k (50bp B.m Xanthine..........GGCAGAAATAGAATGAATTC.....
                                    dehydrogenase)
1-2 : ......TAAAACTCCGCT......0.4k (70bp B.m L1Bm seq).........GTAGTTGCGBCCTCGAATTC.....
A9  : ......TAACTTAGTTAC......2.8k (130bp B.m L1Bm)............ACTTAGCTAAATGCGAATTC.....
D6  : ......TACTAGGGCTGC......0.3k (250bp B.m enbocin).........GAAGCGGAAGAAGCGAATTC.....
D8  : ......TAAATCTGTTCA......3.0k (60bp D.m E75B).............TTCAGGTTTGGAATGAATTC.....
D13 : ......TAGTGTTCCAGT......2.5k (70bp B.m ErA + 60bp L1Bm)..AGGGTGGATGATGTGAATTC.....
5-5 : ......TATTTAAACCAA......0.3k (130bp D.m.+ 80bp L.t seq)..CTAGAAGGNTACATGAATTC.....
```

Figure 1-13. Mariner self-excision. **A:** Illustration of the Mariner excision-ligation process and the structure of a rescued plasmid. The Mos1 transposase attacks the 5' and 3' IR ends in the plasmid pMCH1 and releases a free-end defective Mariner element, which is then recircularized by the host enzyme system. This process usually results in a three-nucleotide loss at either end of the Mariner element. **B:** The sequencing results of the rescued plasmids which were identified as originating from Mariner excision events, showing the sequences at the junctions of 3'IR and 5' IR. All the plasmids here have same size and contain only one EcoRI site.

Class VI : Mariner Self-Excisison

A



B

```
                Mos1 3' IR              Mos1  5'IR          EcoRI  HmB gene
     _----------------------|----------------------|-----|----------
A5 : ......CATACTTGTACACC___CCAGGTGTACAAGTAGGGAA....GAATTCAGCGAGAGCC...
D1 : ......CATACTTGTACACCTGA___GGTGTACAAGTAGGGAA....GAATTCAGCGAGAGCC...
D15: ......CATACTTGTACACCTGA___GGTGTACAAGTAGGGAA....GAATTCAGCGAGAGCC...
D20: ......CATACTTGTACACCTGA___GGTGTACAAGTAGGGAA....GAATTCAGCGAGAGCC...
                              ------
                           missing 3nt
```

recovered plasmid of this class, a 3 bp piece is missing from either end of the Mos1 inverted repeats. The data suggest that either Mos1 excision is not precise or, more likely, that Mos1 transposase cleavage yields 3-nucleotide 3'-overhangs instead of a 2-nucleotide overhangs as previously thought (see also Fig.1-16). During the mismatch repair process, the host enzyme system may fill one of the two overhangs while cut the other one. The two blunt ends could be circularized to resist further degradation. Consequently, the rescued plasmids from such events should contain a 3'IR, the 5'IR (short of 3 nucleotides at either end), an EcoRI site [EcoRI (b)], a HmB residue, and pBS (see also Fig.1-4 and 1-16). Rescued plasmids D1, D15, D20 and A5 are classified in this category. Each of these plasmids contains only one EcoRI site [EcoRI (5'Mos1) joins with EcoRI (b)], and is 6.4 kb in length.

Class VII: Rescued plasmids originated from complex Mos1 recombination events

The sequences after 3' IR and TA in two rescued plasmids, 2-17 and 2-41(Table 1-2), were unreadable. An attempt to sequence from the other end (from the EcoRI site at the *hmB* gene) also failed. To eliminate the possibility that the DNA samples were contaminated with other plasmid species, DNA from these rescued plasmids was re-transformed into *E. coli*. However, the restriction patterns and sequencing results from the plasmids that were isolated from single colonies were the same as the previous ones, which ruled out the "contamination" hypothesis. The results suggest that these two plasmids may contain more than one 3' IR regions.

In conclusion, for the short-term transfection of Mos1 element, in which plasmid rescue was carried out within a week post-transfection in the absence of any antibiotic selection, 11 rescued plasmids were analyzed. Seven of these plasmids were shown to be derived from unintegrated plasmid self-ligation or random integration events. Of the remaining four clones which were derived from the Mos1-mediated mobility, two were shown to have been derived from genomic integrations (50%).

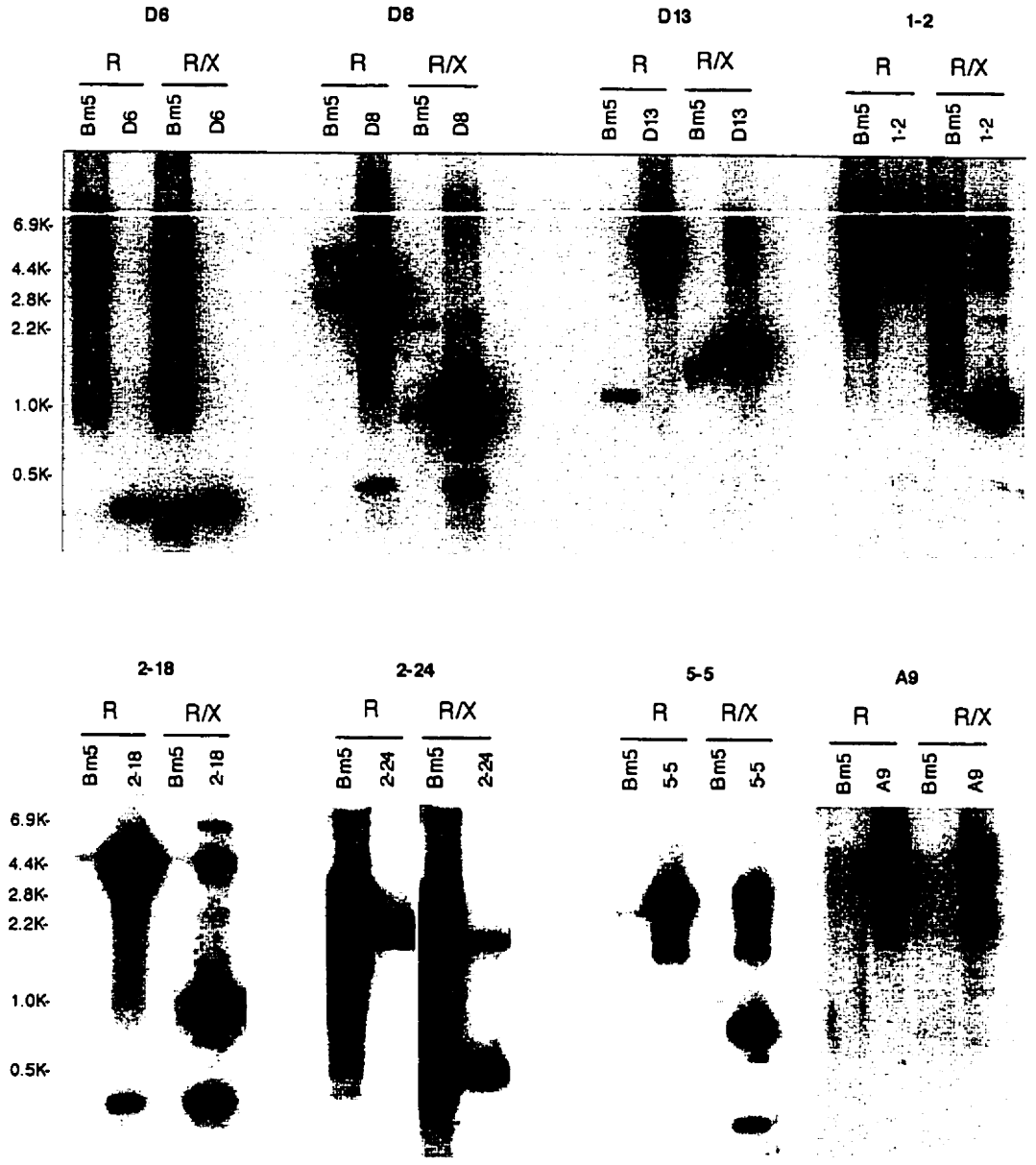
On the other hand, for the "long term" transfection of Mariner elements, plasmid rescue was carried out at the ninth week post-transfection, which included 3 weeks of subculture followed by 6 weeks of HmB antibiotic selection. Six Mos1-mediated genomic transpositions were detected, along with 11 Mos1-mediated inter-plasmid transpositions and four (presumably Mos1-mediated) self-excision-ligation events (table 1-2). The genomic integration rate among the Mos1-mediated transposition events was about 29%. Apparently, the selection pressure did not affect the frequencies of Mos1 transposition to any significant extent.

Consistent with other relevant studies [Gueiros-Filho et al., 1997, Lidlholm et al., 1993, Ivics et al., 1997], in absence of an active Mos1 transposase, no transposition events were observed (table 1-2).

4. Confirmation of Genomic Transposition

To confirm that the postulated genomic insertion sites were actually present in the *Bombyx* genome, the insertion fragments present in the rescued plasmids were used as probes for hybridizations with Bm5 genomic DNA (Fig. 1-14). The insertion

Figure 1-14. Southern blots of *Bombyx* genomic DNA hybridized with the insertion sequences from rescued plasmids. The rescued plasmids A9, D6, D8, D13, 1-2, 2-18, 2-24, 5-5 (20 ng each) and Bm5 genomic DNA (3μg) were digested with EcoRI (R) or EcoRI/XhoI (R/X) respectively. Each rescued plasmid DNA was grouped with Bm5 genomic DNA and hybridized with the probe derived from its own insertion sequences. The results showed that all the probes (indicated at the top of each panel) were able to hybridize with the genomic DNA, indicating these insertion sequences came from the *Bombyx* genome. The A9 probe contained only 1/3 of the radioactivity as that of the other probes used in the hybridizations.

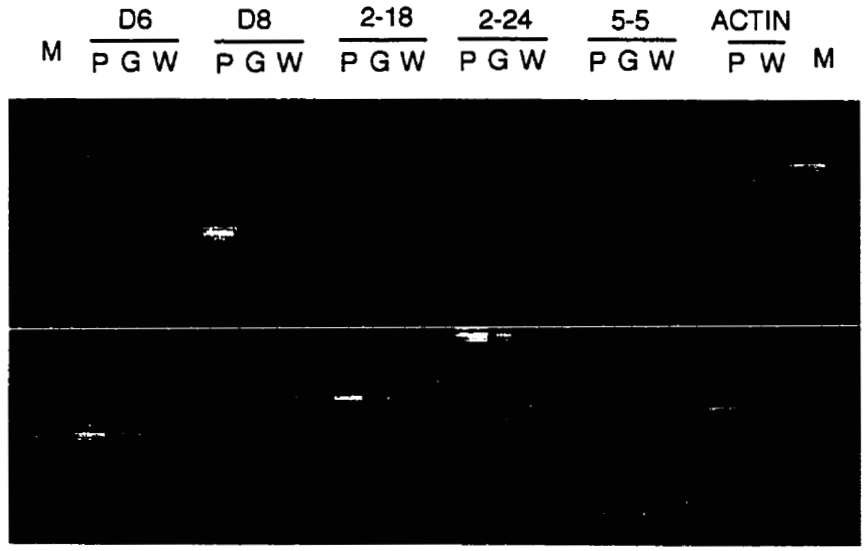

D6
D8
D13
1-2
R
R/X
Bm5
D6
D8
D13
1-2
6.9K-
4.4K-
2.8K-
2.2K-
1.0K-
0.5K-
2-18
2-24
5-5
A9
2-18
2-24
5-5
A9

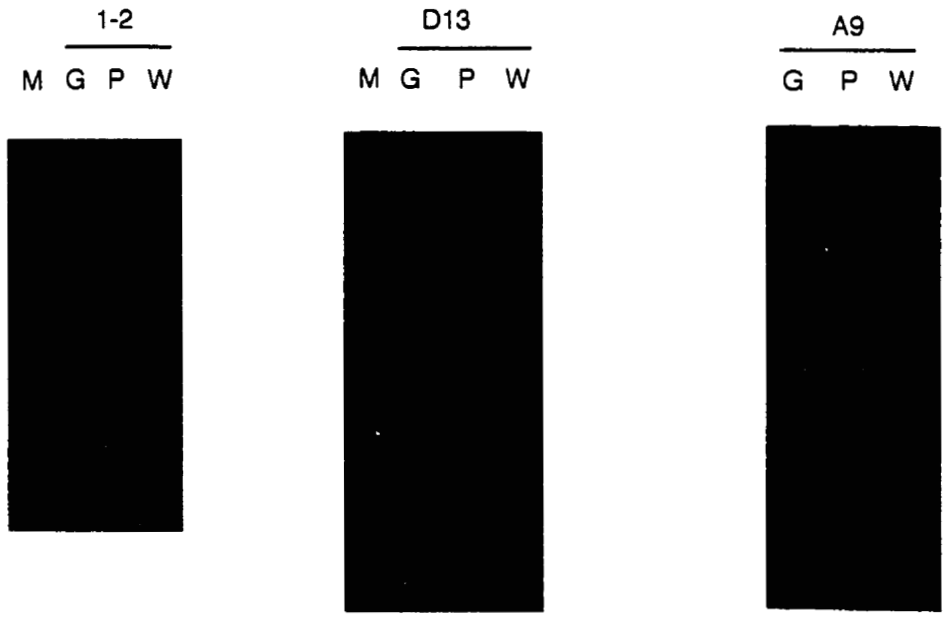
fragments of all rescued plasmids in class V (Mos1 mediated genomic transposition class) were used as probe for hybridization with Bm5 genomic DNA. Five plasmids, A9, D8, D13, 2-18 and 5-5 showed single band patterns in the genomic Southern hybridizations, while the fragments of plasmids D6, 1-2 and 2-24 yielded smearing signals. The intensity of the signals obtained from the hybridizations represents a measure of the copy numbers of these sequences in the Bm5 genome. The intense smear obtained with the 2-24 probe hinted that the rescued insertion sequence may be a mid or highly repetitive element, while the lighter signals obtained with the D6 and 1-2 probes suggest that these are repetitive elements of low copy numbers. The single band signals obtained with the rest of the hybridization probes indicate that these transpositions have occurred at unique sites of the *B. mori* genome.

In addition, we employed primer pairs (Appendix 2) derived from the two ends of each rescued fragment to amplify genomic DNA. As shown in Fig. 1-15, the lengths of the resultant PCR products were exactly the same as those obtained from the plasmids from which the primers were derived, except for that of plasmid D8 which contains a ~3 kb insert (Appendix 2). Sequencing of the PCR fragments that were amplified from genomic DNA showed the exact DNA sequences as those present in the rescued plasmids (data not shown).

In summary, 26 Mos1-mediated transposition events were detected among 54 analysed rescued plasmids. These included 8 genomic transposition events, 4 excision events, 2 complex transpositions and 12 inter-plasmid transpositions. There were also

Figure 1-15. PCR analysis. The insertion fragments shown to represent *Bombyx* genomic DNA based on Southern hybridizations (Fig. 1-14) were examined by PCR amplification to see if the same fragments can be obtained from *Bombyx* genomic DNA. Each primer pair was derived from the two ends of each insert fragment. M: λ DNA double-digested with HindIII and HincII; P: the rescued plasmids templates; G: Bm5 genomic DNA templates; W: water control, containing same components except no template DNA. The genomic DNA templates produce the same fragments as the plasmid insertions with the A9, D6, D13, 1-2, 2-18, and 2-24 primers, a ladder with the 5-5 primers, and no products with the D8 primers.


M
D6
P G W
D8
P G W
2-18
P G W
2-24
P G W
5-5
P G W
ACTIN
P W
M


1-2
M G P W
D13
M G P W
A9
G P W

1 random integration event and 27 partial donor vector self-ligation or random integration events that were rescued. No Mos1-related transposition was detected in the absence of the Mos1 transposase (Table 1-2).

IV. Discussion

We have demonstrated that the Mos1 Mariner transposable element of *D. mauritiana* is capable of conducting gene integration in a *Bombyx* cell line. DNA sequencing, Southern hybridizations and PCR amplifications were employed in order to prove that Mos1 Mariner can traverse different taxonomic orders and function in *B. mori*.

Efficiency is always a major concern in the development of a practical gene transformation system. In our system, there are two factors that could contribute to the transposition efficiency: the host endogenous mariner elements, and the transformation vector itself.

The *Bombyx* genome contains approximately 2400 copies of Mariner-like elements, Bmmar1, which are thought to be inactive [Robertson et al., 1997, Tomita et al., 1997]. Significant differences have been found between the IRs and ORF of Mos1 and those of Bmmar1. Because the regulatory aspects of mariner transposition remain unclear [Hartl et al., 1997, Lohe et al., 1996], we do not know whether the endogenous Bmmar1 affects Mos1 transposase activity or not. It is possible that endogenous inactive Mariner transposases in the silkmoth cells may compete with

active Mos1 transposase for IR substrates; or that endogenous Mariner sequences, in the form of DNA substrates, can compete with the donor vectors for the active transposases. In both cases, the efficiency of Mos1 mediated transposition would be reduced.

The donor vector itself may also play role in the regulation of the transposition. Our current vector was basically made from a *D. mauritiana* Mos1 element that was previously manipulated to allow insertion of a foreign DNA fragment [Lidholm et al, 1993]. The region of that insertion is relatively conserved among the mariner elements and may be important for transposition activity [Lohe et al., 1997].

