ANALYSIS OF SEQUENCES CONTROLLING EXPRESSION OF THE *I* FACTOR OF *Drosophila melanogaster*

APINUNT UDOMKIT

PhD THESIS

INSTITUTE OF CELL AND MOLECULAR BIOLOGY
UNIVERSITY OF EDINBURGH

1996

I dedicate this thesis to my parents

and all my teachers.

Acknowledgements

There are so many people I am in their debt over the period of my study at the University of Edinburgh.

Firstly, I would like to express my deepest gratitude to my supervisor, Prof. David Finnegan, for his advice, encouragement and enduring patience. Without him this thesis would not be possible.

I would like to thank Prof. Mary Bownes for acting as my second supervisor and for her help and advice. Thanks also go to the people in her group for allowing me to use some facilities in their lab.

Many thanks to Carol and Gerry who gave me valuable help and suggestion during my first year. I also would like to extend my thanks to everybody in my group for their help and hospitality. These people are Glen, Susan, Angela, Eve, Maki, Petra, Ruth, Hongmei, Jill, Laura and Ivan. Special thanks to Susan for allowing me to mention her unpublished results in this thesis and to use her *Drosophila* strains in β-galactosidase activity assays. I wish to thank Lynn Powell for her advice on gel retardation and DNA footprinting techniques.

Thanks to Graham Brown and Frank Johnston for their help in photography work. I would like to thank Kathleen who kindly provided *Drosophila* stocks whenever I needed. Thanks to Helen who kept my work going by supplying clean glassware and to Frances for providing fly food during the years.

I would like to express my overwhelming gratitude to Mrs.Cousland for her kindness, encouragement and for giving me a wonderful opportunity to learn many things about Scotland, mostly during the afternoon tea every Sunday, over the years. Special thanks to Bumroong Puangkird for his friendship and for his help and suggestion on

computer work. I am very grateful to Komson Pirapatrungsuriya whose understanding and sincere friendship always encouraged and cheered me up. I wish to extend my appreciation to all Thai students in Edinburgh for their help, friendships and for the enjoyable badminton games we had together.

I would like to thank the DPST scheme (Thailand) and the officers of the Thai government Students' Office in London for giving the financial support and for their help throughout my study.

Finally, I owe tremendous debt to my father, my mother and my sisters for their eternal love, support and understanding. No words can express my gratitude for everything they have been giving me all my life.

ABBREVIATIONS

A	adenosine
ATP	adenosine 5'-triphosphate
bp	base pair
BSA	bovine serum albumin
C	cytosine
°C	degree Celsius
CAT	chloramphenicol acetyltransferase
Ci	curie
cm	centimetre
d	deoxy
dd	dideoxy
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxyribonucleoside triphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
g	gram
G	guanosine
IAA	isoamyl alcohol
kb	kilobase
kDa	kilodalton
LTR	long terminal repeat
M	molar
µg	microgram
µl	microlitre
µM	micromolar
mg	milligram
ml	millilitre

mM	millimolar
min	minute
mRNA	messenger ribonucleic acid
ng	nanogram
OD	optical density
ONPG	O-Nitrophenyl-β-galactopyranoside
Pu	purine
Py	pyrimidine
rpm	revolution per minute
PCR	polymerase chain reaction
RNase	ribonuclease
SDS	sodium dodecyl sulphate
SV40	simian virus 40
T	thymine
TAF	TBP associated factor
TBP	TATA binding protein
TdT	terminal deoxynucleotydyl transferase
TEMED	N,N,N'-tetramethylethylenediamine
TFIIA-H	transcription factor IIA-H
TLC	thin layer chromatography
Tris	tris (hydroxymethyl) aminomethane
U	uracil
u	unit
UTR	untranslated region
UV	ultra violet
V	volt
v/v	volume per volume

ANALYSIS OF SEQUENCES CONTROLLING EXPRESSION OF THE *I* FACTOR OF *Drosophila melanogaster*

Abstract

The *I* factor is a LINE-like transposable element in the genome of *Drosophila melanogaster*. Transposition of the *I* factor occurs by reverse transcription of an RNA intermediate and is activated in the germline of female progeny of crosses between males of an inducer strain that contains complete elements and females of a strain that does not, reactive strain, resulting in a phenomenon called I-R hybrid dysgenesis. Previous data have shown that the promoter of the *I* factor lies between nucleotides 1 and 30. Expression from this promoter is regulated by sequences between nucleotides 41 and 186 that function as an enhancer.

The work present in this thesis has identified within this region a sequence called site 1, encompassing nucleotides 139-157, that is recognised by a sequence specific binding protein. This binding protein is present in ovaries and non-ovarian tissues of both inducer and reactive strains. Site 1 is required for basal promoter activity as well as the enhanced expression in the ovaries of reactive females. We suggest that the level of expression from the *I* factor promoter is dependent on the ratio of site 1 and its sequence specific binding factor or other proteins that interact with it.

TABLE OF CONTENTS

Chapter 2: Materials and methods

CHAPTER 1

Introduction

Transposable elements are discrete sequences in the genome that are able to move themselves to other locations within the genome. In eukaryotes, transposable DNA sequences occur in families of repeated element that may make up 10-15% of the genome (Finnegan 1990). They can be classified according to their presumed mechanism of transposition as elements that transpose directly from DNA to DNA and elements that are related to retroviruses, called retrotransposons, transposing by reverse transcription of an RNA intermediate (Finnegan 1989a). Retrotransposons are of two types, the elements that resemble retroviruses in having long terminal repeats (LTR or retrovirus-like retrotransposons) and the elements that contain no terminal repeats and have A-rich sequences at the 3' end of their coding strands (non-LTR retrotransposons). Non-LTR retrotransposons are often referred as LINE-like elements because the first examples of this type detected are mammalian LINEs or L1 elements.

Transposition rates of transposable elements are difficult to determine as the elements normally move infrequently. For instance, transposable elements in *Drosophila melanogaster* are thought to transpose at a rate of 10^{-4} to 10^{-5} per generation (Charlesworth and lapid 1989). However, there are some particular conditions under which transposable elements are activated and move at high frequencies. One of these is known as hybrid dysgenesis in *Drosophila melanogaster* where transposition of transposable elements induces genetic abnormalities in hybrids between certain strains (Kidwell, *et al.* 1977)

1.1 Hybrid dysgenesis

Hybrid dysgenesis is the occurrence of abnormal characteristics in the progeny of crosses between particular strains of *Drosophila melanogaster* (Kidwell, *et al.* 1977).

Abnormalities produced during hybrid dysgenesis, known as dysgenic traits, include sterility, mutation, and chromosomal aberration. So far, three independent dysgenic systems have been identified. Strains of *Drosophila melanogaster* can be categorised into two types with respect to each system. Inducer (I) and reactive (R) strains are responsible for I-R hybrid dysgenesis (Picard, *et al.* 1978), Paternal (P) and Maternal (M) strains contribute to P-M hybrid dysgenesis (Kidwell, *et al.* 1977) and H and E strains lead to H-E dysgenic system (Blackman, *et al.* 1987, Yannopolous, *et al.* 1987).

1.2 I-R hybrid dysgenesis

I-R hybrid dysgenesis is produced only in the germ-line of female progeny from crosses between males of an inducer strain and females of a reactive strain. The female progeny from an I-R dysgenic cross are known as SF females and are partially or, in some cases, completely sterile as they lay a normal number of fertilised eggs only a proportion of which hatch (Picard, *et al.* 1977). Female progeny from the reciprocal cross are fully fertile and are named RSF females. The progeny of crosses between inducer or reactive strains are also fertile (Picard 1976).

1.2.1 Dysgenic traits induced in I-R hybrid dysgenesis

Sterility

SF females of I-R dysgenic crosses have apparently normal ovaries. The failure of certain percentages of the eggs they lay to hatch is due to developmental arrest during early cleavage divisions (Lavige 1986, Picard, *et al.* 1977). Bregliano *et al.* (1980) suggested a threshold effect resulting from some specific biochemical deficiency in the oocyte as a biological cause of embryo death, though there has not yet been direct evidence clearly demonstrated so far.

SF sterility is measured by the proportion of the hatched eggs to the total numbers of eggs laid by SF females. This value is called 'hatchability' and is independent of the males with which SF females mate but can vary from cross to cross depending on the inducer and reactive parents (Bucheton, *et al.* 1976). Inducer or reactive strains that produce SF females with low hatchability are called strong strains. On the other hand, the strains that give more fertile SF females belong to weak strains. The hatching percentage of the eggs increases with the age of SF females and can reach normal level in 15-20 day old adults. Another factor that also has an effect on hatchability, albeit reversibly, is the temperature. Hatchability varies depending on the stage of oogenesis at which the heat treatment is applied (Bucheton 1978). More details of these effects will be discussed in section 1.2.2.2.

Chromosomal aberration and Mutations

The progeny of SF females that are able to go through development and survive display some genetic abnormalities including mutation and chromosome non-disjunction. The frequencies of these genetic traits are correlated with the level of SF sterility (Picard, *et al.* 1978). High frequency of X chromosome non-disjunction was observed in SF females bearing strongly reduced fertility but not in RSF females nor in inducer and reactive females. The frequency of chromosome non-disjunction decreased in older SF females. Similarly, high rates of mutation were observed in SF females while mutations were significantly less frequent in some RSF females. These mutations were suggested to be the results of small deletions or insertions rather than point mutations. A similar correlation between mutations induced in I-R hybrid dysgenesis and SF sterility has also been reported by Proust and Prudhommeau, 1982.

1.2.2 Inducer and reactive determinants

1.2.2.1 Inducer state

The inducer state in *Drosophila melanogaster* is controlled by genetic determinant called *I* factors. Any of the four chromosomes of *Drosophila melanogaster* may contain *I* factors in which case their chromosomes are said to be i^+. *I* factors are stable and are inherited in a Mendelian fashion in inducer strains as well as in male progeny of crosses between inducer and reactive strains (Picard 1976, Picard and Pélisson 1979). However, *I* factors conduct themselves differently once they are introduced into reactive environment of SF females by I-R dysgenic crosses. They are unstable and can transpose from chromosomes of the inducer parent to either homologous or nonhomologous chromosomes of the reactive parent (Picard 1976).

The process by which reactive chromosomes acquire *I* factors is called chromosome contamination and can occur at the frequency as high as 100% per generation. After chromosome contamination the contaminated reactive chromosomes, called *rc+*, behave like i^+ chromosomes, that is they result in sterile females when they are paternally transmitted to a reactive oocyte and they are able to induce chromosome contamination in females heterozygous for rc^+ and reactive chromosomes (Picard 1979).

Not all chromosomes of inducer strains have i^+ characteristics . Some lack inducer potential and are noticed as i^0 chromosomes (Picard and Pélisson 1979). It has been demonstrated that crosses between males from the strains with a genome comprising entirely of i^0 chromosomes and reactive females are unable to produce SF females (Pélisson and Bregliano 1981). The i^0 chromosomes are, nevertheless, not reactive since i^0 females produce fertile daughters when they are crossed to inducer males. A spontaneous change from i^0 to i^+ chromosomes can occur at remarkably high frequencies especially when present in a reactive cytoplasm. This phenomenon

suggests that i^0 chromosomes carry *I* factors that are silent due to some kinds of repression. However, no direct evidence has been produced to support this assumption.

1.2.2.2 Reactivity

Reactivity is a cytoplasmic condition that, during I-R hybrid dysgenesis, allows transposition of *I* factors to take place. A difference in the levels of reactivity can be seen from one reactive strain to another. The strength of reactivity is defined by the reduction in fertility of SF daughters produced by crosses between reactive females and standard inducer males. Strong-reactive females produce SF daughters that lay eggs with lower hatchability than those of SF females from a weak reactive strain when crossed to males from the same inducer stock (Bucheton, *et al.* 1976).

Hereditary transmission of reactivity

Bucheton and Picard (1978) showed that a reduction of fertility in SF females did not depend on the reactive chromosomes acquired from their mothers but appeared to be determined by a cytoplasmic factor in the oocyte of the reactive parent. The influence of individual paternal chromosome, chromosomes X, 2 and 3, on the level of reactivity has been tested. Females progeny coming from strongly reactive mothers and fathers homozygous for strong or weak reactive chromosomes were crossed with inducer males and their female progeny were tested for fertility. Most female progeny were infertile as they were descended from a strong-reactive maternal line. Variation in the degrees of fertility among these females was observed and this could be correlated with the presence of weak reactive chromosomes that their mothers inherited paternally. A reduction in reactivity was noticed in an equivalent experiment where the genotype of a strong reactive strain was replaced by weak-reactive chromosomes.

Therefore, reactivity is maternally inherited through a cytoplasmic factor but its level can be influenced, though at lower degree, by nuclear genome of the fathers. The action of paternal chromosomes on the level of reactivity becomes greater over generations and eventually overcomes that of the maternal origin. This process takes at least 10 generations (Bucheton and Picard 1978).

Effects of aging and thermic treatment on reactivity

In addition to the influence of genetic factors on reactivity, its level can also be modified by non-genetic factors such as aging and temperature.

Bucheton (1978) followed the level of reactivity as reactive females aged. SF females from young mothers showed high levels of sterility with the hatching percentage not higher than 20%. Older reactive mothers produced more fertile SF daughters. At the age of about two weeks, the reactivity of these females rose up to the level similar to that of RSF females showing 90 to 95% of hatchability. In contrast, age of inducer males did not have any significant effect on the sterility of SF females.

A change in temperature also induces an alteration in the level of reactivity. When SF females normally maintained at 20°C were subjected to higher temperature at 29°C during the 5^{th} and the 9^{th} days after eclosion their sterility decreased within three days after the heat treatment. The hatching percentages during this period was higher than those of SF females of the same ages continuously maintained at 20°C . This consequence was reversible as the sterility increased again after the flies were returned to 20°C (Bucheton 1978). The action of temperature on reactivity was found to depend on the stage at which an increase in temperature was applied. SF sterility decreased only when the heat treatment occurred during late oogenesis for instance 5-9 days after the flies emerged. On the other hand, an increase in sterility was observed when SF females were submitted to heat treatment at all other stages of development (Bucheton 1979).

The effect of aging and temperature on reactivity are partially inheritable (Bucheton 1978). The fertility of young SF females gradually increased when they were raised from older reactive mothers though to a lower degree than those observed in aging SF females themselves. A slight difference of sterility was also detected in SF females descended from old grandparents. The sterility of SF females from reactive ancestors that were subjected to heat treatment was affected in a similar manner.

The effects of these two non-genetic factors are cumulative and reversible as they were both able to modify the level of reactivity when their actions were given to flies at each generation. The flies could acquire the initial level of reactivity again when they were returned to a standard breeding condition (Bucheton 1979).