Random integration has been widely reported in cases where foreign DNA was introduced into cultured cells [Moritz & Williams, 1994]. However, our rescue process did not identify many random transposition events. This was due to the structure of donor vector and the choice of the restriction enzyme for the rescue experiments, which selected against the majority of the random integration events. Although random integration may have occurred after the transfection of the vectors, the rescue assay was only able to detect such events, if they occured at the 300 bp spacer region which was directly adjacent to the 3' IR in the plasmid pMCH1 and ended at the first EcoRI site[EcoRI (a)]. Considering that random integration may occur at any single nucleotide in the plasmid pMCH1 (14 kb in size), only 2% of the possible random integration events could be determined directly by our assay. The remaining 98% would be classified under Class I and would be undistinguishable

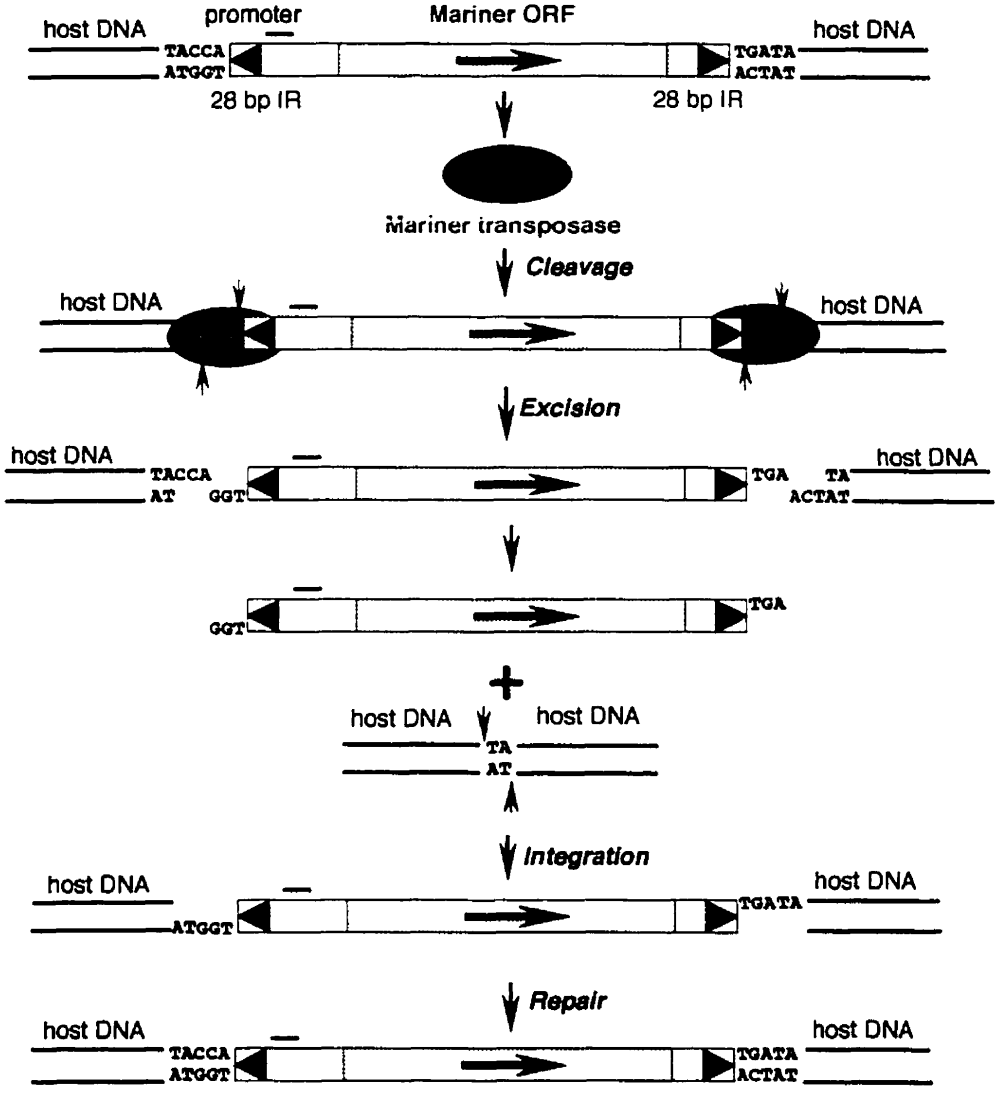
from the self-ligation events involving unintegrated plasmid sequences recovered by the rescue system (see the text about Class I rescued plasmids).

When the transposition frequencies for the short and long term transpositions are compared, it appeared that the application of the drug does preferentially select cells carrying the donor vectors, in either integrated or unintegrated fashion. A higher ratio of donor self-ligation/random integration event relative to Mos1 transpositions was observed with the short-term transfection. These may result from the presence of increased numbers of unintegrated donor plasmids in the cells due to lack of opportunity of disappearance of these plasmids through extended culture periods.

Recent footprinting studies at empty excision sites in the *Drosophila* genome have demonstrated that Mariner is capable of excising its IR ends and generating a 3 nucleotide overhang at their boundaries [Bryan et al., 1990, Hartl et al., 1997(a)]. Our rescued fragments actually displayed such excision intermediate products which were ligated in an "end-to-end" fashion by the host mismatch-repair mechanism, ie, either end of 3' overhangs was removed while the other end remained intact. This may indicate that if integration does not occur immediately after the excision, the free ends of the elements can be re-circularized by the host repair system to resist degradation. Therefore, based on the published data and our own results, we propose a modified model for Mos1 mediated transposition (Fig.1-16)

To investigate whether the defective Mariner elements present in the silkmoth genome can mediate the transposition in absence of the active MosI transposase, the control cell lines were also transfected with donor vector pMCH1 alone. Our results

Figure 1-16. Model for Mos1 mariner transposition. This figure is based on Lampe *et al.*, 1996 and our own results. The red ovals representing transposase molecules are removed from subsequent steps for clarity but are assumed to be involved at each step of the transposition process. The shaded red box shows the Mos1 ORF, the green arrowheads and green letters represents the non-perfect inverted terminal repeats. The TA dinucleotides shown in blue are the genomic target sites. The black lines represent the genomic DNA, and the short bars show the MosI promoter sequences.


host DNA
promoter
Mariner ORF
host DNA
TACCA
ATGGT
TGATA
ACTAT
28 bp IR
28 bp IR
Mariner transposase
Cleavage
host DNA
host DNA
Excision
host DNA
TACCA
AT
GGT
TGA
TA
ACTAT
host DNA
GGT
TGA
host DNA
host DNA
TA
AT
Integration
host DNA
ATGGT
TGATA
host DNA
Repair
host DNA
TACCA
ATGGT
TGATA
ACTAT
host DNA

have shown no evidence that an endogenous *Bombyx* Mariner-like transposase can transpose the donor pMCH1 vector (Table 1-2). Active Mos1 transposase was absolutely required for transposition involving the *Drosophila* IR sequences.

In summary, our results show that the *Drosophila* Mariner transposase can mediate transposition in a *B. mori* cell line. For all Mos1-mediated transposition events that we observed, the insertion sites were at TA sites without exception. No apparent consensus sequences and orientation preferences have been found around the insertion sites. Also our work has shown that the plasmid DNA can remain unintegrated for more than 6 weeks (Fig. 1-5 and 1-6).

Although this system needs to be tested in silkmoth embryos, it has set an example for the development of a universal gene transformation system. The success of this system does not only add another example illustrating how far the Mariner elements can navigate in foreign genomes, it also provides a realistic application for the study and manipulation of such a genetically and economically important insect.

CHAPTER TWO

THE I-Ppo ENDONUCLEASE AS A TOOL FOR GENETIC TRANSFORMATION OF *Bombyx mori*

I. Introduction

One disadvantage of transposon-mediated recombination is that it allows only random integration. Homologous recombination methods for targeting specific genes, analogous to those developed in mouse embryonic stem (ES) cells [Bronson and Smithies, 1994], would be extremely useful. Homologous recombination (HR) between chromosomal and exogenous DNAs is at the basis of methods for introducing genetic changes into the genome [Bollag et al., 1989]. The enzyme-catalyzed site-specific recombination systems (e.g., yeast-based FLP-FRT and PI phage Cre-loxP, see ABBREVIATION for detail), which are promoted by double-strand breaks in DNA [Deng and Capecchi, 1992], have been limited by low transformation efficiencies. Therefore, a promising alternative approach is the use of the mobile intron homing system.

1. Group I introns

In several lower eukaryotes, the genes encoding ribosomal RNA are located on extrachromosomal DNA molecules. Their rDNA shares several unusual features, one of which is the presence of intervening sequences in the coding region for rRNA. These introns are designated either as group I or as group II, based on several conserved consensus sequences that allow the RNA within each group to assume a characteristic structure required for splicing [Cech, 1990, also see fig.2-1]. Although both groups often contain ORFs capable of encoding a protein, group I introns are mainly

distinguished from group II introns by specific secondary structural features. Moreover, group I introns splice themselves out by a transesterification mechanism [Cech, 1990], whereas group II introns splice by a reaction pathway more closely resembling nuclear pre-mRNA splicing [Cech and Sharp, 1987].

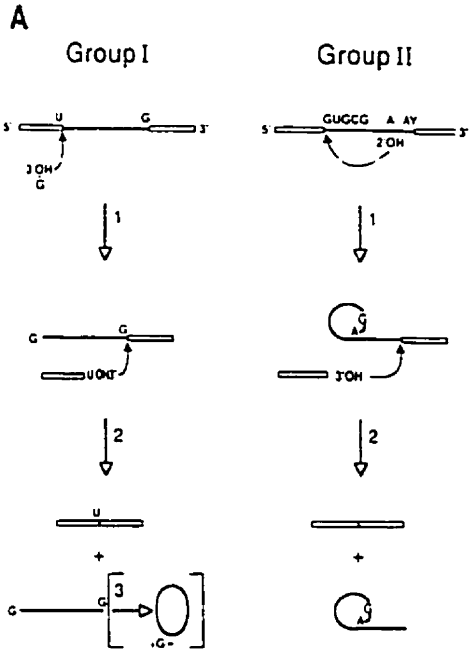
The distribution of group I introns is uneven: they are common in mitochondrial and chloroplast genomes of eukaryotes, but are rare in nuclear DNA. In the few known cases of nuclear group I introns, the insertion site of the intron is always in the genes for ribosomal RNA, and these rDNAs exist as extrachromosomal elements instead of the more usual tandemly repeated copies on one or more chromosomes [Lambowitz, 1989].

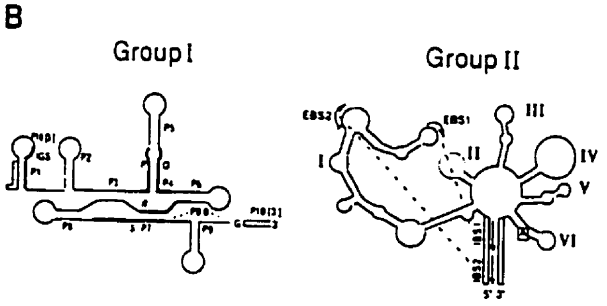
Some group I introns are mobile. They have the ability to transpose in a highly specific manner to an unoccupied homologous target site. In all known examples, mobility is mediated by a specific endonuclease encoded by the intron itself. The endonuclease cuts the target DNA near the site of intron insertion, and then a gene conversion event is initiated at this break and leads to "replication" of the intron into the unoccupied site [Lambowitz & Belfort, 1993]. To date about a dozen intron-encoded endonucleases have been characterized to some extent. Most of them are associated with group I introns in mitochondria of yeast, chloroplasts of *Chlamydomonas ssp*, or T-even bacteriophages [Shub et.al, 1988; Belfort, 1990].

2. Mechanism of Group I "homing"

Unlike most other transpositions, which involve non-homologous donor and

Fig.2 -1. Intron structure and splicing pathway (adapted from Lambowitz, 1993). **A:** Splicing pathways. Exons are shown as boxes, introns as solid lines or circles. Group I pathway: nucleophilic attack at the 5' splice site by the 3'-OH of guanosine; the diagram shows the conserved U and G residues at the 3' ends of the 5' exon and intron, respectively. Step 2, the second transesterification reaction leading to exon ligation. Step 3, cyclization coupled to cleavage releasing a short 5' oligonucleotide with the 5'-terminal G. Group II pathway: Step 1, lariat formation; the diagram shows conserved boundary nucleotides and the A residue forming the 2',5' linkage. Step 2, exon ligation. **B:** Group I and II Intron DNA secondary structures. The conserved sequences in the group I intron structure are in boldface.


A
Group I
Group II
U
G
3'OH
G
GUCCG A AV
2'OH
1
1
G
G
UOH
3'OH
2
2
U
+
+
3
G
.G-

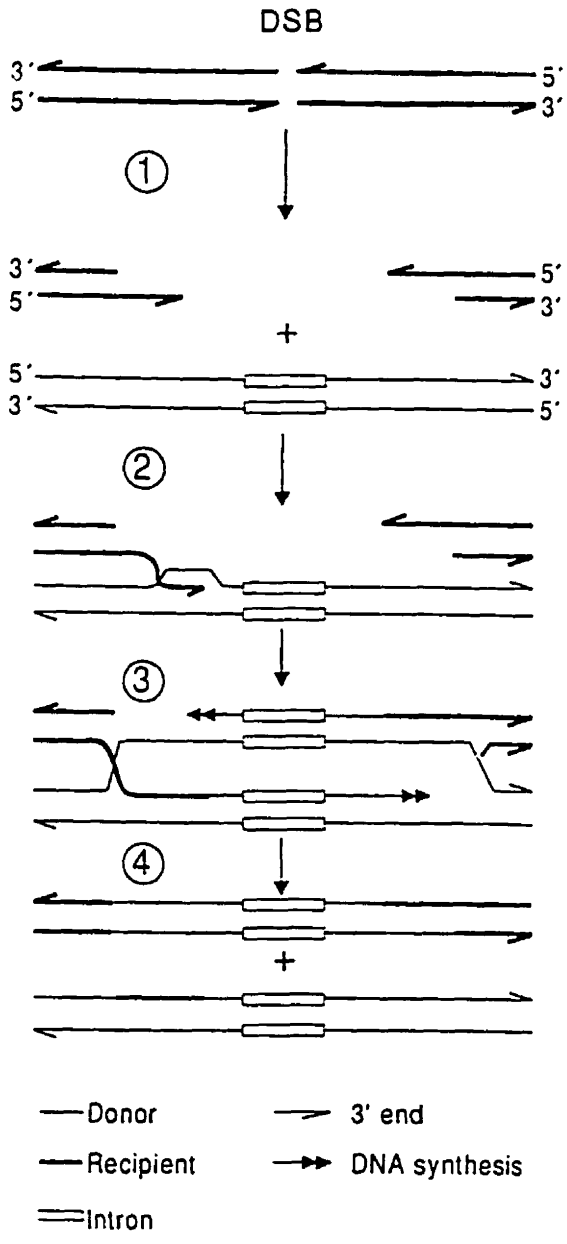

B
Group I
Group II
EBS2
EBS1
III
I
II
IV
V
VI
5' 3'

recipient sites, group I intron mobility is site-specific in that it is restricted largely to exchanges between alleles of genes that contain or lack the intron. This processes, previously referred to as unidirectional intron conversion or site-specific intron transposition, is now termed "homing" [Belfort & Perlman, 1995]. The exchanges are characteristically non-reciprocal and efficient, that is , they are unidirectional gene conversions whereby intron-lacking (I^-) alleles are depleted from a population of I^- and intron-containing (I^+) DNA molecules increase as a result of repeated rounds of intron insertion [Lambowitz & Belfort, 1993]. A further distinction between group I intron mobility and most "conventional" transpositions is that the recipient DNA sequences that flank the inserted intron are not duplicated. In all cases, intron movement requires the activity of a site-specific, double-strand DNA endonuclease encoded by the intron whose cleavage site specificity is considerably more complex than that of typical restriction endonucleases [Perlman & Butow, 1989].

The intron homing cycle is initiated and targeted by a site-specific DNA double strand break, which is made by the intron-encoded endonuclease in a homing site allele that lacks the intron (see fig.2-2). A copy of the intron containing the endonuclease ORF is transferred to the cleaved homing site by DNA double strand break repair or gene conversion [Lambowitz and Belfort 1993; Belfort and Perlman 1995; Belfort and Roberts 1997]. The inserted intron is replicated within the host DNA and expresses the intron-encoded endonuclease, thus retaining the potential for homing to other endonuclease-sensitive, intron-minus alleles.

The prototype for all group I introns is the well-known *Tetrahymena*

Fig.2-2 The double-strand-break (DSB) repair pathway for intron mobility (adapted from Lambowitz & Belfort). The four steps that ensure after the DSB is made are described in the text.


DSB
3′
5′
5′
3′
①
3′
5′
5′
3′
+
5′
3′
3′
5′
②
③
④
+
Donor
Recipient
Intron
3′ end
DNA synthesis

thermophila rDNA intron Tt LSU 1, whose autocatalyitc properties have been studied in detail [Cech, 1988; Cech and Bass, 1986]. Mechanistically, splicing is accomplished by a series of transesterification reactions (Fig. 2-1A), which begin with the attack at the 5' splice site by the 3' OH group of a G residue that is specifically bound to a site in the folded RNA, that is also called the "ribozyme" [Alberts et al., 1989]. The linear intron is released by a second transesterification reaction, involving the G residue that is found as the last residue of all group I introns. The relased intron may undergo further reactions, including cyclization and hydrolytic opening of the circular RNA [Grabowski et al., 1981; Zaug et al., 1984]. All of the reactions carried out by the ribozyme are reversible under appropriate conditions [Woodson et al., 1989].