1.3 *I* factor

1.3.1 Isolation of the *I* factor

Mutations arose in SF females were argued to be the consequence of *I* factor insertions as two independent mutations in the *white* locus, w^{IR1} and w^{IR3}, induced by I-R dysgenic crosses were shown to be genetically linked to *I* factor activity. The mutations could not be separated from *I* factor by recombination (Pélisson 1981). This observation suggests that w^{IR1} and w^{IR3} mutations were the result of insertion of an *I* factor into the *white* gene.

Soon after, the structure of the *white* locus on chromosomes carrying eight *white* mutations, w^{IR1}-w^{IR8}, were determined (Bucheton, *et al.* 1984, Sang, *et al.* 1984). Two of them are due to deletions in mRNA coding sequences. The other six are the consequences of individual insertion of indistinguishable 5.4 kb elements. Three of these insertions were present at the same site within the last intron of the *white* gene that appeared to be a hot-spot for insertion.

The 5.4 kb element that was found inserted in *white* mutations is likely to be a functional *I* factor. In 1984, Bucheton *et al.* have successfully cloned *I* factor sequences present in w^{IR1} and w^{IR3}. One of these clones, pI104, contains full length *I* factor sequences from w^{IR3} DNA.

1.3.2 Molecular structure of the *I* factor

A complete sequence of the *I* factor associated with w^{IR1} has been determined (Fawcett, *et al.* 1986). It is 5,371 bp long and is flanked by direct repeats of 12 bp target genomic sequence. There are four copies of the sequence TAA at the 3' end of one strand. Neither direct nor inverted repeats was found in the sequences at both ends of the *I* factor. Nevertheless, the sequences at the termini of the *I* factors inserted into the *white* gene on chromosomes carrying w^{IR2}-w^{IR6} mutations were determined and found to be highly conserved. The only apparent differences are the third nucleotide at the left-hand end that can be either G or T and the variation in the number of TAA repeats at the right-hand end. These elements are flanked by target sequence duplications varying in length from 10 to 14 bp.

The *I* factors contain two long open reading frames, ORF1 and ORF2, that are located 53 bp apart. The first ORF is 1,287 bp long. The predicted polypeptide encoded by this ORF contains one complete and one partial copy of a highly conserved motif $CX_2CX_4HX_4C$ found in basic nucleic acid binding regions of retroviral *gag* polypeptides, *gag*-like polypeptide of transposable element *copia* in *Drosophila melanogaster* and the coat protein of cauliflower mosaic virus (Covey 1986, Fawcett, *et al.* 1986).

The presence of nucleic acid binding domains in ORF1 polypeptide indicates that this protein may play a role in regulation of *I* factor transposition by interacting with the regulatory sequence on *I* factor DNA itself. However, no specific nucleic acid binding

activity of ORF1 polypeptide of the *I* factor has been detected so far (A. Dawson, personal communication).

The second ORF of the *I* factor is 3,675 bp in size and encodes a polypeptide that contains sequences corresponding to the conserved domains of reverse transcriptases (Fawcett, *et al.* 1986). At least three activities have been found to be contributed by *pol* gene products. These activities are RNA-dependent DNA polymerase, RNase H and endonuclease (Simonelig, *et al.* 1987, Temin and Mizutani 1970, Verma 1975). The conserved domains in ORF2 of the *I* factor show similarity to those of RNA-dependent DNA polymerase and RNase H activities of reverse transcriptases. No endonuclease-related domain has been detected so far (Finnegan 1988).

The structure of the *I* factor (figure 1.1) indicates that it belongs to a family of transposable elements called LINE-like elements. LINEs (long interspersed nuclear elements) are repeated elements found in a wide range of mammals (Singer 1982) and in other eukaryotic species (Hutchison, *et al.* 1989). These elements will be discussed in details in section 1.8.

1.3.3 A cloned *I* factor as a determinant of I-R hybrid dysgenesis

Pritchard *et al.* (1988) have provided evidence confirming that the 5.4 kb *I* factor associated with the w^{IR3} mutation is fully functional. The cloned w^{IR3}-*I* factor (Bucheton, *et al.* 1984) was introduced into reactive flies by *P*-element mediated transformation. The transformed flies carrying an incoming *I* factor were said to be analogous to the progeny of an I-R dysgenic cross providing a condition that allowed the *I* factor to transpose. *I* factors were detected at different sites on the chromosome arms of individual transformed line apart from the original copy from the vector. The presence of these copies suggests that transposition of the *I* factor from the site at which it integrated had occurred in the transformed flies.

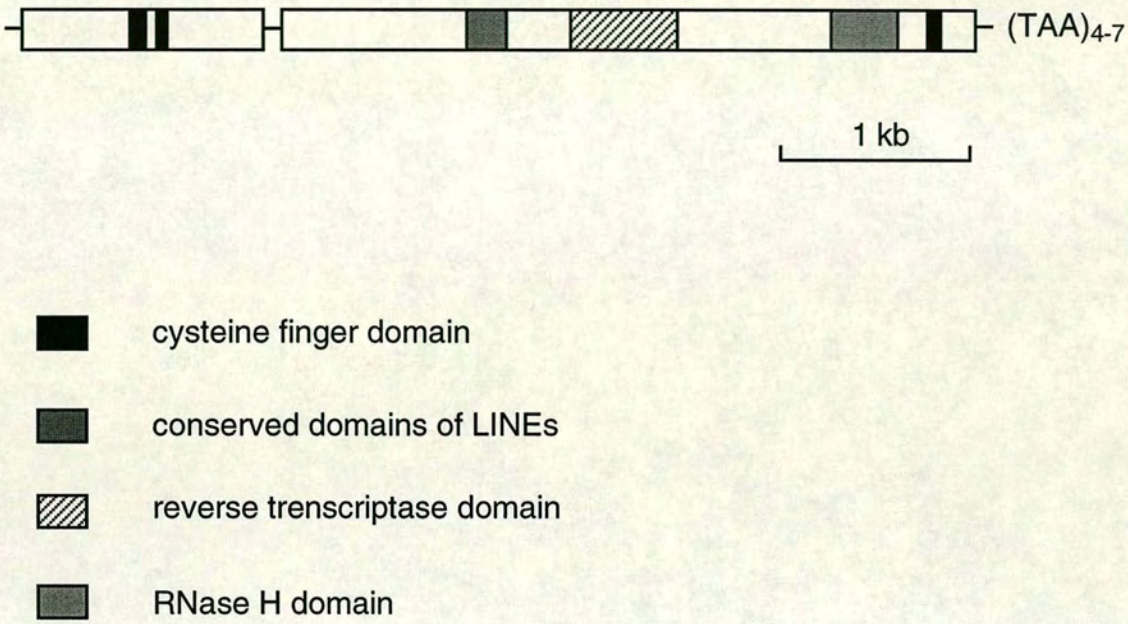


Figure 1.1 The structural features of the I factor: untranslated regions are shown as a single line.

The transformed lines in this case should become inducer as a result of an increasing number of complete *I* factors. This has been confirmed by the ability of transformed males to induce SF sterility when crossed with reactive females.

A transposed copy of *I* factors was cloned from one of the transformed lines. The sequences at its ends were compare to that of the w^{IR3}-*I* factor. The new copy was flanked by a duplication of an 11 bp target sequence that differed from the flanking sequence of w^{IR3}-*I* factor. The element contained an extra copy of TAA at its right hand end while the left hand ends of the two elements were identical. An additional copy of TAA must have been acquired during the transposition presumably by polymerase slippage during replication as has also been reported by Jensen and Heidmann (1991).