3. I-Ppo Intron-encoded Endonuclease

In only a few instances have group I or II introns been found in a nuclear gene, and in all these cases the introns reside in rDNA [Cech, 1988]. The only described example of the group I mobile nuclear intron is the Pp LSU 3 (the third intron in the *P. polycephalum* gene coding for Large Sub-Unit RNA) gene, which is found in the extrachromosomal rDNA of the *Carolina* strain of the acellular slime mold *Physarum polycephalum* (Fig. 2-3). Of the three intervening sequences of the 26S rRNA coding region (Fig. 2-3 and Muscarella et al., 1990), two are group I introns. The Pp LSU 3 mobile intron is the third intron that undergoes self-splicing. It is integrated in a region of the rDNA that is highly conserved. The 3' half of the Pp LSU 3 sequence is closely related to that of Tt LSU1, while the 5' half is novel, encoding the I-Ppo endonuclease

[Muscarella et al., 1990]. Upon formation of a diploid plasmodium containing an I-Ppo$^+$ rDNA and an I-Ppo$^-$ rDNA, the I-Ppo intron efficiently "colonizes" the rDNA lacking this intron [Muscarella et al., 1989]. This transposition is mediated by a site-specific endonuclease, namely I-PpoI (or briefly I-Ppo), encoded in the intron. The DNA sequence and predicted amino acid seqence of the endonuclease are illustrated on Fig.2-4.

It is not known how the I-PpoI endonuclease gene is expressed. Since the gene is embedded in rDNA, presumably the mRNA for I-Ppo is derived from the pre-ribosomal RNA, making this a unique example of a protein-coding gene naturally transcribed by RNA polymerase I. After the self-splicing of Pp LSU 1 *in vitro*, this catalytic RNA can process itself to liberate an RNA fragment that is the mRNA for I-Ppo *in vivo*.

Two protein products have been predicted from the sequence of the Pp LSU3 intron and the adjacent 26S rRNA gene sequences [Muscarella et al., 1990]: a 163 amino acid "short" form of I-PpoI encoded entirely by the Pp LSU3 intron (Fig.2-3B and 2-4), and a 185 amino acid "long" form of I-PpoI containing 22 additional N-terminal residues encoded by adjacent rDNA exon and intron sequences (Fig. 2-3B). This "additional" sequence precedes the ATG of the short form intron, which is located at nucleotide 14-16. Both predicted proteins contain a single copy of the His-Cys box motif (Fig. 2-4). The His-Cys motif has the potential to bind one or more zinc atoms and may associated with DNA-binding function or intron mobility [Flick et al., 1997]. Several additional mobile intron endonuclease families have been identified that share

Figure 2-3. Schematic representation of DNA constructs (modified from Ruoff *et al.*, 1992).

Panel A is a diagram of a portion of extrachromosomal rDNA of the *Carolina* strain of *Physarum polycephalum*. P, promoter; SSU, coding region for the small subunit RNA; LSU, coding region for the large subunit RNA; 1, 2 and 3, introns.

Panel B is a diagram of Pp LSU3 and surrounding DNA. Black rectangles, exons; line and hatched box, intron sequences; the lines above the intron show the open reading frames that can be translated into the long and short forms of the I-PpoI endonuclease. RPZ, ribozyme portion of the intron; 3'UTR, untranslated region between the ORF and the ribozyme, ALT 5'SS, alternative 5' splice site; 5'SS, splice site, IPS, internal processing site; 3'SS, 3' splice site.

Panel C shows the pI3-d9 plasmid structure. pI3-d9 was made by subcloning a MseI-ClaI fragment of the Pp LSU3 into Bluescript (Stratagene) XbaI-ClaI sites (MseI and XbaI sites were filled-in and blunt-end ligated). The fragment is 932 bp. There is a reading frame coding for the short from I-PpoI in the 5' side. At the 3' end, there are 24 bp of exon and 3 bp of neighbouring intron sequences.

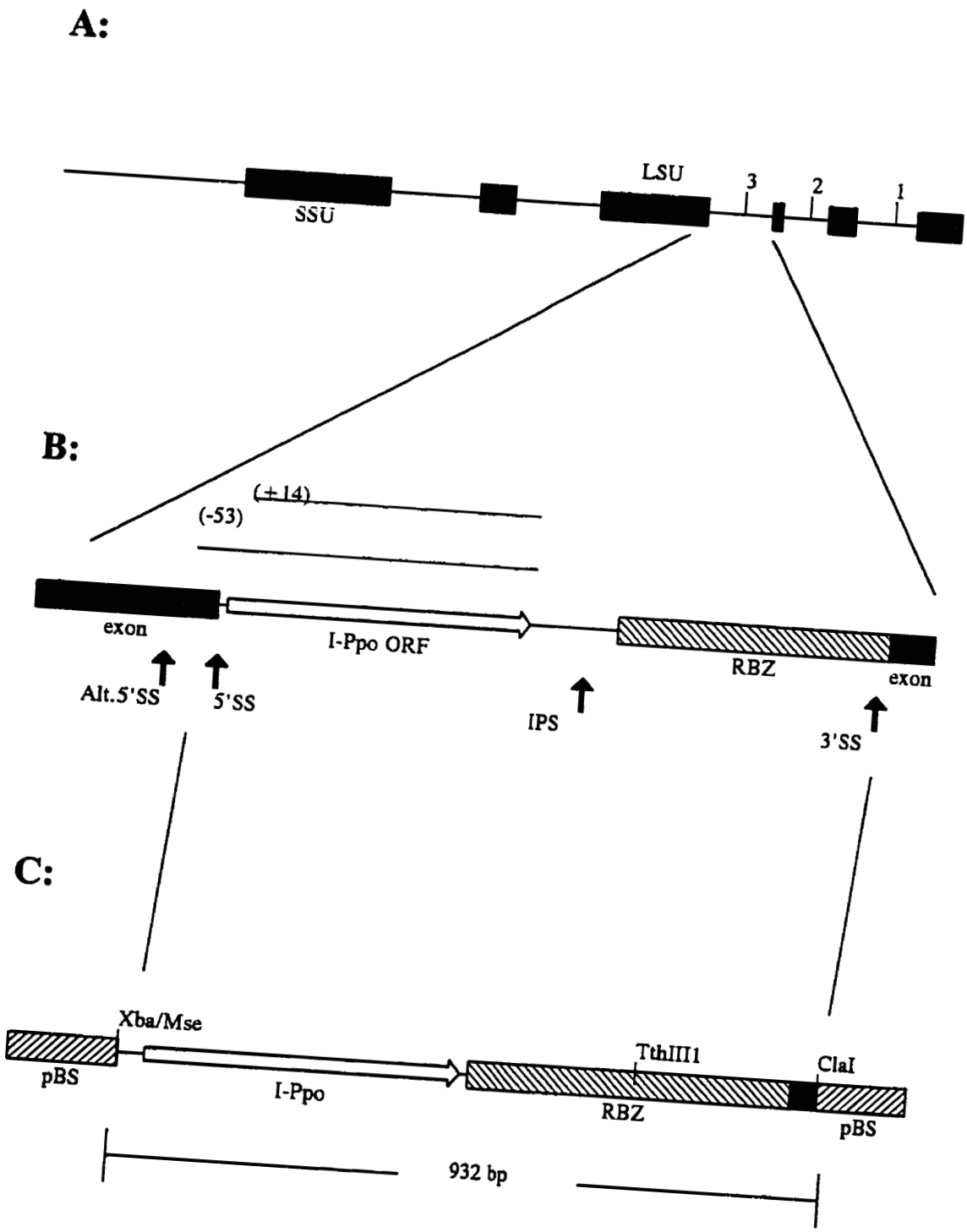
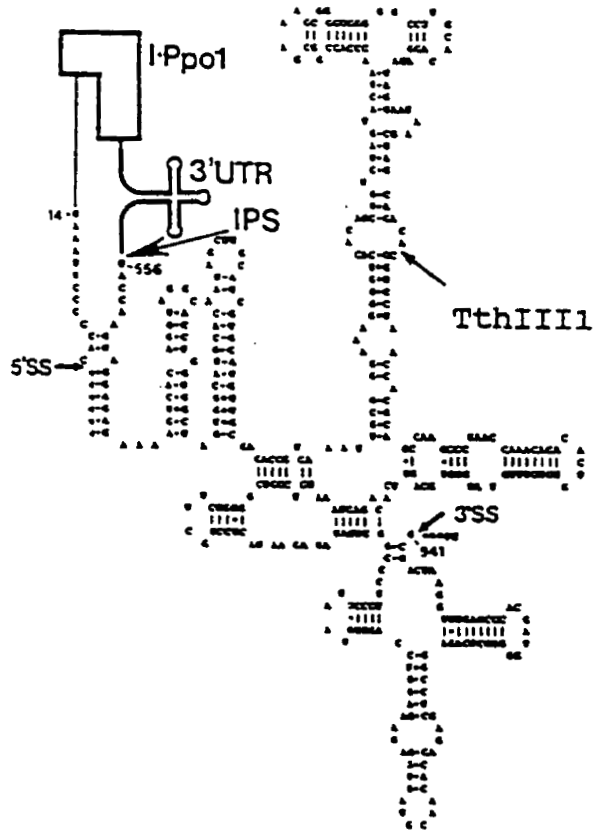

A:
LSU
3
2
1
SSU
B:
(+14)
(-53)
exon
I-Ppo ORF
RBZ
exon
Alt.5'SS
5'SS
IPS
3'SS
C:
Xba/Mse
TthIII1
ClaI
pBS
I-Ppo
RBZ
pBS
932 bp

Figure.2-4. The nucleotide and deduced amino acid sequences of the PpLSU intron (from Gene Bank, accession number: L03183, Muscarella et al., 1990). For the DNA sequence, capital letters correspond to intron sequences; the lower case letters correspond to the upstream and downstream exon sequences. For the amino acid sequences, the short form of I-Ppo endonuclease is in uppercase, while the 5' part of the long form of I-Ppo endonuclease is in lowercase. The His-Cys motif is in bold and underlined. The MseI-ClaI fragment was cloned into pBS plasmid. In the expression vector pBmA.I-Ppo', the TthIII-ClaI fragment was removed in order to destroy the ribozyme core.

```
cagtgctctg gatgttaaaa tggcgaaatc caaccaagct cgggtaaacg
                    m   a   k   s   n   q   a   r   v   n   g
                  MseI ↓
gcgggagtaa ctatgactct ctCACCCCCT TAAATATGGC GCTCACCAAT
  g   s   n   y   d   s   l   t   p   l   n   M   A   L   T   N
GCTCAAATCT TGGCTGTGAT TGACAGCTGG GAAGAAACAG TCGGTCAGTT
A   Q   I   L   A   V   I   D   S   W   E   E   T   V   G   Q   F
TCCAGTGATA ACGCACCATG TACCATTAGG TGGCGGTCTG CAAGGAACGC
  P   V   I   T   H   H   V   P   L   G   G   G   L   Q   G   T
TCCATTGTTA CGAGATCCCC CTAGCAGCTC CTTATGGGGT TGGCTTTGCT
L   H   C   Y   E   I   P   L   A   A   P   Y   G   V   G   F   A
AAGAATGGGC CTACCCGCTG GCAATACAAA CGGACAATCA ATCAAGTCGT
 K   N   G   P   T   R   W   Q   Y   K   R   T   I   N   Q   V   V
CCACAGATGG GGATCCCACA CAGTCCCTTT TCTATTAGAA CCGGATAACA
 H   R   W   G   S   H   T   V   P   F   L   L   E   P   D   N   I
TCAACGGCAA AACCTGCACA GCATCGCACC TATGTCATAA TACTCGATGC
  N   G   K   T   C   T   A   S   H   L   C   H   N   T   R   C
CACAATCCCT TGCACTTGTG CTGGGAGTCA CTAGACGACA ACAAAGGCAG
H   N   P   L   H   L   C   W   E   S   L   D   D   N   K   G   R
AAACTGGTGC CCGGGTCCAA CGGGGGATGT GTCCATGCGG TGGTTTGTTT
 N   W   C   P   G   P   T   G   D   V   S   M   R   W   F   V   -
AAGGCAGGGT CCGTTGTACG CCCGGGGGCG ACTGTGGCAG GTCCTCAACA
AAGGGGCAGT CACTTTGTGG TATAAGGTCT GAGGTTACTA GATCGGGTGG
GTCAGCCCAC GGGCAAAACG CTGACCGTAC CAAGTAGGAG AGAAAAGTCA
CAGGTAGGCA CCTGGTGGCT AGTGGCATGC ACAATAGCTT GCATCGTGCA
TGTTGCGAGA CCGTCAAATT GCGGGAAAGG GGTCACACAG CTGTCTAGTA
CTAACCCACC GGCGAAAGCG GGGTGGGGGC CTTGCACAGG ATATAGTAAT
AAGCTAGCAG ACACGGTCCT AACCACGCAG CCAAGCCCTA ACCAAACACA
                ↑ TthIII 1
CACTGTGTGT TTGTATGGGT GCAGTTCACA GACTAAATGT CGGTCGGGGG
CTTCGGCCTC ATAAGATATA GTCGGCCCTT CCCTTAATGG GGACCTGGAG
AAGAGGTGAT GCAACACCAG AGCTCCGGAG ACTCGGGGGT AGCACCCGAG
TTTGGAATCA GCGtaaggta gccaaa.....atcgat
                                      ↑ ClaI
```

Fig. 2-5. A structural model of the PpLSU intron (adapted from Ruoff et al., 1992). Secondary structure of the group I intron ribozyme (RBZ) region. Arrows indicate the 5' and 3' splice sites (5' and 3' SS) as well as the internal processing site (IPS) [Zaug et al., 1983]. The P1-extension containing the I-Ppo coding region and the 3' untranslated region (3'UTR) is also shown .


I-Ppo1
3'UTR
IPS
TthIII1
5'SS
3'SS

the conserved His-Cys box motifs [Belfort & Perlman, 1995; Belfort & Roberts, 1997].

Both short form and long form endonucleases are enzymatically active when synthesized *in vitro* and have the same recognition site. It is not known whether one of the two possible protein species is preferentially synthesized or is active in *Physarum* [Muscarella et al., 1990].

The short form of I-Ppo exists as a small homodimer of apparent molecular mass 34-39 kDa [Ellison & Vogt, 1993]. The endonuclease binds and induces a shallow bend in the partially symmertric, 15 bp homing site in the presence of EDTA, and can be activated by divalent metal ions, including Mg^{2+}, Mn^{2+}, Ca^{2+}, Co^{2+} and Zn^{2+} [Mucsarella et al., 1990; Ellison & Vogt, 1993; Wittmayer & Raines, 1996]. Homing site binding and cleavage are salt dependent, with K_ds ranging from 3.3 to 112 nM in the presence of 33-275 mM NaCl. Cleavage is particularly efficient, with a k_{cat}/K_m of $10^8 M^{-1}s^{-1}$, and occurs at the centre of the 15 bp homing site to generate four base, 3' overhangs of the sequence 5'-TTAA-3' [Wittmayer & Raines, 1996]. The interaction of I-Ppo with its homing site has been well characterized, although the degree of sequence degeneracy tolerated within the site has not yet been defined. The enzyme appears to share some functional characteristics with a number of Type II restriction endonucleases, such as high sequence specificity and catalytic efficiency (Fig. 2-5) [Mueller et al., 1993].

4. Regulation of Intron Mobility

Transposition is highly regulated in all biological systems. Mobile introns are

under selective pressure to control their movement and safeguard the viability of their host. One way to control intron movement is the loss or inactivation of the ORFs that encode the endonucleases, as was demonstrated for the nuclear large rRNA introns in *P. polycephalum* and *T. thermophila* [Muscarella & Vogt, 1989]. For example, in the case of *T. thermophila*, the initial insertion was probably promoted by a site-specific endonuclease that was subsequently lost in later generations.

The mobility can also be controlled by limiting the expression of the intron ORFs at the transcriptional or translational level, and by limiting recombination, i.e. by regulation of the site-specificity of the endonucleases [Wernette et al, 1992; Monteilhet et al, 1990; Thierry et al, 1991] or the homology-dependence of intron homing [Quirk et al, 1989]. Another barrier to mobility may be imposed by the absence of necessary recombination functions in the host. In addition, repair of the double-stranded break may be too inefficient to maintain viability when sustained in a single-copy genome, so that endonuclease-containing group I introns could be confined to multi-copy genomes or genes [Lambowitz, 1989].

5. Horizontal Transmission of I-Ppo Elements

Group I introns are present at different locations in the genome of different organisms and are more strongly conserved than the surrounding exon sequences. This strongly suggests that group I introns conduct horizontal transmission between organisms. There may exist many pathways for intron horizontal transfer [Heinemann, 1991], including transfer of naked or packaged DNA with or without concomitant

DNA replication. In a symbiotic association, the introns are provided within a genetic environment that is consistent with preserving gene expression, while the ORFs promote intron propagation. However, only specific intron-ORF associations result in a functional mobile elements compatible with the viability of the host. The nature of the association may be influenced by such factors as intron structure, splicing pathway, organellar *versus* nuclear compartmentalization, and residence in single *versus* multicopy genomes [Lambowitz, 1993].