Another interesting point reported by Pritchard *et al.* (1988) is an inability of the *I* factor to transpose on its own from the injected plasmid without *P*-mediated transformation sequences. This could be explained by late transposition of the *I* factors. As mutations associated with I-R hybrid dysgenesis mostly appear as single events (Picard, *et al.* 1978), it is most likely that *I* factors transpose just shortly before meiosis. The injected DNA may have been degraded or too dilute by that time so that the frequency of successful transformantion becomes very low. In contrast, transposition of *P* elements starts earlier probably in late embryogenesis (Engels 1988).

1.3.4 Defective *I* elements

Sequences homologous to *I* factors have been found in the genomic DNA of both inducer and reactive strains (Bucheton, *et al.* 1984). The two strains share sequences of *I* elements that are concentrated in the pericentromeric heterochromatin. These elements must be non-functional as they are present in reactive strain as well as in inducer strain. In addition to these common defective elements, inducer strains also

contain copies of *I* factors on the chromosome arms that are not present in reactive strains. The number and location of these euchromatic elements vary from one inducer strain to another suggesting that the elements include active *I* factors.

DNA probes containing the left-hand and the right-hand ends of the *I* factor were used to screen *Drosophila* genomic libraries of both inducer and reactive strains. Several clones from both strains were chosen and their structures determined by restriction mapping (Crozatier, *et al.* 1988).

I elements in the clones obtained from reactive strain were found inserted into repeated sequences that are themselves inactive transposable elements (Crozatier, *et al.* 1988, Vaury, *et al.* 1989). This is not surprising as the *I* elements in reactive strain are located in β-heterochromatic regions that are composed virtually entirely of repetitive sequences (Miklos, *et al.* 1988).

Restriction map analysis shows that these elements contain deletions, rearrangements or substitutions at different places throughout their length. None of the elements sequenced at their both ends is flanked by a target site duplication. The elements with an intact 3' end have a distinctive pattern of an A-rich sequence made up of $TAA(TAAA)_n$ rather than TAA repeats at the 3' end of functional *I* factor (Crozatier, *et al.* 1988).

Sequence comparison between the 5' end of four defective *I* elements and the 5' end of active *I* factor shows an average homology of 94 per cent. There are several deletions and insertions that disrupt the ORF1 of the *I* factors and many base pair substitutions leading to amino acid changes in this regions (Vaury, *et al.* 1990).

Among the clones obtained from inducer strain, three contain single copy of *I* element flanked by non-repeated sequences that mapped to different sites on the chromosome arms. Restriction map and sequence analysis showed that two of these elements

appeared to have a structure similar to that of functional *I* factors. The other element was truncated at the 5' end. The sequence at its 3' end was identical to that of active *I* factors with four TAA repeats. The element starts at position 2,316 and has lost ORF1 and part of ORF2. The truncation at the 5' end seems to be a common occurrence as it has been found in a large proportion of euchromatic *I* sequences (Busseau, *et al.* 1989, Crozatier, *et al.* 1988, Jensen and Heidmann 1991).

1.4 Mechanism of *I* factor transposition

As mutations induced by *I* factor transposition during I-R hybrid dysgenesis never revert and there is no evidence of i^+ chromosomes becoming i^0 chromosomes (Pélisson 1981, Picard 1979), transposition of the *I* factor probably occurs in a replicative manner rather than by an excision process. The replication of a new copy should occur by reverse transcription of an RNA intermediate as suggested by the presence of reverse transcriptase-like sequence in ORF2 of the *I* factor. This is also supported by the experiments showing that an intron inserted into the *I* factor sequence was removed during transposition so that it was not present in the transposed copy of the *I* factor (Jensen and Heidmann 1991, Pélisson, *et al.* 1991). Pritchard *et al.* (1988) have cloned and characterised a transposed copy of the *I* factor obtained by introducting the 5.4 kb *I* factor associated with w^{IR3} mutation into the genome of a reactive strain. This copy has an expected size for a full length *I* factor and the sequences at both ends are identical to those of the original copy except for an extra TAA repeat at the 3' end of the transposed copy. This evidence suggests that the RNA intermediate for transposition of the *I* factor contains the entire sequence of a donor element. The promoter for the synthesis of this RNA has been proposed to be carried within the *I* factor itself as the *I* factor associated with w^{IR1} mutation can transpose to new sites (Pélisson 1981). The *white* gene promoter cannot be involve in this process as the *I* factor was inserted in the opposite direction.

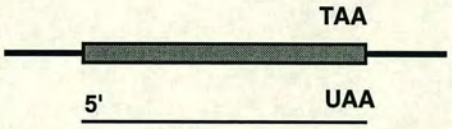
It is still unclear how the reverse transcription is primed, or how *I* factors integrate at new sites. A mechanism proposed for transposition of Cin4 elements of *Zea mays* that have similar structural organisation to *I* factors may be applied to *I* factor transposition. The 3' end of Cin4 transcripts are thought to associate with a 5' end of chromosomal DNA at a site of staggered single strand breaks so that the 3' end of the opposite strands can act as primer for reverse transcription (Schwarz-Sommer, *et al.* 1987). Since ORF2 of the *I* factor appears not to encode endonuclease activity, it is likely that *I* element integration takes place at spontaneous staggered nicks in chromosomal DNA. A possible mechanism for *I* factor transposition based on those proposed by Pritchard *et al.* (1988) and Bucheton (1990) is summarised in figure 1.2. The reverse transcriptase encoded by *I* factors presumably binds specifically to the UAA repeats of the full-length *I* factor RNA. This complex could then associate with the 3' overhang single stranded DNA of staggered nicks in chromosomal DNA. The mechanism also provides evidence for the generation of 5' truncated elements. This could occur by premature termination of reverse transcription as illustrated in figure 1.3.

1.5 A potential RNA intermediate for *I* factor transposition

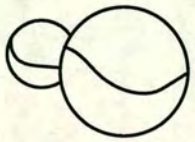
A 5.4 kb *I* factor transcript has been identified (Chaboissier, *et al.* 1990). Northern blot analysis showed that this transcript was only detected in the ovaries of SF females. Very small amounts of the transcript, however, could be detected in the ovaries of RSF females after longer exposure time. The presence of the 5.4 kb RNA therefore correlates with the frequency of *I* factor transposition that occurs five times less often in the ovaries of RSF females than in the ovaries of SF females (Picard 1976). The abundance of the RNA also varies with the age of dysgenic flies in a manner that correlates with SF sterility and *I* factor transposition. The 5.4 kb RNA is presumably synthesised from complete *I* factors and probably represents full-length RNA intermediate for *I* factor transposition. The presence of this RNA only under the conditions that permit *I* factor transposition suggests that *I* factor transposition is regulated at the level of transcription.

Figure 1.2 A proposed model for *I* factor transposition

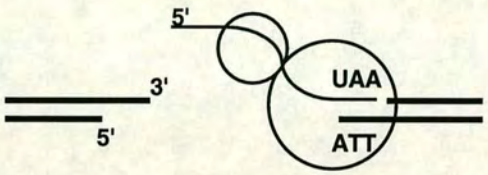
The full-length *I* factors are shown as shaded boxes. The thin lines represent *I* factor RNA and the thick lines represent chromosomal DNA. The products of *I* factor ORFs are shown as circles.



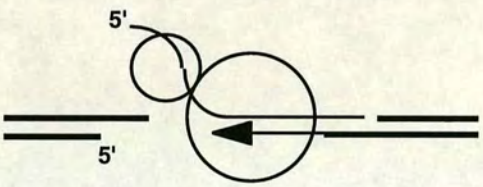
1. A full-length *I* factor RNA is transcribed from an internal promoter.



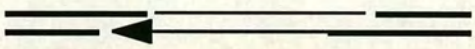
2. The products of ORF1 and/or ORF2 bind the RNA from which they are transcribed.



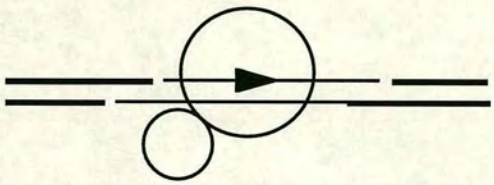
3. Nucleoprotein complex associates with the 3' overhang of staggered nick in chromosomal DNA presumably by homology between the UAA repeats of the *I* factor RNA and the sequence at the nick.



4. The 3' end of the protuding DNA strand is used as a primer for synthesis of the first *I* factor strand by reverse transcriptase activity of ORF2 products.



5. The synthesis of the first *I* factor DNA strand is complete resulting in an RNA-DNA duplex.



6. The RNA template is degraded by RNase H activity of the ORF2 products and the second strand of the *I* factor synthesised.

7. The synthesis of the second strand is complete. The strands ligated to chromosomal DNA creating a new copy of the *I* factor flanked by a target site duplication from the sequences between the staggered single strand break.

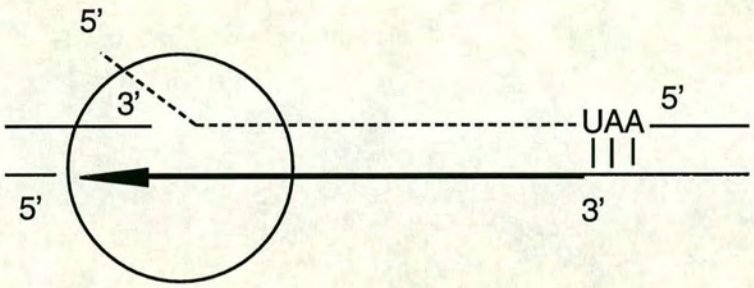


Figure 1.3 Generation of 5' truncated *I* element

The 5' truncated copy of the *I* factor is generated by premature termination of reverse transcription. This process is facilitated by template switching from the RNA intermediate (dashed line) to the left hand 3' overhang of the chromosomal DNA.

The intervening sequence between ORF1 and ORF2 of the *I* factor is not spliced out during transcription and is included in the 5.4 kb transcript. Translation of the ORF2 product must therefore be initiated by frameshifting or by a process of reinitiation at the first AUG codon of ORF2. These processes are inefficient, therefore translation of ORF2 may be one of the possible mechanisms for the regulation of *I* factor transposition.

1.6 Identification of *I* factor promoter

Functional *I* elements that are inserted downstream of the sequences that apparently lack promoter properties have been described (Pélisson 1981, Pritchard, *et al.* 1988). This suggests that transcription of the *I* factor is not dependent on flanking DNA sequences but is initiated from a promoter that lies entirely within the *I* factor sequence. Though this property of the *I* factor promoter is similar to that of promoter recognised by RNA polymerase III, it is unlikely that *I* factor is transcribed by RNA polymerase III as the *I* factor sequence lacks some characteristics of this type of promoter. The transcribed strand of the *I* factors contains several oligo(T) sequences that serve as the signal for RNA polymerase III transcription termination (Bogenhagen and Brown 1981). The first 200 bp of *I* factor sequence also lack any regions similar to consensus sequences of polymerase III promoter (Galli, *et al.* 1981). These characteristics suggest that *I* factor contains an internal RNA polymerase II promoter (Fawcett, *et al.* 1986). Internal promoter have been described for human LINE element L1Hs and the other two *D. melanogaster* LINE-like elements, *jockey* and *F* (Minchiotti and Di Nocera 1991a, Mizrokhi, *et al.* 1988, Swergold 1990). At least L1Hs and jockey are transcribed by RNA polymerase II.

An internal promoter of the *I* factor has been recently identified (McLean, *et al.* 1993). The strategy for detection of promoter activity was to link the putative promoter to a reporter gene and assay the activity by transient expression in tissue culture cells or in flies. Nucleotides 1-186 of the *I* factor were selected as a putative promoter. This

sequence makes up the entire 5' untranslated region of the *I* factor (Fawcett, *et al.* 1986) and, hence, is the most likely candidate for a promoter.

1.6.1 Promoter activity in *Drosophila* Schneider Line 2 cells

The putative promoter sequence was amplified by PCR from plasmid pI954 as this construct contains a complete *I* factor that, when introduced into a reactive flies, can transpose and induce hybrid dysgenesis (Pritchard, *et al.* 1988). The amplified product was cloned into plasmid pCAT.1 carrying the reporter CAT gene. The resulting construct, so called p186T.1, contains the putative promoter, nucleotides 1-186 of the *I* factor with a T at position +3, placed immediately upstream of CAT gene (figure 1.4).

Promoter activity was tested by transient expression of the CAT gene in *Drosophila* Schneider Line 2 cells (McLean, *et al.* 1993). The construct p186T.1 directs high CAT activity relatively to low activity from pCAT.1 (table 1.1). This is also true when the fragments containing putative promoter linked to the CAT gene, or the gene alone, were inverted suggesting that expression of CAT was not due to the sequences upstream of the *I* factor sequence. This result indicates that the 5' UTR of the *I* factor contains a promoter (McLean, *et al.* 1993).

In order to determine nucleotide sequences that represent a minimal promoter of the *I* factor, several 3' deletions of the putative *I* promoter fragment were constructed. CAT activity directed from these deletions are compared in table 1.1 (McLean, *et al.* 1993). A construct p100T with a deletion of the sequences to position +100 establishes similar level of CAT activity to that from p186T.1. However, further deletion to position +40 (p40T) increases the activity two fold . This suggests suppression of promoter activity from sequences between positions +40 and +100. The activity drops again with a deletion to position +28 (p28T) to the same level as that from p186T.1 and p100T suggesting that sequences between positions +28 and +40 positively act on transcription from *I* factor promoter. Nucleotides 1-12, carrying a G at position +3, in

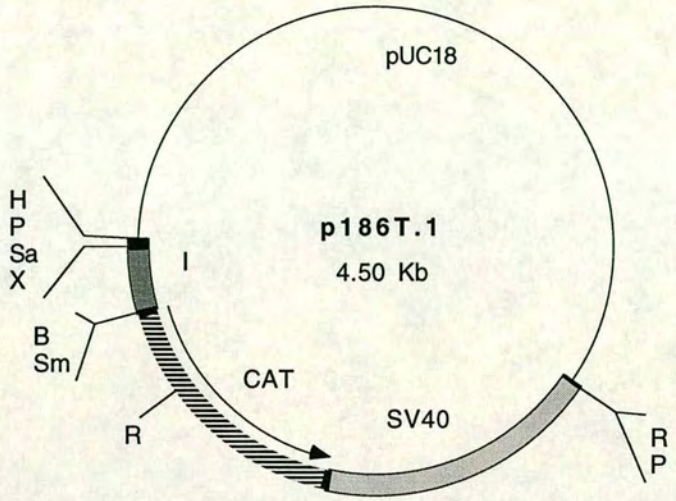


Figure 1.4 Physical map of p186T.1 (McLean *et al.* 1993)

An *Xba*I-*Bam*HI fragment containing nucleotides 1-186 of the *I* factor was placed adjacent upstream of the CAT gene of pCAT.0 (McLean, *et al.* 1993). The arrow represents the direction of CAT transcription.

the construct p12G do not give any significant higher CAT activity than the background level obtained from pCAT.1. Thus, the 3' limit of the *I* factor promoter lies between positions +12 and +28.