The PpLSU3 intron has been reported to be able to mediate homing to the intronless rDNA genes of *P. polycephalum*. When the strain with the PpLSU3 intron (I^+ strain) is mated with the other strain (I^- strain), the intron rapidly colonizes almost all of the 150 previously intronless rDNA molecules [Muscarella, 1993]. Muscarella and Vogt (1993) have developed an experimental system for introduction of Pp LSU3 into *S. cerevisiae*, an organism that does not normally contain any nuclear group I introns. Transformation with plasmids carrying Pp LSU3 was found to be lethal in yeast cells due to the cleavage of some of the 150 tandem rDNA repeats by I-Ppo. However, transformants that were resistant to the lethal effects of I-Ppo appeared at a surprising high frequency. These transformants were shown to be of two classes. The first class acquired point mutations at the I-Ppo cleavage site. The second class acquired the PpLSU3-element in all its rDNA repeats. Cells of the second class were shown to produce significant levels of I-Ppo. Furthermore, mating experiments indicate that Pp LSU 3 was mobile in yeast cells [Muscarella, et al, 1993].

6. I-Ppo Endonuclease and *Bombyx mori*

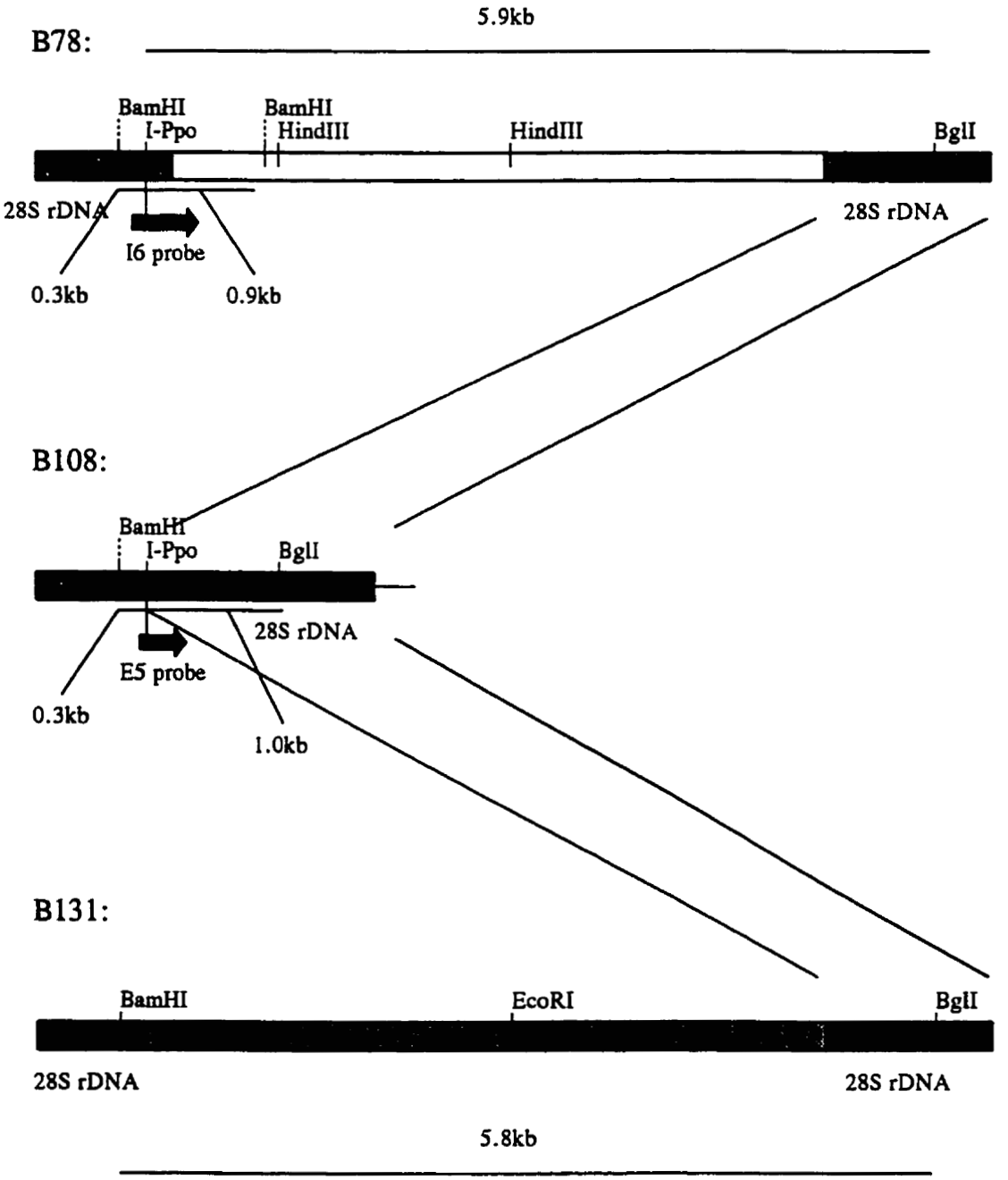
Recent reports have demonstrated that the intron-encoded endonuclease can induce recombination at a high efficiency in yeast and mammalian nuclei [Michel et al., 1998; Fairhead et al., 1993]. The double-stranded breaks (DSBs) made by the endonuclease can be repaired with a donor molecule having homologous regions flanking the break. In addition, the endonuclease recognition sequences are very rare in genomic DNA. Since the endonuclease protein is not a recombinase, its potential for chromosome engineering is larger than that of systems with target site requirement on both host and donor molecules.

The rDNA of *B. mori* consists of 240 repeating units [Gate, 1974] that form the nucleolar organizers (NO) in the genome. Three length classes of the repeating units have been identified. About 83% of the rDNA repeats are 10.6 kb in length and contain no intronic interruptions (represented by clone B108, Fig.2-6). About 8% of the repeats contain a 5.0 kb insertion (represented by clone B131, Fig.2-6), while 2% of the population possess a 4.5kb insertion (represented by clone B78, Fig.2-6) [Eickbush et al., 1985; Fujiwara et al, 1984]. The two insertion sites are located close to the middle of the 28S rDNA gene units. Through the sequence homology searches, we found I-Ppo intron-encoded endonuclease recognition sequences in the 28S rDNA gene of *Bombyx* [Eickbush et al., 1985; Fujiwara et al, 1984; Fig.2-6]. The I-Ppo cleavage site is located between the two insertion sites of the I^+ 28S rDNA genes. Actually, one of the insertion (B131) destroys the I-Ppo recognition sequence.

This finding raised the possibility that I-Ppo endonuclease could be used to

Figure 2-6.. **Panel A**, The physical maps of the three genomic clones coding for three different classes of *Bombyx mori* 28S rDNA genes (modified from Eickbush *et al.*, 1985). B108 contains no intron, while B78 and B131 contain introns inserted at different sites close to the 28S rDNA gene. Solid boxes: 28S rDNA, open box: B78 intron, shaded box: B131 intron. The EcoRI, BamHI, HindIII and BglI sites are shown. E5 and I6 probes are generated from PCR products which share the common 5' ends but differ at their 3' parts. Both probes start from the exon sequence (at nt 1 in B108, see Panel B). However, E5 ends in the exon sequence (nt 276 in B108, see Panel B), while I6 ends in the intron sequence (nt 456 in B78, see Panel B). **Panel B**, The sequences of the three different clones from *Bombyx mori* 28S rDNA genes showing the I-Ppo recognition sites . B108 represents mature rRNA sequences. The insertions in B78 and B131 occur at nt 64 and nt 138 respectively in the intron-less version gene. Only the clones B108 (without intron) and B78 (with intron) contained I-Ppo sites. The 28S rDNA sequences are uppercased, while the insertion sequences are lowercased. The I-Ppo recognition sequences are located from nt 50 to nt 72, and underlined. The cleavage sites are pointed by the arrow.

A: Three forms of B.m. 28S rRNA genes



B: I-Ppo recognition sequences in the Bombyx mori 28S rDNA.

```
B78: GTCAACGTGA AGAAATTCAA GCAAGCGCGG GTAAACGGCG GGAGTAACTA
B108:GTCAACGTGA AGAAATTCAA GCAAGCGCGG GTAAACGGCG GGAGTAACTA
B131:GTCAACGTGA AGAAATTCAA GCAAGCGCGG GTAAACGGCG GGAGTAACTA
               ↓ I-Ppo cut site
B78: TGACTCTCTT AAGG
B108:TGACTCTCTT AAGG
B131:TGACTCTCTT AAGcgggagt aactatgact ctcttaggg....~4.4kb..

B78:                TAGCCA AATGCCTCGT CATCTAATTA GTGACGCGCA
B108:               TAGCCA AATGCCTCGT CATCTAATTA GTGACGCGCA
B131:..cggcgatg aaaaTAGCCA AATGCCTCGT CATCTAATTA GTGACGCGCA

B78: TGAATGGATT AACGAGATTC CCACTGTCCC TATCTACTtg acttcgccgt
B108:TGAATGGATT AACGAGATTC CCACTGTCCC TATCTACT
B131:TGAATGGATT AACGAGATTC CCACTGTCCC TATCTACT

B78: cggcct....~4.5kb......tgtccctatc tactATCTAG CGAAACCACA
B108:                                    ATCTAG CGAAACCACA
B131:                                    ATCTAG CGAAACCACA

B78: GCCAAGGGAA CGGGCTTGGG AGAATCAGCG GGGAAAGAAG ACCCTGTTGA
B108:GCCAAGGGAA CGGGCTTGGG AGAATCAGCG GGGAAAGAAG ACCCTGTTGA
B131:GCCAAGGGAA CGGGCTTGGG AGAATCAGCG GGGAAAGAAG ACCCTGTTGA

B78: GCTTGACTCT AGTCTGGCAT TGTAAGGAGA CATGAGAGGT GTAGCATAAG
B108:GCTTGACTCT AGTCTGGCAT TGTAAGGAGA CATGAGAGGT GTAGCATAAG
B131:GCTTGACTCT AGTCTGGCAT TGTAAGGAGA CATGAGAGGT GTAGCATAAG

B78: TGGGAGATCG TTTCGCGGC
B108:TGGGAGATCG TTTCGCGGC
B131:TGGGAGATCG TTTCGCGGC
```

direct homing at specific sites (rDNA genes) in the *Bombyx* genome. Therefore, we constructed I-Ppo expression vectors that provide the site-specific endonuclease, and a donor vector that provides the alignment sequence and genetic markers for homologous recombination events. It has been predicted that upon co-transfection of the two plasmids into a *Bombyx* cell line, Bm5 cells, the I-Ppo endonuclease would be expressed and transported into the host nuclei, where it would create double strand breaks specifically at the 28S rDNA genes. The cloned I-Ppo site flanking sequences in the donor vector would be expected to align to the host 28 S rDNA genes, and the sandwiched reporter or other foreign genes in the donor vector would be inserted into the host genome at the "homing" sites by the homing (gene conversion) mechanism. Based on this assumption, we examined the *in vivo* function of the expression and donor vectors, and analysed the *in vivo* endonuclease activity of the transfected expression vectors. However, we were unable to demonstrate directly that homing did occur in the transfected cells.

II. Materials and Methods

1. Construction of I-Ppo Expression Vectors with different promoters

To provide the *Bombyx* cell line with I-Ppo endonuclease, the I-Ppo expression vector was constructed from plasmid pI3-d9 (Fig. 2-3C; see Muscarella et al., 1990), which contained a Pp LSU 3 intron short of 9 nucleotides at its 5' end and its intact ORF encoded the short form of the I-Ppo endonuclease. The 1 kb NotI-SalI fragment

containing the I-Ppo ORF was cloned into the BamHI site of pBmA plasmid, resulting in pBmA.I-Ppo plasmid with an expression cassette under the control of Bombyx actin promoter (Fig. 2-7A).

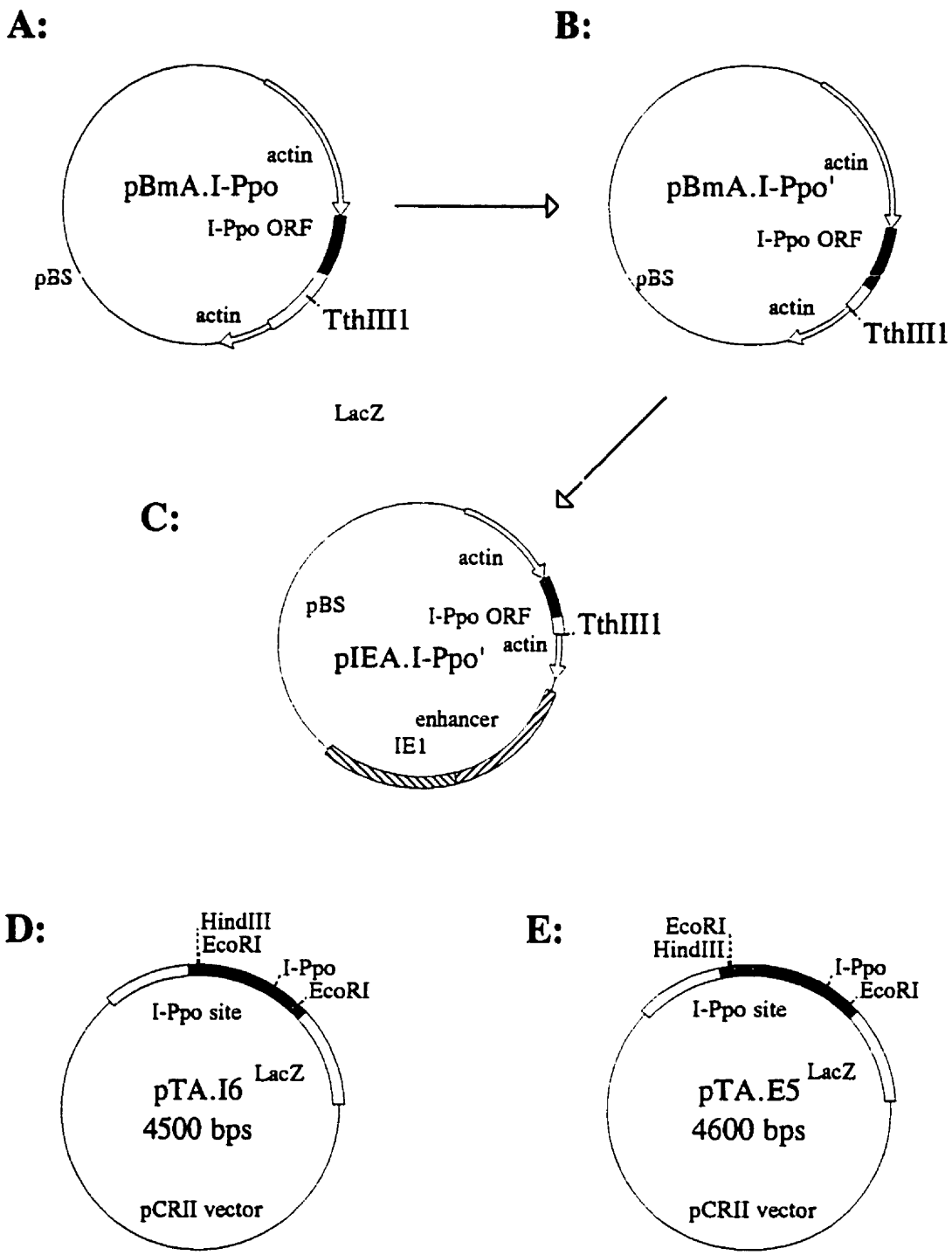
In order to control the lethal effect of I-Ppo cleavage, the pBmA.I-Ppo vector was modified to produce different levels of cutting activity. First, to avoid possible mRNA cleavage by the ribozyme portion formed at the 3' side of the Pp LSU 3 intron, a 200bp TthIII-ClaI (741-941 bp of the intron) fragment was removed from pBmA.I-Ppo, therefore the ribozyme core (638-854bp) was destroyed. The resulting plasmid was pBmA.I-Ppo' (Fig. 2-7B).

To increase the I-Ppo expression level, a 2.7 kb SacI fragment from pBmA.I-Ppo' containing the actin promoter, the I-Ppo ORF (without the ribozyme structure) and the actin terminator sequences was subcloned into the SacI site of the over-expression plasmid pIEA that contains two additional genetic elements, the enhancer element HR3 and the IE1 trans-activator [Farrell et al., *in press*] (Fig.2-7C). The IE1 gene product and the HR3 enhancer can each increase 100-fold the level of expression of the actin promoter. The combined presence of the IE1 gene and HR3 enhancer in the same plasmid results in an increase of around 5,000-fold in the level of expression of a basic actin promoter-based vector [Lu et al., 1998]. The new plasmid pIEA.I-Ppo' is shown in Fig. 2-7C.

2. Construction of the I-Ppo Donor Vectors

A 3 kb MboI/PstI (from an intronless 28S rDNA gene of *Bombyx*; Eickbush et

Figure 2-7. Physical maps of the plasmids constructed for the I-Ppo project. Panels A, B and C, illustrate the series of the I-Ppo expression vectors (the "helper" vectors) pBmA.I-Ppo, pBmA.I-Ppo' and pIEA.I-Ppo. The solid boxes represent the I-Ppo ORF; the open box, the non-coding region of the I-Ppo element; the open arrow, the actin promoter and actin terminator sequences; Panels D and E illustrate the cloned I-Ppo target site vector pTA.I6 and pTA.E5; the shaded box represents the I-Ppo recognition sequence from *B. mori* rDNA, I-Ppo cleavage sites in both vectors are 69 nt away from the EcoRI site at the right side (see text for explanations regarding the differences between pTA.E5 and pTA.I6 target vectors); solid boxes, polylinkers; open boxes, Lac Z gene. pCRII vectors are PCR TA cloning vectors which were purchased from INVITROGEN.


A:
B:
actin
pBmA.I-Ppo
I-Ppo ORF
pBS
actin
TthIII1
pBmA.I-Ppo'
LacZ
C:
pIEA.I-Ppo'
enhancer
IE1
D:
HindIII
EcoRI
I-Ppo
I-Ppo site
pTA.I6
4500 bps
pCRII vector
E:
pTA.E5
4600 bps

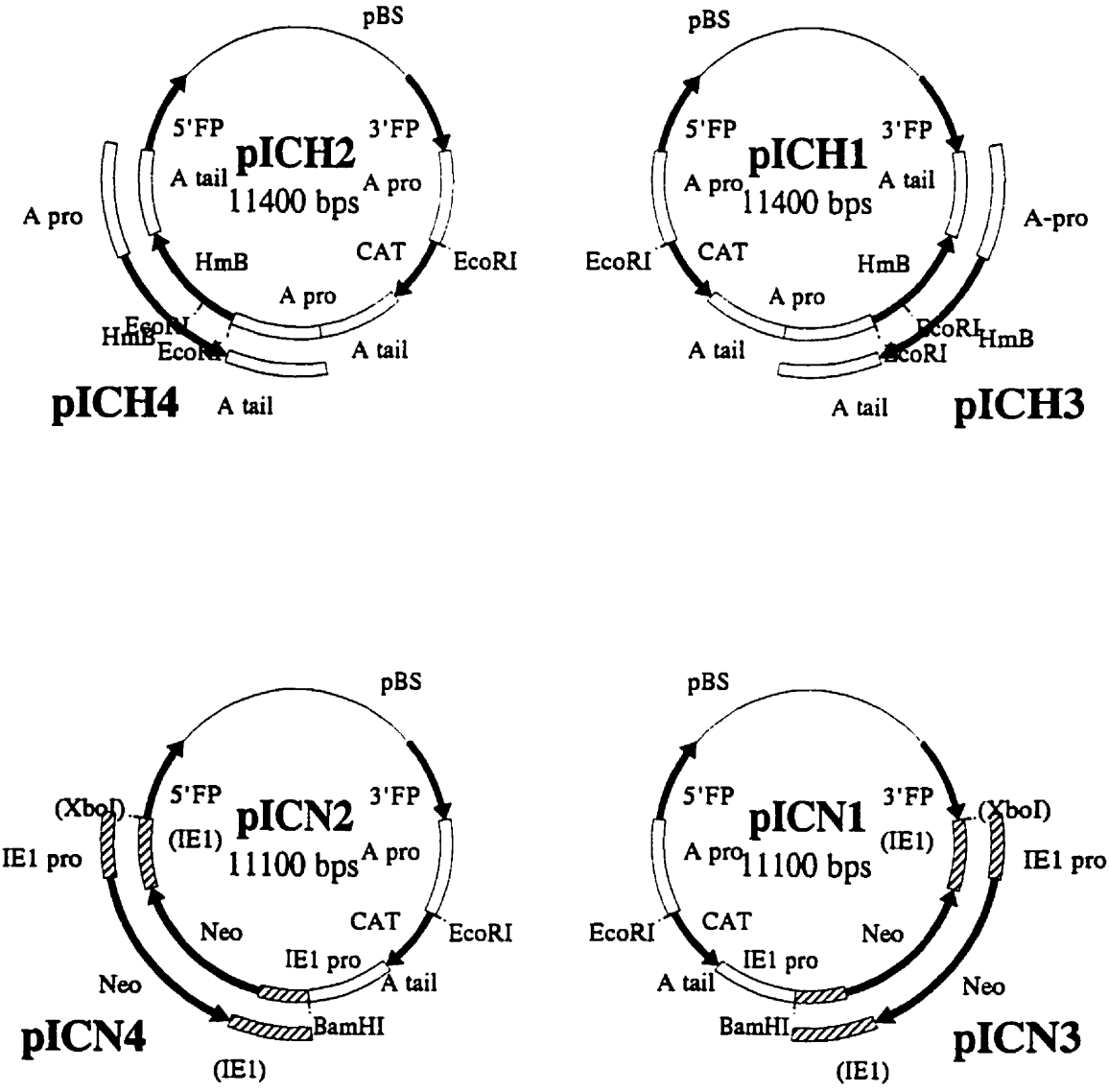
al., 1985) and a 2.8 kb MboI/EcoRI (from an intron-containing 28S rDNA of *Bombyx*), both of them containing I-Ppo recognition sequences and their flanking sequences in Bombyx rDNAs [Fujiwara et al., 1984; Eickbush et al., 1985] were subcloned from their genomic clones (pBmR161, pBmR145, Dr. H. Maekawa's gift; B108 and B78 are derived from them respectively, also see Fig.2-6) into the SacI site of Bluescript (Stratagene). DNA from the resultant plasmids pBS161 and pBS145 was cleaved by commercial I-Ppo endonuclease, and this was followed by insertion of a 6.7 kb SacI fragment containing the reporter CAT gene with actin promoter control and a Neomycin resistance gene under IEI promoter control, resulting in plasmids pICN1, pICN2, pICN3, and pICN4 (the numbers here indicate the different orientations of the reporter and selection genes inside the vectors; Fig. 2-8); or a 6.5 kb Sac I fragment containing the actin-CAT cassette and a Hygromycin B resistance gene driven by the actin promoter (the same fragment used for construction of pMCH1,2, see Fig 1-3A), resulting in plasmids pICH1 pICH2, pICH3 and pICH4 (Fig. 2-8). These donor vectors would allow alignment of the sequences at the I-Ppo target site upon double strand breakage introduced by the expressed I-Ppo endonuclease. Additional sets of donor vectors were also constructed in a similar way. These contained different components (the *B.mori* genomic fragments for I-Ppo recognition with or without intron) at different orientations in the vectors (see Fig 2-8 and table 2-1 for details).

3. Construction of I-Ppo Target Test Vectors

To examine the I-Ppo endonuclease activity *in vivo* (expression in tissue culture

Figure 2-8: I-Ppo donor vectors. The solid arrows represent the 5' and 3' flanking regions of the I-Ppo recognition site in the *B. mori* rDNA genes (without intron); the shaded arrows: CAT, HmB or Neo ORFs; the open boxes: actin promoter and terminator sequences driving the expression of the CAT; the backward slanted boxes: actin promoter and terminator with HmB; the forward slanted boxes: IE1 promoter and "terminator" sequence residues, the IE1 sequences are shown as IE1. The fragments outside the plasmid maps represent the plasmids with the same DNA components but in different orientations.

I-Ppo Donor Vectors



cells), I-Ppo target test vectors were constructed. The I-Ppo recognition sequence and flanking sequences in *Bombyx* 28S rDNA genes were obtained through genomic PCR, and cloned into pCR™ II vectors (Invitrogen) directly. The resulted vector pTAE5 (4.6 kb) contained a ~400 bp I-Ppo target sequence (nt 7-389 in B108; see Fig. 2-6; Eickbush et al, 1989) without any intron in its 3' side, while the vector pTAI6 (4.5 kb) contained a 450 bp I-Ppo target sequence (nt 7-478 in clone B78, see Fig. 2-6; Eickbush et al, 1989) with an intron located in its 3' end. Therefore, the two plasmids shared the same 5' side (~350 bp sequence) but differed at their 3' ends.

4. Other Methods

All other methods were as described in the General Methods section or in Chapter 1.

III. Results

1. Cleavage of the I-Ppo recognition sequence on plasmid DNA

The target vector pTAI6 (Fig. 2-7D) was co-transfected into Bm5 cells, together with pBmA.I-Ppo, pBmA.I-Ppo', pIE1 + pBmA.I-Ppo, or pIEA.I-Ppo' (Fig. 2-7A, B, C). Three days post-transfection, total DNA of the transfected cells was isolated, digested with HindIII and resolved on an agarose gel, which was then subjected to Southern hybridization using a 450 bp EcoRI fragment from plasmid pTAI6 as a probe, which contained the I-Ppo recognition sequence . The result showed that in

Table 2-1. The Components of the I-Ppo Transformation Vectors

A set of 14 I-Ppo donor vectors were constructed according to their conponents and orientations. The four thick columns in middle showed what genes or regions the plasmid contained, the column at the right indicated the relative positions and and orientations of these genes (referring to the 5' I-Ppo flanking sequence, also see Fig 2-8 for detail).

No.	plasmid	Reporter CAT (➜)	Selective Genes		Flanking Sequences		Orientation (refering to 5' I-Ppo flanking sequence)
			Neo (⇨)	HmB (→)	no intron	with intron	
1	pICN1	✓	✓		✓		➜ ⇨
2	pICN3	✓	✓		✓		➜ ⇦
3	pICN2	✓	✓		✓		⇦ ⬅
4	pICN4	✓	✓		✓		⇨ ⬅
5	pICH1	✓		✓	✓		➜ →
6	pICH3	✓		✓	✓		➜ ←
7	pICH2	✓		✓	✓		← ⬅
8	pICH4	✓		✓	✓		→ ⬅
9	pI'CN1	✓	✓			✓	➜ ⇨
10	pI'CN2	✓	✓			✓	⇦ ⬅
11	pI'CH1	✓		✓		✓	➜ →
12	pI'CH3	✓		✓		✓	➜ ←
13	pI'CH2	✓		✓		✓	← ⬅
14	pI'CH4	✓		✓		✓	→ ⬅

presence of I-Ppo expression vectors, the 4.5 kb linerized pTAI6/HindIII fragment was cut into a 4.1 kb and a 0.35 kb fragment, which fits the I-Ppo cutting pattern (see Fig.2-9). Furthermore, the stronger promoter used in expression vector, the stronger signals obtained on the Southern blot. This shows that I-Ppo expression vectors are capable of expressing the endonuclease and these enzymes are able to cleave the target site in the plasmid DNA.

2. I-Ppo Cleavage of Genomic rDNA in Bm5 Cells

The expression vectors pBmA.I-Ppo, pBmA.I-Ppo' and pIEA.I-Ppo' were also transfected into Bm5 cells separately. One week post-transfection, total DNA of the transfected cells was isolated, digested with BamHI and BglI, resolved on the agarose gels, and subjected to Southern hybridizations using a 450 bp (with intron, from pTAI6) and a 550 bp (without intron, from pTAE5) fragment containing the I-Ppo recognition sequences as probes (see Materials and Methods). The results showed that upon transfection of the expression vectors, the 1.3 kb BamHI/BglI fragment generated from intron-free genomic rDNA genes, which contained the I-Ppo site, was cleaved into a 0.9 kb and a 0.4 kb fragment (see Fig. 2-10A). In addition, the 1.2 kb BamHI/BglI fragment from intron-containing genomic rDNA gene, which also contained an I-Ppo site, was cleaved into a 0.9 kb and a 0.3 kb fragment. Again, the stronger the promoter used in the expression vectors, the stronger the signals obtained from cleaved products in the Southern blots (Fig. 2-10B). This further proved that the expressed I-Ppo endonuclease was capable of being transported into the host nuclei and

Figure 2-9: Southern blot of HindIII digested total DNA from the transfected Bm5 cells, probed by the I-Ppo target site DNA sequence. pTAI6 plasmid containing I-Ppo target site was introduced into Bm5 cells and cleaved by the co-transfected I-Ppo endonuclease. A 3 μg of total DNA was restricted by HindIII and the 4.5 kb pTAI6 plasmid was linearized. Beside the major 4.5kb bands, a 4.1 kb and 0.3 kb bands were observed only in the I-Ppo co-transfected samples, which proved that I-Ppo expression vector can produce functional I-Ppo endonuclease in Bm5 cells. It was noticed that the stronger promoter the expression vector has, the stronger cleavage observed. pIEA-I-Ppo vector with IE-1 enhanced actin promoter yielded the strongest signal, while the pBmA.I-Ppo and pBmA. I-Ppo' gave the weaker singals at the compariable intensity. The bands shown in the middle of gel (~1.2kb) may come from the *Bombyx* 28S rDNA short repeat sequences which share some homology with the probe.

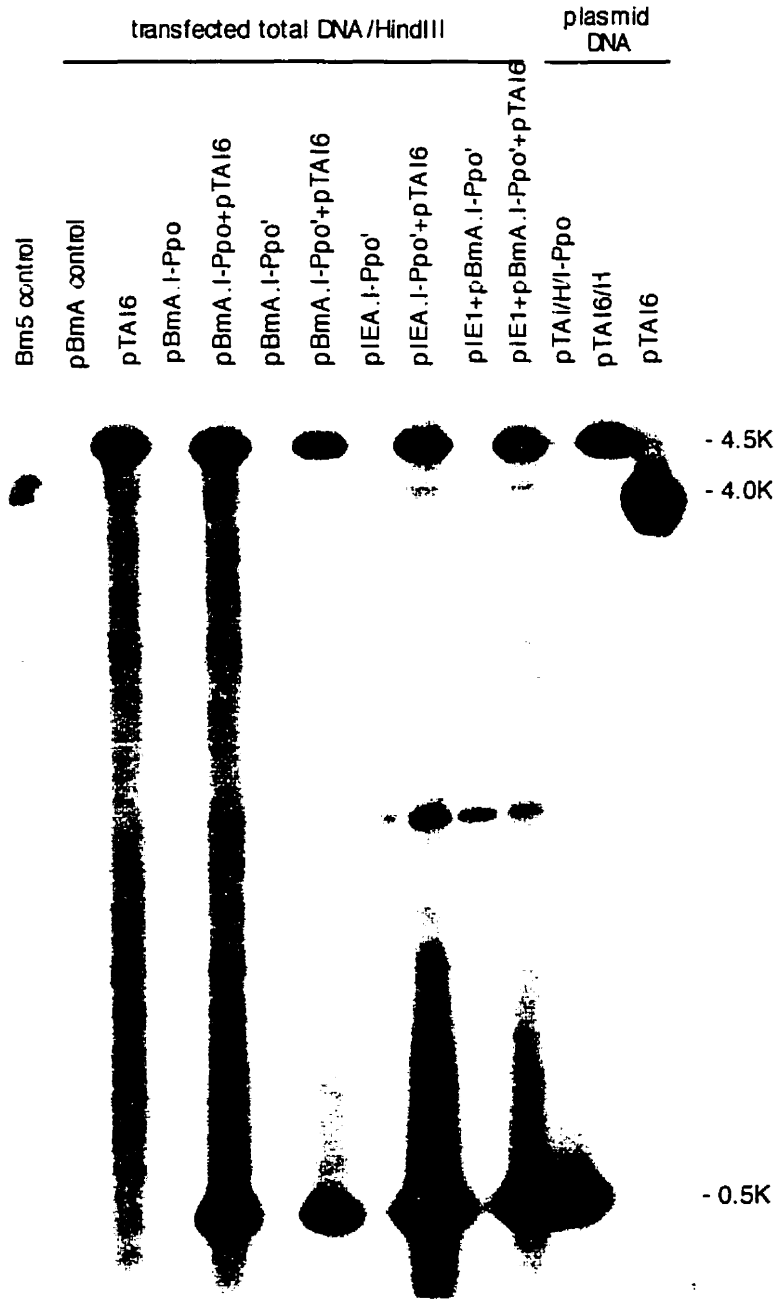
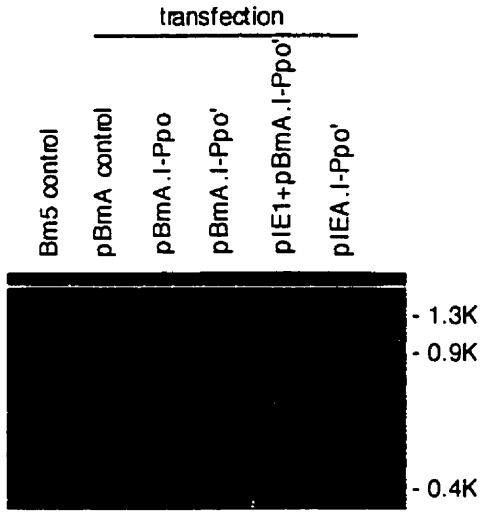
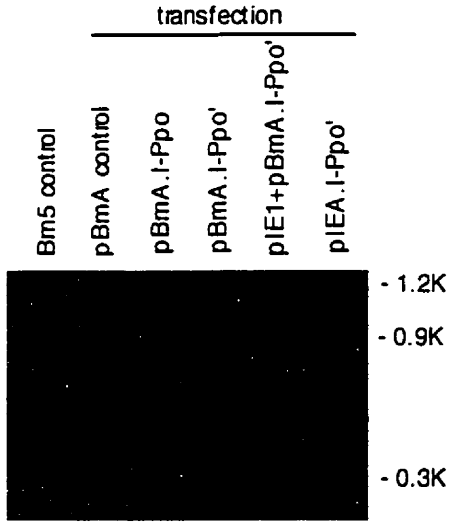

transfected total DNA/HindIII
plasmid DNA
Bm5 control
pBmA control
pTA16
pBmA.I-Ppo
pBmA.I-Ppo+pTA16
pBmA.I-Ppo'
pBmA.I-Ppo'+pTA16
pIEA.I-Ppo'
pIEA.I-Ppo'+pTA16
pIE1+pBmA.I-Ppo'
pIE1+pBmA.I-Ppo'+pTA16
pTA1/H/I-Ppo
pTA16/I1
pTA16
- 4.5K
- 4.0K
- 0.5K

Figure 2-10: Southern blots of BamHI/BglI digested genomic DNAs from the I-Ppo transfected Bm5 cells, probed by the *Bombyx* 28S rDNA DNA sequence containing I-Ppo target site. The I-Ppo expression vectors were transfected into Bm5 cells and the expressed I-Ppo endonuclease cleaved the target sites within *B. mori* 28S rDNA. Three μg of total DNA was restricted by BamHI/BglI and the 1.3 kb (without intron downstream of the target site, Panel A) or 1.2 kb (with intron, Panel B) fragments were produced (see also clones B108 and B78 in Fig.2-6). Besides the major 1.3 kb and 1.2 kb bands, a pair of 0.9 kb and 0.4 kb bands or a pair of 0.9 kb and 0.3 kb bands were observed only in the I-Ppo transfected samples (see also Fig.2-6). This proves that the I-Ppo expression vector can produce functional I-Ppo endonuclease in Bm5 cells. Again, it was noticed that the stronger promoter the expression vector has, the stronger cleavage observed, which was consistent with the plasmid cutting patterns. The pIEA-I-Ppo vector with IE-1 enhanced actin promoter yielded the strongest signal, while the pBmA.I-Ppo and pBmA. I-Ppo' vectors gave the weaker expressions. The bands between 0.5 -0.9 kb may be due to *Bombyx* 28S rDNA short repeat sequences which share some homology with both probes. A: two days exposure; B: a 12 hour exposure.