Since nucleotide at the third position of *I* factors can be either T or G, the effect of this flexibility on the promoter activity was determined (McLean, *et al.* 1993). The promoters with a G as the third nucleotide, for instance p186G and p40G, express about 60-80% higher CAT activity than that from the corresponding promoters with the T at the same position (Table 1.1).

Most of the RNA polymerase II promoters require either an upstream TATA box or a CpG rich sequence for transcription initiation. Neither of these sequences is present in the promoter of the *I* factor. Therefore, the promoter of the *I* factor falls into a group of RNA polymerase II promoters that lack both TATA box and CpG rich sequence. Instead of upstream elements, this group of promoter requires sequences at and downstream of transcription start site for basal transcription. Examples of the promoters in this group are those of the elements in LINE-like family such as *F* (Di Nocera, *et al.* 1983), *G* (Di Nocera 1988), *jockey* (Priimagi, *et al.* 1988), *Doc* (Driver, *et al.* 1989) and some developmentally regulated genes such as *Ultrabithorax* (Biggin and Tjian 1988) and *Antennapedia* (Perkins, *et al.* 1988) of *D. melanogaster* and murine terminal deoxynucleotidyl transferase (Smale and Baltimore 1989).

Nucleotides 1-40 of the *I* factor contain two motifs that are found conserved amongst the promoters of this type. The first motif, sharing a consensus CA(G/T)T, lies at the transcription initiation site or near the start of the element. The second motif centred around position +30 with a consensus AGACGTGPyPy (McLean, *et al.* 1993).

The importance of the CA(G/T)T motif present at the beginning of the *I* factor in promoter activity was tested by comparing CAT activity directed from the constructs having different nucleotide sequence in the first four positions (C. McLean, personal

Table 1.1 CAT activity directed from the *I* factor 5' UTR (From McLean, *et al.* 1993)

Construct	Mean% CAT activity
p186T.1	100
p100T	112
p40T	217
p28T	102
p186G	161
p40G	397
p12G	14
pCAT.1	11

Constructs p186T.1, p100T, p40T and p28T contain the full-length *I* factor 5' UTR (nucleotides 1-186) and the deletions (nucleotides 1-100, 1-40 and 1-28) with a T at position +3 inserted into pCAT.1 vector (McLean, *et al.* 1993) respectively. Constructs p186G, p40G, p12G contain the full-length *I* factor 5' UTR (nucleotides 1-186) and the deletions (nucleotides 1-40 and 1-12) with a G at position +3 inserted into pCAT.1 vector respectively.

communication). Mutations in this sequences result in reduction in CAT activity with the most severe effect when the consensus is replaced by ACCG in which case no activity could be detected at all. These results suggest an essential function of the CA(G/T)T motif for *I* factor promoter activity.

Construct p28G lacks the second conserved motif but retains ability to express CAT activity at the level of twofold lower than that from p40T which contains the conserved motifs. The AGACGTGPyPy is, therefore, required for high level of but not essential for expression from the *I* factor promoter in tissue culture.

Transcription of the *I* factor initiates at position +1

Primer extension analysis using transcripts synthesised from p40G showed that transcription from *I* factor promoter initiates at position +1 (McLean, *et al.* 1993). The construct p40G was chosen as a template for transcription because it retains both conserved motifs and direct highest CAT activity. No transcript was detected in the presence of 4 μg/ml of α-amanatin. The position of transcription initiation and the sensitivity of the transcript to α-amanatin are in agreement with the assumption that transcription is initiated from an internal promoter by RNA polymerase II.

1.6.2 Promoter activity in transformed flies

The putative promoter fragments containing nucleotide sequences within the 5' UTR of the *I* factor were tested for their ability to direct expression of the CAT gene in reactive flies. *P*-mediated transformation was used to make transformed lines (see method 2.2.4.2).

An *I* factor promoter on the transgene introduced into a reactive strain should behave in the same way as it does in dysgenic females because the promoter is entering a reactive background in both situations. Thus, the expression from the *I* factor promoter in the

transformed lines should be a good representation of the actual promoter activity during I-R hybrid dysgenesis.

Developmental profiles of the expression from the *I* factor promoter showed a steady increase in CAT activity throughout development in males but a burst of activity after eclosion in females (McLean, *et al.* 1993). Removal of nucleotides 101-186 changed the developmental profile. It exhibited high CAT activity during pupal stage in both males and females suggesting suppression of promoter activity from sequences between positions +101 and +186 during pupal development (C. McLean, unpublished data). Deletion of promoter sequences to positions +40 and +28 showed similar developmental profiles.

As *I* factor transposition occurs at high frequency in the germline of dysgenic females, its promoter must be regulated so that it is highly active under this condition. This seems to be the case when the expression from the *I* factor promoter in ovaries and in non-ovarian tissues (carcass) of transformed lines was compared (table 1.2). The promoter sequence comprising nucleotides 1-186 showed enhanced expression in ovaries. The activity is about twenty-fold higher in ovaries than in carcass. This value is presented as a ratio between the activity in ovaries and the activity in carcass (O/C). The ratio dropped about seven-fold when nucleotides 101-186 were deleted from the promoter sequences. The enhancement of the activity in ovaries could not be observed when the 3' deletion extended to position +40. These results indicate that the enhanced activity in ovaries required sequences downstream of position +40 (C McLean, unpublished data).

1.7 Regulation of I-R hybrid dysgenesis

The complexity of I-R dysgenic system is not yet clearly understood though some hypotheses have been proposed in order to explain how the system is regulated (Bucheton and Picard 1978, Finnegan 1989b).

Table 1.2 Ratio of CAT activities expressed from the *I* factor 5' UTR in ovaries and carcass of transformed flies (C. McLean, unpublished results)

Promoter	CAT activity (O/C)
1-186	20
1-100	3.8
1-40	0.11

SF sterility, the most prominent symptom of I-R hybrid dysgenesis, seems to be the outcome of an interaction between inducer and reactive states. A reduction in fertility by this manner is not confined only to SF females but also to their daughters. The daughters of SF females are less fertile when they inherit the chromosomes from fathers of a reactive origin than when they acquire the chromosomes from inducer fathers (Picard 1978).

The chromosomes of a reactive strain bear genetic elements that allow the cytoplasmic reactive state to be maintained and thus exhibit reactivity (Bucheton and Picard 1978). SF sterility could be observed as long as reactive state retains in the cytoplasm as in the case of the daughters of SF females that inherit from their mothers the cytoplasm of a reactive origin. Crosses between SF females and inducer males reduce the number of chromosomes of reactive origin in the offspring. A cytoplasmic state therefore changes to a lower level of reactivity resulting in more fertile daughters as compared to the daughters of crosses between SF female and reactive males in which reactive state is still maintained by the incoming paternal chromosomes.

RSF females, though themselves normally fertile, can sometimes produce daughters showing a low level of fertility depending on the genotype of their mates (Picard 1978). As RSF females contain an inducer cytoplasm that is maternally transmitted it is unlikely that they will produce female progeny suffering from SF sterility. However, proposing that reactive chromosomes carry a gene or group of genes whose products determine the cytoplasmic state, RSF females might have some potential for reactivity by receiving reactive chromosomes from their father. This can also explain low frequency of *I* factor transposition detected in RSF females (Picard 1976).

The regulation of *I* factor expression during I-R hybrid dysgenesis was followed by the use of *I* factor-*lacZ* translational fusion k160 in which the expression of *E coli.* β-galactosidase is controlled by the 5' untranslated region and part of ORF1 of the *I* factor (Lachaume, *et al.* 1992). Expression of this transgenic fusion was observed in

the germ-line of reactive and SF females but was barely detectable in RSF females while no expression was detected in inducer females. This correlates well with the level of dysgenic symptoms in each category of females.

Moreover, the expression of k160 was modified upon aging and heat treatment in a similar manner as was the reactivity (Lachaume and Pinon 1993). The k160 fusion was expressed in proportion to the level of reactivity, that is high level of expression was induced in strong reactive females whereas the expression was undetectable in weak reactive females. These correlations indicate that the level of reactivity positively regulates the expression of k160 fusion and thus of the *I* factor.

In addition, inducer state could also play a role in the regulation of I-R hybrid dysgenesis. It has been proposed that inducer phenotype is conferred by the components encoded by the *I* factors as strong reactive strain can be converted into inducer after being introduced into their genome a cloned *I* factor (Pritchard, *et al.* 1988). The regulatory molecules that regulate *I* factors could be the product of ORF1 that contains potential nucleic acid binding domains (Finnegan 1989b). Therefore *I* factor activity is inhibited by these molecules in inducer strain but is no longer inhibited in SF female since they are entering a reactive cellular environment in which the regulatory molecules are not present. However, this hypothesis can not explain clearly for the low level of *I* factor transposition in RSF females.

1.8 LINEs and LINE-like elements

The structure and properties of the *I* factors place them into a class of transposable elements that is known as non-viral retrotransposons or, more commonly, LINEs (Weiner, *et al.* 1986). LINEs are a family of long interspersed repetitive sequences found in a wide range of eukaryotic organisms. The elements in this family share several important features. Complete elements are about 5 to 7 kb long containing neither direct nor inverted terminal repeats. They are bordered by short direct repeats

of variable sequences that have been shown in many cases to be target site duplications. The 3' terminus of one strand is composed of an adenine rich sequence. The same strand also contains one or two open reading frames (ORFs). The putative polypeptide product of one of the ORFs shows homology to retroviral reverse transcriptase. These structural features suggest that LINEs transpose via RNA intermediates.

1.8.1 Mammalian LINEs

A family of LINEs, so called LINE-1 or L1, has been found widespread in mammals (Burton, *et al.* 1986). There are 10^4 to 10^5 copies of these elements per haploid genome making up about five per cent or more of the total DNA. The complete elements are 6 to 7 kb in length but the majority of the elements present in the genome are truncated at their 5' ends or are internally rearranged (Fanning and Singer 1987b). The central region of the elements contains two open reading frames and is flanked by 5' and 3' non coding regions that vary in size and sequence from the element of one species to another. In contrast, the sequences comprising open reading frames of the elements are highly homologous amongst the elements of several species from mouse to human.

Human L1 (L1Hs)

Despite the high copy number of L1 elements in the human genome, only a small proportion are full-length and presumed to be functional. In recent years, mutations in some human genes have been reported to be the consequence of insertions of L1Hs elements providing evidence that L1Hs have recently transposed in human genome.

Two insertion events of L1Hs were found in the *factor VIII* gene on the X chromosome of two individual patients suffering from haemophilia A (Kazazian Jr., *et al.* 1988). As the mothers of both patients have normal *factor VIII* genes, the

insertion must have occurred in the germline of their mothers. The insertion copies in both cases are truncated, containing only the 3' portion of L1Hs that includes much of the second ORF and polyadenylic acid tail (Woods-Samuels, *et al.* 1989). At least four alleles were identified at the *LRE-1* (L1 retrotransposable element-1) locus on chromosome 22 that might have been a progenitor of the inserts. One of them, L1.2B, was isolated from the genome of the mother of patient JH-27 by its sequence identity to the factor VIII insertion copy. This is therefore the likely original copy of the newly transposed element (Dombroski, *et al.* 1991).

Transposition of L1Hs is also possible in somatic cells as indicated by two insertions recently identified. One of these was found in a patient with adenocarcinoma of the breast where an element transposed into one *myc* allele of the carcinoma tissues but not into the corresponding allele in surrounding normal tissues (Morse, *et al.* 1988). More recently, an insertion into an exon of the *APC* gene in the cells of a colon tumour was identified (Miki, *et al.* 1992). The insertion was also found only in the affected cells as in the case of insertion into *myc* locus.

A full length L1Hs transcript has been detected in a human teratocarcinoma cell line, NTera2D1 (Skowronski and Singer 1985). This transcript is about 6.5 kb long, polyadenylated and represents the coding strand of the element. The transcript initiates at the nucleotide residue defined as the 5' end of the genomic L1Hs (Skowronski, *et al.* 1988). Several cDNA clones isolated from a library of L1Hs mRNA of NTera2D1 cells have been characterised. Slight differences among the sequences of these cDNAs suggests that they were transcribed from different L1Hs elements in the genome. The entire sequence of one of the clones, cD11, has been determined. It contains about 800 and 200 bp of the 5' and 3' non coding regions. Two ORFs exist, ORF1 is 1,122 bp long separated from 3,852 bp ORF2 by 33 residues (Skowronski, *et al.* 1988).

Swergold (1990) has identified in the 5' untranslated region of L1Hs a promoter potential that can promote transient expression in NTera2D1 cells with the most critical

sequence for transcription located in the first 100 bp of L1Hs. The transcription is initiated at or near the beginning of the element. The L1Hs-encoded proteins are preferentially expressed from this promoter in some cell types such as teratocarcinoma cells, NTera2D1 and 21201Ep, and choriocarcinoma cell line JEG-3 but not in HeLa cells (Leibold, *et al.* 1990, Skowronski, *et al.* 1988). The 5' region identified as the promoter of L1Hs has been shown to harbour three prominent transcriptional regulatory elements (Minakami, *et al.* 1992). The sequence of this region is comparatively rich in CpG dinucleotides (Hohjoh, *et al.* 1990, Skowronski, *et al.* 1988) and is differentially methylated in cells that exhibit different level of L1Hs expression (Thayer, *et al.* 1993). Thayer *et al.* (1993) suggested that the differential methylation of L1Hs regulatory sequences may be one possible explanation for cell type specific manner of L1Hs expression.

Translational products of ORF1 have been detected in extracts of teratocarcinoma cell lines NTera2D1 and 2102Ep , as well as in extracts of choriocarcinoma cell line JEG-3, as a 38 kDa polypeptide (Leibold, *et al.* 1990). Translation was initiated at the first AUG codon in ORF1. Despite their corresponding positions, no homology between ORF1 polypeptide of L1Hs and gag polypeptide of retroviruses have been found (Singer, *et al.* 1993). No DNA binding activity has been detected for ORF1 product (Holmes, *et al.* 1992), therefore the function of this polypeptide is still to be determined.

Deragon *et al.* (1990) have identified a reverse transcriptase activity that associated with a macromolecular complex in the microsomal fraction of NTera2D1 cells. The L1 RNA was detected in this reverse transcriptase-containing fraction suggesting that the activity identified could originate from L1 elements in NTera2D1 cells. As a predicted polypeptide encoded from L1Hs ORF2 exhibit homology to retroviral reverse transcriptase (Skowronski, *et al.* 1988), ORF2 of the element is likely to encode the identified reverse transcriptase activity. The L1Hs ORF2 has been expressed as a fusion protein using expression system based on *Saccharomyces cerevisiae*

retrotransposon Ty1 and has been shown to have reverse transcriptase activity (Mathias, *et al.* 1991).

Since the two ORFs of L1Hs are separated by a 33 bp inter-ORF region with an in-frame stop codon at both ends (Skowronski, *et al.* 1988) and there is no evidence for an ORF1/ORF2 fusion protein so far, the most likely mechanism for translation of ORF2 is an internal initiation by newly attached ribosomes (McMillan and Singer 1993, Singer, *et al.* 1993). The translation initiation of ORF2 was shown to be independent of that of ORF1 (Singer, *et al.* 1993).