A: E5 probe



B: I6 probe



cleaving the target sites on genomic DNA.

3. Co-transfection of the I-Ppo Expression Vectors and the Donor Vectors

The expression vectors pBmA.I-Ppo or pIEA.I-Ppo' (Fig. 2-7)were co-transfected with the donor vectors pICN1, 2 and pICH1, 2 respectively into Bm5 cells. Four weeks after the initial transfection, Neomycin and Hygromycin selection, respectively, were applied for 6 weeks. However, none of the co-transfection groups survived the selection process, indicating that the I-Ppo endonuclease level might have been too high and might have caused too much damage in the host genomes.

IV. Discussion

In this chapter, we have explored the use of I-Ppo endonuclease to introduce DSBs in the *Bombyx* genome to induce gene conversion. Our data shows that the I-Ppo expression vectors work very efficiently in the *Bombyx* cell line. In this expression system, the I-Ppo gene, originated from *P. polycephalum*, can use the host protein synthetic apparatus to produce the endonuclease protein. It seems that the foreign I-Ppo endonuclease can keep its cleavage function in the trans-kingdom host. It was shown to be able to cleave extra-chromosomal DNA (transfected plasmid in host cells), and to be transported into the host nuclei and made cleavages on genomic DNA.

Our data also indicates that the extent of site-specific DSBs in plasmid or genomic DNA paralleled the I-Ppo levels of expression controlled by the promoter

elements present in the expression vectors. Southern analysis showed that the degree of DSB at the I-Ppo recognition site was directly related to the strength of the promoter that controls the I-Ppo endonuclease level in the host cells. These results suggest that the DSBs were directly caused by the transfected I-Ppo endonuclease gene.

The I-Ppo expression level obtained from the pBmA.I-Ppo and pIEA.I-Ppo' plasmids in Bm5 cells was, however, apparently very high. This was indicated by the fact that the host cells that were transfected with such plasmids could not survive the selection pressure. To obtain the balance between sufficient level of DSB and host tolerance, lower activity I-Ppo expression vectors and/or amounts of DNA used for transfection should be tested in the future.

Rare-cutting endonucleases provide a powerful tool for genome manipulation. Several studies have shown that expression of such endonucleases in mammalian cells "stimulated" recombination by initiating gene conversion at DSBs [Choulika et al., 1995; Cohen-Tannoudji et al., 1998]. Here we constructed a series of I-Ppo expression vectors and the donor vectors, and tested their functions in *Bombyx* cell lines. Our strategy was to introduce the I-Ppo endonuclease and the I-Ppo recognition site, interrupted with the desired genes, into Bm5 cells, and produce DSBs at the known sites in genomic DNA and facilitate the alignment of homologous sequences at these sites.

Despite the success in expressing I-Ppo in Bm5 cells, a preliminary PCR screening (see Appendix 3 for detail) failed to detect foreign gene "homing" events in long-term (4 weeks) co-transfection of I-Ppo expression and donor vectors. This may be caused by trivial problems related to the PCR process such as the condition of the reaction for long

product amplification or the choice of the primers. Besides the possible artifactual errors, the results suggest that the principle of the I-Ppo system works (it does produce DSBs in the *Bombyx* genome). However, we anticipate that problems may appear when such a system is employed as a practical gene transformation tool. One of the potential obstacles is that the homing sites are located within a rRNA gene and will undergo RNA processing. Therefore, the integrated foreign gene may be cut out along with the introns and not be expressed properly. The other problem is that the I-Ppo cleavage activity may cause the cell death, as we and other researchers [Muscarella et al., 1993] have encountered: without transfection of I-Ppo, DSBs will not occur in the genome, and the directed integration of antibiotic gene will not happen, therefore the cells will die under the selection pressure. On the other hand, with transfection of I-Ppo, the cells may gain the antibiotic genes through the I-Ppo directed integration and survive the antibiotic selection. However the cells will still die because of the lethal effect of I-Ppo expression (cleavage of rDNA genes).

The dilemma can be solved by applying the same strategy using other rare restriction enzymes which have less lethal effect on host. A rare restriction enzyme can be employed to make DSB at a limited number of sites yielding two 3'-protruding free ends which will be the primers for the gene conversion process. A cloned genomic sequence containing such sites with foreign genes inserted in them can be used for alignment of homologous sequences (prerequisite for gene conversion). Provided the rare endonuclease can be shown to be capable of introducing double-stranded breaks at the target sites in the *Bombyx* genome, and a selection of a spectrum of the enzymes that may reduce the lethal

effect can be defined, this alternative strategy may be more applicable.

PERSPECTIVES AND FUTURE DIRECTIONS

We have established a *Bombyx mori* gene transformation system using the Mariner transposable element, and explored the feasibility of I-Ppo endonuclease mediated gene transformation system. In principle, neither system has much restriction in term of host choices, as long as the transposase or endonuclease can be expressed in the host. This means that they can be used as a universal gene transformation system.

Although both systems initiate the transposition by introducing the DSB at the insertion sites, they use quite different mechanisms to accomplish the transposition process. Mariner element integrates at TA sites, and this essentially corresponds to random integration in terms of site-specificity. Mariner may also have a size restriction in terms of the foreign gene it can carry.

On the other hand, the I-Ppo mediated system has rather high site-specificity and is highly efficient in initiating gene conversions in the natural systems. This proved that the principle is sound; however, in practice, it may not be feasible. An alternative approach is to use the rare restriction endonucleases with relevant low lethal effect. Actually the DSB lethal effect on the host can be overcome by adjusting the I-Ppo expression level or amount of transfected DNA. However, relative to Mariner, more effort may be required to optimize the transfection condition for every specific host.

The results presented in this thesis show significant advances in insect gene

transformation. Although the Mariner gene transformation system has been successfully established in a *Bombyx* cell line, there are still a lot of unanswered questions. The most urgent ones relate to a possible increase in the transformaton efficiency and to the application of this system to embryo manipulations for establishing transgenic animals. Attempts can be made to alter the insertion site of foreign genes in the Mariner vector, and to try the DNA microinjection process on *Bombyx* eggs. For the I-Ppo system, the PCR detection of integrated sequences in the short term transfected cells must be completed in order to finish the feasibility study. By adjusting the amount of transfected DNA, a carefully controlled balance between the expression level of the enzyme (maybe lethal if too high) and sufficient homing efficiency could be reached. The integration events could be selected by going through the selection process. There is also a possibility of using other rare restriction enzymes with less lethal impact on the host as an alternative approach. Finally, both elements might be integrated into virus transformation vectors to establish a novel gene transformation system, in which a virusl element is used to overcome the barrier of DNA entry, while transposable elements are used to facilitate genomic integration.

BIBLIOGRAPHY

Akins RA, Lambowitz AM. A protein required for splicing group I introns in Neurospora mitochondria is nitochondrial tyroyl-tRNA synthetase or a derivative thereof. Cell. 50:331-345, 1987.

Alberts B. Bray D. Lewis J. Raff M. Roberts K. Watson JD. Molecular Biology of the Cell. New York: Garland, 2nd ed, 1989.

Anziano PQ, Hanson DK, Mahler JR, Perlman PS. Functional domains in introns: tras-acting and cis-acting regions of intron 4 of the cob gene. Cell 30:925-932, 1982.

Bailey JM, Davidson N. Methylmercury as a reversible denaturing agent for agarose gel electrophoresis. Analyt. Biochem. 70:75-85, 1976

Belfort M, Perlman PS. Mechanisms of intron mobility. J. Biol. Chem. 270:30237-40, 1995.

Belfort M, Reaban ME, Coetzee T, Dalgaard JZ. Prokaryotic introns and inteins: A panoply of from and function. J Bacteriol 177:3897-3903, 1995.

Belfort M, Boberts R. Homing endonucleases --- Keeping the home in order. Nucleic Acids Res. 25:3379-3388, 1997.

Berg DE. Howe MM. eds. Mobile DNA. *Washington. DC. Am. Soc. Microbiol.* 1989.

Bollag RJ, Waldman AS, Liskay RM. Homologous recombination in mammalian cells. Annu. Rev. Genet. 23:199-225, 1989.

Boussy IA., Healy MJ, Oakeshott JG, Kidwell MG. Molecular analysis of the P-M

gonadal dysgenesis subline in Australian Drosophila melanogaster. Genetics 119:889-902

Bradford MM. A rapid and sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem.72:248-254, 1976.

Bronson SK, Smithies O. Altering mice by homologous recombination using embryonic stem cells. J.Biol. Chem. 269:27155-8, 1994.

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

Bryan GJ. Jacobson JW. Hartl DL. Heritable somatic excision of a Drosophila transposon. Science. 235:1636-8, 1987.

Buck W.D., Eickbush D.G., Xiong Y., Jakubczak J. & Eickbush T.H. 1993 Sequence relationship of retransposable elements R1 and R2 within and between divergent insect species. Mol.Biol. Evol. 10:163-185

Buck W.D., Calang C.C. & Eickbush T.H. 1987 The site-specific ribosomal insertion element type II of *Bombyx mori* (R2Bm) contains the coding sequence for a reverse transcriptase-like enzyme. Mol. Cell. Biol. 7:2221-2230

Capy P. Koga A. David JR. Hartl DL. Sequence analysis of active mariner elements in natural populations of *Drosophila simulans*. Genetics. 130:499-506, 1992,

Capy P. Langin T. Bigot Y. Brunet F. Daboussi MJ. Periquet G. David JR. Hartl DL. Horizontal transmission versus ancient origin: mariner in the witness box. Genetica. 93:161-70, 1994.

Cech T.R. Self-splicing of group I introns. Annu. Rev. Biochem. 59:543-568, 1990.

Cech TR. Conserved sequences and structures of group I introns: building an active site for RNA catalysis. Gene 73:259-271, 1988.

Cech TR, Bass BL. Biological catalysis of RNA. Annu. Rev. Biochem. 55:599-629, 1986.

Choulika A, Perrin A, Dujon B Nicolas JF. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. Mol. Cell. Biol. 1968-73, 1995.

Coates CJ. Turney CL. Frommer M. O'Brochta DA. Warren WD. Atkinson PW. The transposable element mariner can excise in non-drosophilid insects. Mol. & Gen. Genet. 249:246-52, 1995.

Coates CJ. Turney CL. Frommer M. O'Brochta DA. Atkinson PW. Interplasmid transposition of the mariner transposable element in non-*drosophilid* insects. Mol. & Gen. Genet. 253:728-33, 1997.

Cohen RS, Messelson M. Separate regulatory elements of the heat-inducible and ovarian expression of Drosophila hsp26 gene. Cell 43:737-746, 1985

Cohen-Tannoudji M, Robine S, Choulika A, Pinto D, Marjou FE, Babinet C, Louvard D, Jaisser F. I-SceI-induced gene replacement at a natural locus in embryonic stem cells. Mol. Cell. Biol. 18:1444-1448, 1998.

Crabb DW. Minth CD. Dixon JE. Assaying the reporter gene chloramphenical acetyltransferase. Methods Enzymol. 168:690-701, 1989.

Deng C, Capecchi MR. Reexamination of gene targeting frequency as a function of

the extent of homology between the targeting vector and the target locus. Mol. Cell. Biol. 12:3365-3371, 1993.

Doak TG. Doerder FP. Jahn CG Herrick AG. A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. Proc. Natl. Acad. Sci. USA 91:942-946, 1994.

Eickbush T.H. & Robins B. *Bombyx mori* 28s ribosomal genes contains insertion elements similar to the type I and II elements of Drosophila melanogaster. EMBO J. 4:2281-2285, 1985.

Ellison EL, Vogt VM. Interaction of the intron-encoded mobility endonuclease I-PpoI with its target site. Mol. Cell. Biol. 13:7531-7539, 1996.

Engels WR. P elements in Drosophila melanogaster. in Mobile DNA. Berg & Howe eds. *Washington DC. Am. Soc. Microbiol.* 437-484. 1989.

Fairhead C, Dujon B. Consequences of unique double -stranded breaks in yeast chromosomes: death or homozygosis. Mol. Gen. Genet. 240:170-180, 1993.

Fajas L. Staels B. Auwerx J. Cycle sequencing on large DNA templates. Biotechniques. 23:1034-1036, 1997.

Farrell PJ, Lu M, Prevost J, Brown C, BehieL and Iatrou K. High-level expression of secreted glycoproteins in transformed Lepidopteran insect cells using a novel expression vector. Biotechnol. & Bioengin. in press.

Feinberg AP. Vogelstein B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Analyt. Biochem. 132:6-13, 1983.

Finnegan DJ. Eukaryotic transposable element and genome evolution. Trends Genet.

5:103-107, 1989.

Finnegan DJ. Transposable elements: how non-LTR retrotransposons do it. Curr. Biol. 7:R245-R248, 1997.

Fisher JA, maniatis T. Regulatory elements involved in Drosophila Adh gene expression are conserved in divergent species and separate elements mediated expression in different tissues. EMBO J. 5:1275-1289, 1986.

Flick KE. McHugh D. Heath JD. Stephens KM. Monnat Jr. RJ. Stoddard BL. Crystallization and prelimary X-ray studies of I-PpoI: A nuclear, intron-encoded homing endonuclease from *Physarum polycephalum*. Prot. Sci. 6:2677-80, 1997.

Fotaki ME. Iatrou K. Identifcation of a transcriptionally active pseudogene in the chorion locus of the silkwoth *Bombyx mori*: regional sequence conservation and biological function. J. Mol. Biol. 203:849-860, 1988.

Fujikawa H., Ogura T., Takada N., Miyajima N., Ishikawa H. & Maekawa H. Introns and their flanking sequences of *Bombyx mori* rDNA. Nucl. Acids Res. 12:6861-6869, 1984.

Gage LP. Polyploidization of the silk gland of B. mori. J. Mol. Biol. 86:97-108, 1974.

Garcia-Fernandez J. Bayascas-Ramirez JR. Marfany G. Munoz-Marmol AM. Casali A. Baguna J. Salo E. 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, 1995.

Garza D. Medhora M. Koga A. Hartl DL. Introduction of the transposable element mariner into the germline of Drosophila melanogaster. Genetics. 128:303-10, 1991.

Golic MM. Rong YS. Petersen RB. Linqusit SL. Golic KG. FLP-mediated DNA mobilization to specific target sites in Drosophila chromosomes. Nucl. Acids Res. 25:3665-3671, 1997.

Grabowski PJ. Zaug AJ. Cech TR. The intervening sequence of the ribosomal RNA precursor is converted to a circular RNA in isolated nuclei of *Tetrahymena*. Cell, 23:467-476, 1981.

Grace TDC. Establishment of a line of cells from the silkworm *Bombyx mori*. Nature. 216:613, 1967.

Grindley NDF. Leschziner AE. DNA transposition: from balck box to color monitor. Cell 83:1063-1066, 1995.

Gritz L. Davies J. Plasmid encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in Escherichia coli and Saccharomyces cerevisiae. Gene 25:179-188, 1983.

Gueiros-Filho FJ. Beverley SM. Trans-kingdom transposition of the Drosophila element mariner within the protozoan Leishmania. Science. 276:1716-1719, 1997.

Handler AM. McCombs SD. Fraser MJ. Saul SH. The lepidopteran transposon vector, *PiggyBac*, mediates germ-line transformation in the *Mediterranean* fruit fly. Proc. Natl. Acad. Sci. USA. 95:7520-7525, 1998.

Hartl DL. Lohe AR. Lozovskaya ER. Modern thoughts on an ancyent marinere: function, evolution, regulation. Ann. Rev. Genet. 31:337-358, 1997.

Hartl DL. Lozovskaya ER. Nurminsky DI. Lohe AR.(1) What restricts the activity of mariner-like transposable elements. Trends Genet. 13:197-201, 1997.

Heinemann J.A. 1991 Genetics of gene transfer between species. (Review) Trends Genet. 7:181-185

Iatrou K. Ito K. Witkiesicz H. Polyhedrin gene of *Bombyx mori* nuclear polyhedrosis virus. J. Virol. 54:436-445, 1985.

Iatrou K. Meidinger RG. Tissue-specific expression of silkmoth chorion genes in vivo using *Bombyx mori* nuclear polyhedrosis virus as a transducing vector. Proc. Natl. Acad. Sci. USA. 87:3650-3654, 1990.

Iatrou K. Meidinger RG. *Bombyx mori* nuclear polyhedrosis virus-based vectors for expressing passenger genes in silkmoth cells under viral or cellular promoter control. Gene 75:59-71, 1989.

Ivics Z. Hackett PB. Plasterk RH. Izsvak Z. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell 91:501-510, 1997.

Jacobson JW. Medhora MM. Hartl DL. Molecular structure of a somatically unstable transposable element in Drosophila. Proc. Natl. Acad. Sci. USA. 83:8684-8688, 1986.

Jarvis DL. Effects of baculovirus infection on Ie1-mediated foreign gene expression in stably transformed insect cells. J. Virol. 67:2583-2591, 1993.

Johnson R. Meidinger RG. Iatrou K. A cellular promoter-based expression cassette for generating recombinant baculoviruses directing rapid expression of passenger genes in infected insects. Virology. 190:815-823, 1992.

Keeler KJ. Dray T. Penney JE. Gloor GB. Gene trageting of a plasmid-borne sequence to a double strand DNA break in Drosophila melanogaster. Mol. Cell. Biol. 16:522-

528, 1996.

Kidwell MG. Evolutionary biology. Voyage of an ancient mariner. Nature 362:202, 1993.

Lambowitz A.M. Infectious introns. Cell 56:323-326, 1989.

Lambowitz A.M., Belfort M. Introns as mobile genetic elements. Ann. Rev. Biochem. 62:587-622, 1993.

Lampe DJ. Churchill ME. Robertson HM. A purified mariner transposase is sufficient to mediate transposition in vitro. EMBO J. 15:5470-9, 1996 .

Levis R, O'Hare K, Rubin G. Effects of transposable element insertions on RNA encoded by the white gene of Drosophila. Cell 38:471-482

Lidholm DA. Lohe AR. Hartl DL. The transposable element mariner mediates germline transformation in Drosophila melanogaster. Genetics 134:859-68, 1993.

Lohe AR. De Aguiar D. Hartl DL. Mutations in the mariner transposase: the D,D(35)E consensus sequence is nonfunctional. Proc. Natl. Acad. Sci. USA. 94:1293-1297, 1997.

Lohe AR. Sullivan DT. Hartl DL. Subunit interactions in the mariner transposase. Genetics 144:1087-1095, 1996.

Lohe AR. Hartl DL. Autoregulation of mariner transposase activity by overproduction and dominant-negative complementation. Mol. Biol. & Evol. 13:549-555, 1996.

Lohe AR. Hartl DL. Reduced germline mobility of a mariner vector containing exogenous DNA: effect of size or site?. Genetics 143:1299-1306, 1996.

Loukeris TG. Livadaras I. Arca B. Zabalou S. Savakis C. Gene transfer into the

medfly, *Ceratitis capitata*, with a *Drosophila hydei* transposable element. Science 270:2002-2005, 1995.

Lu M, Farrell P, Johnson R, Iatrou K. A baculovirus (Bombyx mori nuclear polyhedrosis virus) repeat element functions as a powerful constitutive enhancer in transfected insect cells. J. Biol. Chem. 272:30724-30728, 1997.

Lukacsovich T, Yang D, Waldman AS. Repair of a specific double-strand break generated within a mannalian chromosome by yeast endonuclease I-SceI. Nucl. Acids Res. 22:5649-5657, 1994.

Marshall A. The insects are coming. Nature Biotech. 16:530-533, 1998.

Maruyama K. Schoor KD. Hartl DL. Identification of nucleotide substitutions necessary for trans-activation of mariner transposable elements in Drosophila: analysis of naturally occurring elements. Genetics 128:777-784, 1991.

Medhora MM. MacPeek AH. Hartl DL. Excision of the Drosophila transposable element mariner: identification and characterization of the Mos factor. EMBO J. 7:2185-2189, 1988.

Mizuuchi K. Transpositional recombination: mechanistic insights from studies of Mu and other elements. Ann. Rev. Biochem. 61:1011-1051, 1992.

Monteilhet C. Perrin A. Thierry A. Colleaux L. Dujon B. Purification and characterization of the in vitro activity of I-Sce I, a novel and highly specific enconuclease encoded by a group I intron. Nucleic Acids Res. 18:1407-1413, 1990.

Moritz T. Williams DA. Gene transfer into the hematopoietic system. [Review] Curr. Op. Hematology, 1:423-428, 1994.

Mueller JE. Bryk M. Loizos N. Belfort M. Homing endonucleases. In Nucleases, Linn SM. Lloyd RS. Roberts RJ. Eds., Cold Spring Harbor Press, Cold Spring Harbor, NY, 111-143, 1993.

Muscarella D.E., Ellison E.L., Ruoff B.E. & Vogt V.M. Characterization of I-Ppo, an intron-encoded endonuclease that mediates homing of a group I intron in the ribosomal DNA of Physarum polycephalum. Mol. Cell. Biol. 10:3386-3396, 1990.

Muscarella D.E. & Vogt V.M. A mobile group I intron in the nuclear rDNA of Physarum polycephalum. Cell 56: 443-454, 1989.

Muscarella D.E. & Vogt V.M. A mobile group I intron from Physarum polycephalum can insert itself and induce point mutation in the nuclear ribosomal DNA of Sacchrauomyces cerevisiae. Mol. Cell. Biol. 13:1023-1033, 1993.

Nassif N. Penney J. Pal S. Engels WR. Gloor G. Efficient copying of non-homologous sequences from ectopic sites via P-element-induced gap repair. Mol. Cell. Biol. 14:1613-1625, 1994.

O'Brochta DA. Atkinson PW. Transposable elements and gene transformation in non-drosophilid insects. Insect Biochem. & Mol. Biol. 26:739-53, 1996.

Orr-Weaver CJ, Spradling AC. Drosophila chorion gene amplication requires an upstream region regulating s18 transcription. Mol. Cell. Biol. 6:4624-4633, 1986.

Plasterk RH. The Tc1/mariner transposon family. Curr. Top. Microbiol. & Immunol. 204:125-143, 1996.

Plasterk RH. The origin of footprints of the Tc1 transposon of *Caenorhabditis elegans*. EMBO J. 10:1919-1925, 1991.

Rasmunson KE. Raymon JD, Simmons MJ. Repression of hybrid dysgenesis in Drosophila melanogaster by individual naturally occurring P elements. Genetics 133:605-622, 1993.

Rio DC. Regulation of *Drosophila* P element transposition. Trends Genet. 7:2822-87, 1991.

Robbins PD. Tahara H. Ghivizzani SC. Viral vectors for gene therapy. Trends In Biotechnol. 16:35-40, 1998.

Robertson HM. The mariner transposable element is widespread in insects. Nature 362:241-5, 1993.

Robertson HM. Asplund ML. Bmmar1: a basal lineage of the mariner family of transposable elements in the silkworm moth, *Bombyx mori*. Insect Biochem. & Mol. Biol. 26:945-954, 1996.

Robertson HM. Lampe DJ. Distribution of transposable elements in arthropods. Ann. Rev. Entomol. 40:333-357, 1995.

Robertson HM. MacLeod EG. Five major subfamilies of mariner transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. Insect Mol. Biol. 2:125-139, 1993.

Rouete P, Smih F, Jasin M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Proc. Natl. Acad. Sci. USA. 91:6064-6068, 1994.

Russell DW. Hirata RK. Human gene targeting by viral vectors. Nature Genet. 18:325-330, 1998.

Sambrook J, Fritsch EF, Maniatis T. "Molecular Cloning: A Laboratory Manual". Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989.

Schowalter DB and Sommer SS. The generation of radiolabeled DNA and RNA probes with Polymerase Chain Reaction. Analyt. Biochem. 177:90-94, 1989.

Shub DA, Gott JM, Xu MQ, Lang BF, Michel F, Tomaschewski J, Pedersen-Lane J, Belfort M. Structural conservation among three homologous intron of bacteriophage T4 and the group I intron of enkaryotes. Proc. Natl. Acad. Sci. USA. 85:1151-1155, 1988.

Simmons MJ. Bucholz LM. Transposase titration in Drosophila melanogaster: a model of cytotype in the P-M system of hybrid dysgenesis. Proc. Natl. Acad. Sci. USA 82:8119-8123, 1985.

Smith F, Rouet P, Romanienko J, Jasin M. Double -strand breaks at the target locus stimulate gene targeting in embryonic stem cells. Nucl. Acids Res. 23:5012-5019, 1995.

Song SU. Gerasimova TI. Kurkulos M. Boeke JD. Corces VG. An env-like portein encoded by a Drosophila retro element: evidence that gypsy is an infectious retrovirus. Genes Dev. 8:204620-57, 1994.

Thierry A. Perrin A. Boyer J. Fairhead C. Dujon B. Frey B. Schmitz G. Cleavage of yeast and bacteriophage T7 genomes at a single site using the rare cutter endonuclease I-SceI. Nucl. Acids Res. 19:189-190, 1991

Tomita S. Sohn BH. Tamura T. Cloning and characterization of a mariner-like element in the silkworm, *Bombyx mori*. Genes Genet. Systems 72:219-28, 1997.

van Luenen HG. Colloms SD. Plasterk RH. The mechanism of transposition of Tc3 in C. elegans. Cell 79:293-301, 1994.

Vile RG. Russell SJ. Retroviruses as vectors. British Medical Bulletin. 51:12-30, 1995.

Vos JC. De Baere I. Plasterk RH. Transposase is the only nematode protein required for in vitro transposition of Tc1. Genes Dev. 10:755-761, 1996.

Vos JC. van Luenen HG. Plasterk RH. Characterization of the *Caenorhabditis clegans* Tc1 transposase in vivo and in vitro. Genes Dev. 7:1244-1253, 1993.

Weiss SA, Smith GC, Kalter SS, Baughn JL. Improved method for the production of insect cell cultures in large volume. In Vitro 17:495-502.

Welsh S. Kay SA. Reporter gene expression for monitoring gene transfer. Curr. Op. Biotechnol. 8:617-622, 1997.

Wernette C. Saldanha R. Smith D. Ming D. Perlman P.S., Butow R.A. Complex recognition site for the group I intron-encoded endonuclease I-SceII. Mol. Cell Biol. 12:716-723, 1992.

Whittmayer PK, Raines RT. Substrate binding and trunover by the highly specific I-PpoI endonuclease. Biochemistry 35:1076-1083, 1997.

Woodson SA. Cech TR. Reverse self-splicing of the *Tetrahymena* group I intron: implication for the directionality of splicing and for intron transposition. Cell 57:335-345, 1989.

Yamakawa H. Ohara O. A DNA cycle sequencing reaction that minimizes compressions on automated fluorescent sequencers. Nucl. Acids Res. 25:1311-1312, 1997.

Zaug AJ. Kent JR. Cech TR. A labile phosphodiester bond at the ligation junction in a circular intervening sequence RNA. Science 224:574-578.

Zhang J. Kalogerakis N. Behie LA. Iatrou K. Optimum infection conditions for recombinant protein production in insect cell (Bm5) suspension culture. Biotechnol. Progr. 10:636-643, 1994.

APPENDICES

Appendix 1. The DNA sequence and the translated amino acid sequence of Mos1 mariner transposable element [Genebank, accession number M14653; see Jacobson et al., 1986].

The non-perfect terminal inverted repeats are underlined. The sequence between the arrows (the SspI-HindIII fragment) has been inserted into the pBmA to construct a mariner transposase expression vector.

```
1    CCAGGTGTAC AAGTAGGGAA TGTCGGTTCG AACATATAGA TGTCTCGCAA
                ↓ (SspI site)
51   ACGTAAATAT TTATCGATTG TCATAAAACT TTGACCTTGT GAAGTGTCAA
101  CCTTGACTGT CGAACCACCA TAGTTTGGCG CGAATTGAGC GTCATAATTG
151  TTTACTCTCA GTGCAGTCAA CATGTCGAGT TTCGTGCCGA ATAAAGAGCA
                           M  S  S  F  V  P  N  K  E  Q
201  AACGCGGACA GTATTAATTT TCTGTTTTCA TTTGAAGAAA ACAGCTGCGG
      T  R  T  V  L  I  F  C  F  H  L  K  K  T  A  A  E
251  AATCGCACCG AATGCTTGTT GAAGCCTTTG GCGAACAAGT ACCAACTGTG
       S  H  R  M  L  V  E  A  F  G  E  Q  V  P  T  V
301  AAAACGTGTG AACGGTGGTT TCAACGCTTC AAAAGTGGTG ATTTTGACGT
     K  K  C  E  R  W  F  Q  R  F  K  S  G  D  F  D  V
351  CGACGACAAA GAGCACGGAA AACCGCCAAA AAGGTACGAA GACGCCGAAC
      D  D  K  E  H  G  K  P  P  K  R  Y  E  D  A  E  L
401  TGCAAGCATT ATTGGATGAA GACGATGCTC AAACGCAAAA ACAACTCGCA
       Q  A  L  L  D  E  D  D  A  Q  T  Q  K  Q  L  A
451  GAGCAGTTGG AAGTAAGTCA ACAAGCAGTT TCCAATCGCT TGCGAGAGAT
     E  Q  L  E  V  S  Q  Q  A  V  S  N  R  L  R  E  M
501  GGGAAAGATT CAGAAGGTCG GTAGATGGGT GCCACATGAG TTGAACGAGA
      G  K  I  Q  K  V  G  R  W  V  P  H  E  L  N  E  R
551  GGCAGATGGA GAGGCGCAAA AACACATGCG AAATTTTGCT TTCACGATAC
       Q  M  E  R  R  K  N  T  C  E  I  L  L  S  R  Y
601  AAAAGGAAGT CGTTTTTGCA TCGTATCGTT ACTGGAGATG AAAAATGGAT
     K  R  K  S  F  L  H  R  I  V  T  G  D  E  K  W  I
651  CTTTTTTGTT AATCCTAAAC GTAAAAAGTC ATACGTTGAT CCTGGACAAC
     F  F  V  S  P  K  R  K  K  S  Y  V  D  P  G  Q  p
701  CGGCCACATC GACTGCTCGA CCGAATCGCT TTGGCAAGAA GACGATGCTC
       A  T  S  T  A  R  P  N  R  F  G  K  K  T  M  L
751  TGTGTTTGGT GGGATCAGAG CGGTGTCATT TACTATGAGC TCTTGAAACC
```

```
     C  V  W  W   D  Q  S   G  V  I   Y  Y  E  L   L  K  R
801  CGGCGAAACG GTGAATACGG CACGCTACCA ACAACAATTG ATCAATTTGA
      G  E  T   V  N  T  A   R  Y  Q   Q  Q  L   I  N  L  N
851  ACCGTGCGCT TCAGAGAAAA CGACCGGAAT ATCAAAAAAG ACAACACAGG
       R  A  L   Q  R  K   R  P  E  Y   Q  K  R   Q  H  R
901  GTCATTTTTC TCCATGACAA CGCTCCATCA CATACGGCAA GAGCGGTTCG
     V  I  F  L   H  D  N   A  P  S   H  T  A  R   A  V  R
951  CGACACGTTG GAAACACTCA ATTGGGAAGT GCTTCCGCAT GCGGCTTACT
       D  T  L  E   T  L  N   W  E  V   L  P  H   A  A  Y  S
1001 CACCAGACCT GGCCCCATCC GATTACCACC TATTCGCTTC GATGGGACAC
       P  D  L   A  P  S   D  Y  H  L   F  A  S   M  G  H
1051 GCACTCGCTG AGCAGCGCTT CGATTCTTAC GAAAGTGTGA AAAAATGGCT
     A  L  A  E   Q  R  F   D  S  Y   E  S  V  K   K  W  L
1101 CGATGAATGG TTCGCCGCAA AAGACGATGA GTTCTACTGG CGTGGAATCC
       D  E  W   F  A  A  K   D  D  E   F  Y  W  R  G  I  H
1151 ACAAATTGCC CGAGAGATGG GAAAAATGTG TAGCTAGCGA CGGCAAATAC
       K  L  P   E  R  W   E  K  C  V   A  S  D   G  K  Y
1201 TTTGAATAAA TGATTTTTTC TTTTTCCACA AAATTTAACG TGTTTTTTGA
     L  E
1251 TTAAAAAAAA ACGACATTTC ATACTTGTAC ACCTGA.....(~200bp
```

flanking *D. mauritiana* seqence) ..... AAGCTT

↑ (HindIII site)

Appendix 2. The Insertion Sequences in Rescued Plasmids.

The underlined sequences indicate primers used in the genomic PCR reactions as described in the body of the thesis; dots indicate undetermined sequences; bold letters indicate the 5' and 3' ends of the rescued insertion fragments.

The insert in plasmid D6 (322 bp, from the TA site flanking the 3'IR to the EcoRI site in HmB, complete sequence):

TACTAGGGCT GCAGGAATTC GAGTAACGAG TTCTTTAGAA CGTATTGTTA CGCTGGTAAG
AGCAACGCCA TCTATCGCCG AATGGCCGAA CGAGTATCGA ATGATCCAGT AGCACCAAGA
ACGATCGAGA AGTACAACGC CATCTATGTC AGATAGCGGA AACACACAAT ACTCGACTTT
TGTGGAACAT TCTCGATAAT TCTGGGGATG TGGTATCGGC TATAAAAGCG TTGCAGAGAT
GGCGCGCAGT CATTCAGTAA TGGGAACTAC TCGAAGCGAA ACAGCGAACG GATCNCCTGA
AGCGAAGCGG AAGAGCGAAT TC

The insert in plasmid D8 (~3 kb, from the TA site flanking the 3'IR to the EcoRI site in HmB)

TAAATCTGTT CATCAGTGTA CAACAAGAGC TTAGTTTAGC GAGTCGAATT CTCGATAGTG
TATTATCCGA TTACGGCAAC ATTATTTTTA TATTATCAGC GATATTATTA AATTATTAGT
AATTATTGGT TTAAGTTTCC GTTTGAATAT CGAATGCCTA TACATTACAT CTAAACAGAA
GTGTGTTATG TGGGCGGCCG ATCTATTACG AAACTATTGC GCTCAAGTGA GGACGCGTAA
AGAATGGGCA TTCGCGAATA CGCAGATTTA AAAAAAAAAA ATTGCGCATC AGACTCGTAT
ATCGGTCTAG GNGACGATNA ATTTATACNC CANCGCGATC TTTTTATTGC GTTTCCGCAT
AGATACGCNG ACGCTTATCG CGANTACGCG TCGATGCTAA TTCTAATTAC GGGGCCTGTC
GTCCGANACG CCTCNCTTCT AAATCNACAG CTGAAGCATC CTATATCCAA TGTTTAAGCN
NACCCACCTG GATTATTTTA CTTTTATAAC AACTTTCCNT TCCCCTGTGA ACNTATCNAN
GTTATCCANA ANCGNTTCNT TTCCACTTTG AAAANTCGGA CCGGTGAATA AGTTTTTNTN
NTNAAACCCC GGTCCCCCCC CGTTAATTAA TAAACNTGAT TTNGGAACCT CCAATGAATT

CCTTGGNTGN C....~1.8 kb unsequenced segment....AATGACTTCG GGTGGACNAC

AAACTGCTTA AATAAAACNG ATTTAACTTT CCTGAGTGTT TAGCATTGTA ATGAGTTCGN

TTCATTTTTC GGATTCGGCA CACCCATCGC GCGTNTCCGC TNGACGACAG ACCGACCAGA

CCGGTTNTAG GTTAGACATT ATTAATAAAT CGTNATAATA ATTAAATTTA ATGAATGGAA

AATTCTCTCT GATATAACGA AATGATTTAG TGATCGTGNT ACGCCGCGCC GAATCGAATA

TCGAATAGNA CGCGAGTTTA TNTTGTTGCG ATGCAGTTTA TCTTACGTCA ACGCGGCTTT

AGGTTTTTTT CGTGATAAGA CAATACAAAT TTATGTTAGG GTGTCACGTT GTAATAGTAT

TGTTATCGGC ATTGATAATG CGAAGGTCAA ACGGAGGCCA AGATGGCGAC GCGGCCGCAA

TCGCGACACA TATGCGAGAA CATTGTTCAT TTGCCATTGT TTAAACGGAG GGACTTCGGT

TCTGTTTTTT AAACCATAAC TTCAATGAAG CGATCTCATT TTACACTTCT CTCCATCGAA

GCACAGTTNA TCTATAATAC GTAGAAACAC TTCAGGTTTG GAATGAATTC

The insert in plasmid 2-18 (536 bp,from the TA site flanking the 3'IR to the EcoRI site in HmB, complete sequence)

TATACGGCCC GGGCGATCGT GGCTTGTAGT TAGTTAAAGA GCTCATATTA CAACATCAGT

ACTGTAATTT ACTTAAATAT ACGTGAGTTT AATAAAGTAT ACATAAAAGT GAGCGGAGTA

AGTGTTAATT ATGCGTGGTT TATTCTCGAA TTAACTTGTT TCTCCCAAGA GACATACAAC

ATTTGTTTAA TTGGTAATCG TAAAATGTTT AAAAGAAAAT TCAGCAAATC TTTTAATAAG

CTTCTTTTAG GGCTAAAACT ACTATCAGAC TAGTCACTAT CAGATTTACT TTTATGTTAT

AATATTGAAG TCTGAAATGC ATTGATTTAA TGGGTAATGC ACTTCCAATT GGAAAACATA

ATTGGTTCTC GGCATATTTA GGCAGATAAT ATTCNCAATA CGCCTTTTGA GTTATAATAA

AACTATNCCT NAATACCTAA TTCCGTCCCT AACTAAAACA CTACAAAATA AAAAGTTAAT

TCCATCGAAT CATACATCCG AATCGAATAG ATAACATTTG CCTCTAAAAC GAATTC

The insert in plasmid 2-24 (1082 bp, from 3'IR to EcoRI, complete sequence):

TATACTTCGG TAAGCTTGAT ATCGAATTCT GATGGAACTG ATGTTTTAGT AAAATTAGAC
GACCTTTGAG AATCAGGGTC GGTGACGCTA CGTTACTGTT CTAGAAATTA CCAACAATAA
CAACAACAAC ACATTAAAAA CAACAGACAG ACAACCAATG GAGTGGACGT ACGATTGGTC
AAAATTGATG ATAAGGATTG GACGCGAAAG GCAAGATACC GATAGGAATG AAAAAAAAAC
CTAAATCCGG CATTAGGCGG ACATATTTCG AGGAAAAGGT AGTAAAAACG CGTCAGCAAA
GTATATTTGG AGGAAAAGAT AATAAAAACG CGTCACCAAA CTAAGGAAAA CTACAAAAAA
TAACTGTAAA CAAAACTGAA GCGCAATCGA TATTGGACTG ACGTAATTCT GCCGGTTTCA
CAACAAACGT CTCGGGCAGA GGGATTTCAT TCACTAAATT TTCTTGTGTT TTATGAATAT
AATAAAAACT AGATGATGAC CGAACTTTGC TCGTTTTTTT TATTTGATAA CGCCATCTTG
TTGTGTCTTT AAAGCGGTTA GTTGCCCTCA ATTAAGAAAA ATAGTATTAT TATTCGCCAA
TAGATGTCGG GAAGAGTTGA TTATTGAAAA CACGAATAAA AGAACATTTT CTGAAAATAA
ATCGTGGCTA GATCGATTTA TCGCCCCCGA AATCCCATGT ATACTAAATT TTATGAAAAT
CGTTGGAGCC GTTTCCGAGA TTCAGATTAT ATATGTATTA ATATACATGA ATTGCTCGTT
TAAAGGTATA AGATGAATAG TGACATAAGT GCCAAATGTT TCTTCGGTTT TCCTGCAGGA
GCTTTGTACG GGAGGTCAGC TTGATCGGCT GCTTGTTAGC ATTGATATGA AATTACTAAT
AGCTTTACGT TAGATTTGTC TATTTCATAA ATGAATTGAT AATTCGCTGT GATGGATACG
TATGCTCGGT ATGCTCCACG ACTACTATTG ATCATAGATC CGCCCGTACC GAAGTATTAG
AAGTNTCAGC TTCAATCGAA TAACGGATAT CCTTTGCTTT TTGGCAGAAA TAGAATGAAT
TC

The insert in plasmid 5-5 (438 bp, from TA to EcoRI, complete sequence):

TATTTAAACC AAATAAAACT GAACCCGTAA CCCGTAACTT AAGCCCAGAT GGCGAAATAC
ATTGAATTAT AAACGAAAAT GGCTATTTTC TCCTTCAACA ATTCATTAGT TAAATCGTTT
CTATTTGGTA TTAAAGATTG ACAGTCATTT TTCATGTTAT TCTAACTCGT GAATTTTTAA
CAACAAAAAA ACTCCCATAA ATTATGCACA ACGAAGTTGG GATAAATCTT TTTGTTATTC
TTATCTTTGA TATCGAGGTA AACAAACTGA ATTATGACGT NAAAAATAAA AACATATNTA
ATACCCAGCC CATATGTATG TAAATTCGGN AGAATTTTTA TTCCTTATCT AATACAGAAC
ATTTNGTGGC TTATGAATAA ACATAAGANT TATTGGTTCC GACGTATTTT ACGCNTTCCT
AGAAGGNTAC ATGAATTC

The insert in plasmid D13 (~2.5 kb, from TA to EcoRI site):

TAGTGTTCCA GTACCTTGAG ACACGATGTC AAAATCTGAA ATATATTCTT CCCCCATTCT
CCAGTCAAAT TAGAGTGCAA GAGTAGTTAG TGGAAAAAAT AGTCAAAACA CTGGGATACC
AGTGAATATC AATAGCTCCT ATACAAAAAT ATGTCAGAGA CAATTTCTCA ATTTCCCTTA
TTTATCCAAA TCGCTCTCAT AGAAGGGCAT TTCGTGTATT CAGCAAATAA ACTGGGCATA
GAGCATCTAT TACTTTTTAG AACATGTCAA TAATGCTCAT GGTCCACATG AAAGGAACAA
TTTTTTCCGA CCTATTCTAG TAGCCTCTAG GGATTCTACT AGATTCCCAG AATTAGGAAG
AGCACAGTTT AAATTTTTGG GGCNTGACGT TTAGANTGAT GTTNATGGAA GGGTTTGAAT

CAAGCTCTAC NCCCNGCAAA GCAATTCCNC CCNCCCCCCA ACTCTGCGCG TTTTCCCACC

ATTAGCGNTN TTATTATCNT GAATAACCCC ANCCGNTTNT NTCCCCAACC GCTGANTAAA

NCCTTGAAAA NAATGGTTTN TTCNAANCCN ANAGGGNGGA NAANTTATTC CCNAAAACNG

ACTNTGCCNN CCCCCCNGCC GGGGTTCCNT TTGCAAACTG CCTNAACCAA TGCCCCTNTT

CTGCNCCGGN CCNGAGGCAN G...~1.2kb unsequenced segment .....TCCCAGTCAA

ATTAGAGTGC AAGAGTAGTT AGTGGAAAAA ATAGTCAAAA CACTGGGATA CCAGTGAATA

TCAATAGCTC CTATACAAAA ATATGTCAGA GACAATTTCT CAATTTCCCT TATTTATCCA

AATCGCTCTC ATAGAAGGGC ATTTCGTGTA TTCAGCAAAT AAACTGGGCA TAGAGCATNT

ATTACTTTTT AGAACATGTC AATAATGCTC ATGGTCCACA TGAAAGGAAC AATTTTTTCC

GACCTATTCT AGTAGCCTNT AGGGATTCTA CTAGATTCAC AGAATTAGGA GGAGCACAGT

TTAAATTTTT GGGCATGACG TTAAGACTGA TGTAAATGGA AGGGTTTGAA TCAAGCTNTA

CACCCCAGCA AAGCAATTCC ACCCTCCCTC CAAGCTCTGC GCGTGTTCCA GCAATTAGCG

TATTATTATC ATGATTAACC CTAACCAGTT TATATCCTCA AACCGCTGAC TAGACCCTTT

GAAGATGAAT AGGTATAGTT CGAATGCACA GATAAGGGTG GATGATGTGA ATTC

The insert in plasmid A9 (~2.8 kb, from 3'IR to EcoRI, incomplete sequence):

TAACTTAGTT ACNTACTCTT CTGGCTTGTG AATTATTTTA TTGCTTAAGT GAATTTTCCA

TCTTAAGTTG AACTTGGTTT CGAAATGAGA TCTCTAAGTT AGTTCCCAGT AGCTTATCGT

TAGTTTAAAG TTTGCAAAGA TTATTTACTA TTGTAGCTTA CGAGCGTTTA CTTTATTTTT

ATTTTTTTAT AGCTGACGAA CCGATTTACG GATGGGTATC CGTAACCGTN TGGGTANCAG

ACCGGATTAT ATCGGACTCC GGCTACGAGC TATACCTAGG ACTCAGTACG CCACTCGATT

TTCCCTCAAC CCACTTTCAA TGTAGCCAAA TCAGCACAGA CGCTGTGTAA AAAGGACATT

GACCTTATGT AAAATAANAT AAAGCCCANA TTTTTTCACA TTTTTTCCCT TCTACTGTAT

TCAATGCACT GCAAAGAANT CNCTCNCNTA ATTTTTTTAG GTNCCTGANG CCATCGGTCA

TCGTTCTCGC CGGGATCTTC TCAGTGGGTC GCGTTTCCGA TCCGGTGGTA GATTCCTGCG

AAGCACTGCT CTTGCTANGG CCAGTGTTAT CTACGTCCGT CAGGTTTGAA CCCCGTGAGC

TTACCTACTA GTTAAGGTTA CGCTGAAATA TCCTGTCAAG GCTATCAGCT TAGGTTGGGA

AAAATAAACT AAAGAAGTCC ACGAAAAAAC CTGGGTAGTA TATCAATAAT TCTGCTTTAT

CCAAATCANA ATTGTCTATC AAACTCTTAA TGTNTCATCA CATGCAATAT TCNACGAACA

GTATATNGGT CATCGGCGGA NGCCATAACC GTCNTANCAC CCAATTTATT GAACCACCAT

AACTAAAGTN CATCCGCGAT ACNCNTCTNA TTTATGGGTA ACATTAGGGG TNTTTCCCAA

TTAAAATTAT TGCATATTCC NTTTGGATNT NCTCNCCATA GTNANGAAGG AAATTTATNA

AAACCCNCGG GATTTTTTTT TGAATCCTCC TATATTTTAT GAATCCNTAA ANTNACCAAA

NTGCTCCAAA AATGATGAAT CCNTTCNACT TGCTATTTTT TTATANGCNT TTCTTTGGGG

ANCCNGTTAA TCCAATTTCT TGATCATTTA TAAAAAAAAA AAAACCGTTT TGAAATTCTG

CNAATNTTCT TATAATTCCC TTTGAACCTT GTTTTTAACC TTNTTCCAAT ATTCNTTCCT

GTTTTTTATC CNTCCCNNAA AAAAT..... ~0.8 kb unsequenced fragment .....

TCGGTAACAG TATAAACCCA GGACTTAGAC CTCATGTCTA GGTAGTCATT TAGCTAACAT

CCAGTATGCT GTGTGATCAT TCCGTGCAAT AAAAAGGGCT AGAGTTAACC AGGAAGGTTA

ATGAAAGATA TTAATTCTAC AAATGGAATA CTGCGCTGTG CTATTCATAG GAACGTAGGC

TAGGACTATA CATACTCTAC GAAGCCACAG TCACGTATCT TCTGAGTGAT AGTTGAATAT

GTTATAATTA TTAGATATTT GAAGCGATTA GAGGATACTT ATTGAATTAT TTACCTAAAC

TGGCCGATAA AGTCGTCGCT CATTCAGAGG AAATAGATCC TCTCAGTCAT ATGTAAAGCT

TAACATAGTC CGCAAAAGCA CAAATCCACA GTCGACGAAA CGTAAATATA GATTTAATTG

TGCTGTATTC TTTAAACATA GTGTTTTTAG AAATGATTTT AGCTTGACAA TGTTACGGAA

CTTCAAAAAA TAATAATAAT AAATCGGAAA CAAATAGGTT GACAGTTGAT GTACATAAAT

ACATATTTAT TGTGATATCT CATAGGATAT TTTGGTTGAA GTCTGATGTA ATAGAAGTGT

CAATCACGTG CCACCCTCTC TTATTAGATC CGTGATATTC GAATCAGCAT AAAAGAAAAA

AGAAAAGTGC GGCTCTTTTA TTACGTTTGA AAGAAATATG TCGAATATGT TTTATCTTCC

TTAACGCCGA ACTTAGCTAA ATGCGAATTC

The insert in plasmid 1-2 (385 bp, from TA flanking 3'IR to EcoRI at HmB):

TAAAACTCCG CTTCCAGCAG TCAGGAAATA ATACACAGAT TGCAGTAGTT CACTTAATTC

TCTATTGCTT TATTTGTAAC TCTGCTTGTA CAAATCTTTA ACTGAATACA CTGCTTTCTG

TCCCGTTTTT CCANACTACC GCCTTTTTTA TTATGTCCCG TCCCCACTAC TGCCATTGCT

CCTACCAACA GTGTTGCTAT TTGGCTATTC GAAAATTGCT TTAACTTTAA TATGATAAAT

TTTATTGTTA TTATTTTACA ACAAACCTTA CCCTACTAAT AGTGGCCTTG TGACTAAGCC

TGACGTTGAT AATACTAGCC CTAGCTAGAA GCACTGGCNT CGGTGTTNTC TACCACCNGG

ATCGTANTTC GCGAACCTCG AATTC

Appendix 3. PCR Detection of I-Ppo initiated integration

Principle: The I-Ppo endonuclease was shown to be capable of producing DSBs in *Bombyx* 28S rDNA. Once the I-Ppo donor vector integrates into the I-Ppo cutting site in 28S rDNA via gene conversion, a PCR amplification can be employed to detect the integration events. A pair of primers is designed for the detection of the boundary region of the integration site which includes one side of the I-Ppo recognition sequence in 28S rDNA gene followed immediately by the I-Ppo donor vector sequence. Thus, only genomic DNA carrying the integrated vector DNA specifically at I-Ppo site will yield amplification products, while neither extrochromosomal vector nor genomic DNA alone can yield the PCR products.

Integration Map: The map shows the pICN vector integration site in *Bombyx* 28S rDNA. The 5' and 3' primer are derived from 28S rDNA and the pICN vector sequence, respectively. The slanted regions are identical to the two flanking regions in the I-Ppo donor vectors. Shaded box: actin promoter and terminator; Solid arrow: *cat* reporter gene and *neo* selection gene in the vector; open box: the 28S rDNA gene. This drawing is not in scale.

Integration of the I-Ppo vector in 28S rDNA