Mouse L1 (L1Md)

L1Md is a major long interspersed repetitive element in the mouse genome. The full-length elements are about 7 kb long (Fanning 1983). Complete nucleotide sequences of three full-length L1Md elements have been reported (Loeb, *et al.* 1986, Voliva, *et al.* 1984). Unlike L1Hs, The two ORFs of L1Md, 1,137 and 3,900 bp in length, overlap by 14 bp (Loeb, *et al.* 1986, Shehee, *et al.* 1987). The second ORF (3,900 bp) shows amino acid sequence homology to several reverse transcriptases therefore suggesting that it encodes a protein with reverse transcriptase activity. Recent evidence showing retrotransposition of L1Md in baby hamster kidney cells has been reported (Evans and Palmiter 1991). When a putative reverse transcriptase coding region of this L1 element was deleted, the element still underwent retrotransposition in hamster cells suggesting that reverse transcriptase can be supplied by an endogenous enzyme.

In addition to the common features of all LINEs, L1Md elements share some unique features common to elements from this species. They have at least two distinct 5' end structures called A-type and F-type (Fanning 1983, Loeb, *et al.* 1986). The A-type elements contain homologous sequences at their 5' end due to tandem arrays of a monomer of 208 bp. The elements analysed so far have different numbers of these repeating units (Shehee, *et al.* 1987). The 5'-most unit of each element is about two-

thirds the length of the full-length one. The F-type element seems to be a major class of L1 element in the mouse genome. The F-type monomer is 206 bp and share no sequence homology to the A-type monomer (Padgett, *et al.* 1988). The 5' end of the F-type monomer shows 76% homology to a region within the first 70 bp of the downstream L1Md sequence. This homology provides a clue for explaining the existence of two different types of 5' end in L1Md family. An unequal crossing over could have occurred, in the region of shared homology, between a pre-existing F-type monomer or multimer in the genome that is not associated with L1 and an A-type L1 element. This should lead to an L1 element with a new 5' end of an F-type (Padgett, *et al.* 1988).

Despite no sequence homology, The A- and F-type monomers do share the consensus CCGCCC of RNA polymerase II promoter (Loeb, *et al.* 1986, Padgett, *et al.* 1988). It is possible that these tandem repeats play an important role in transcription and hence transposition of L1Md elements by acting as a promoter. A model of transposition has been proposed according to this assumption (Loeb, *et al.* 1986).

Rat L1 (L1Rn)

L1Rn is a LINE element cloned from rat genome (Witney and Furano 1984). About 40,000 copies of the elements make up to 10% of the rat genome. Full length element is 6.7 kb long. L1Rn exhibits common features of LINE family such as a 3' adenine-rich terminus, short direct repeated of the target sequence at both ends and long open reading frames homologous to those of other LINEs (D'Ambrosio, *et al.* 1986). The sequences 5' to the first ORF of rat and mouse L1 elements are completely distinct (Furano, *et al.* 1988). However, both regions are dominated by tandem duplication and contain more than 50% GC content. It is presumed that this sequence also contains promoter-like region of L1Rn. as in the case of L1Md. In fact, a sequence of 600 bp at the 5' end of newly isolated L1Rn has been shown to be able to function as a promoter *in vivo* (Nur, *et al.* 1988).

LINEs in other mammals

L1 elements are also present in several species where their sequences have been analysed. In rabbit, the element (L1Oc) was found located 3' to the β-1 globin gene. It shows sequence homology to human and rodent L1 in an ORF region (Demers, *et al.* 1986). Insertion of dog L1 element into cellular oncogene c-myc created 10 bp target sequence duplication. The element shows 62% homology to human L1 (Katzir, *et al.* 1985). Cat and primates such as slow loris and African green monkey also contain elements of the L1 family (Fanning and Singer 1987a, Grimaldi and Singer 1983, Hattori, *et al.* 1986, Lerman, *et al.* 1983). Sequences of L1 elements from four mammalian orders including primate, rodent (mouse), lagomorph (rabbit) and carnivore (cat) were compared in their predicted amino acid sequences and shown to have strong homology to reverse transcriptases as well as a putative metal binding domain of nucleic acid binding proteins (Fanning and Singer 1987a).

1.8.2 LINEs in Trypanosomatids

Trypanosome repeated sequence (TRS) or *ingi*

This element was found in up to 400 copies in the genome of *Trypanosoma brucei.* The majority are 5.2 kb long (Murphy, *et al.* 1987). Sequence analysis of complete 5.2 kb reveals no direct or inverted repeats but both ends are flanked by different separate halves of the previously described short transposable element RIME (Hasan, *et al.* 1982). *Ingi* contains an open reading frame that occupies almost their entire length encoding a protein homologous to reverse transcriptases (Kimmel, *et al.* 1987).

Transcription of *ingi* is developmentally regulated. The transcripts are much more abundant in the bloodstream form than in cultured procyclies. The majority of poly(A)$^+$ transcripts are longer than the unit length of the element. Therefore, the transcription of *ingi* is probably dependent on external promoters. Taken together, it

would appear that *ingi* has a preference for developmentally regulated promoters or that the element located downstream from such promoters are preferentially transcribed (Murphy, *et al.* 1987).

SLACS (Spliced-Leader Associated Conserved Sequence)

Aksoy *et al* (1987) reported that some of the spliced-leader (SL) RNA genes of the African trypanosome, *T. b. gambiense* contain interrupting sequences. Nine of about 300 copies of SL-RNA sequences are interrupted by insertion elements between 11^{th} and 12^{th} nucleotides from the 5' and of their coding sequence. These elements, so called *SLACS*, are flanked by 19 bp target sequence duplications and contain poly(A) tails at the 3' ends.

Complete nucleotide sequence of *SLACS* implies that the element is 6,675 bp long containing two ORFs that are separated by 79 nucleotides. The products of the two ORFs exhibit homology to *gag* and *pol* polypeptides (Aksoy, *et al.* 1990). Only 9 copies of *SLACS* are present in the genome. Different copies vary at their 5' ends due to different numbers of copy of repeat segments of about 185 bp. A unique property of *SLACS* is that no truncated element has been detected in *T. b. gambiense*.

CRE1

CRE1 interrupts miniexon genes of the trypanosomatid *Crithidia fasciculata* at a specific site in the conserved region of the genes. A complete *CRE1* element is 3,940 bp with single large ORF encoding polypeptide of 1,140 amino acids (Gabriel, *et al.* 1990). It was demonstrated that this ORF encodes a reverse transcriptase activity (Gabriel and Boeke 1991).

1.8.3 *Tad*: a LINE-like element in *Neurospora crassa*

Tad is a transposable element found in the Adiopodoume strain of *Neurospora crassa* collected form the Ivory Coast. *Tad* element was present in multiple copies in this strain (Kinsey 1989). It was detected as a 7 kb element inserting in two independent mutants of the *am* (glutamate dehydrogenase) gene (Kinsey and Helber 1989). These elements have been cloned and further characterised. The insertion of *Tad* resulted in duplication of target sites in both mutants.

Sequence analysis reveals that *Tad* contains general features of LINEs. It has no terminal repeats either in direct or inverted orientation and has two open reading frames with homology to the corresponding ORFs of other elements in this group. *Tad* is not present in laboratory strains but does exist in multiple copies in hybrids between the Adiopodoume strain and laboratory strains that lacks the elements (Kinsey and Helber 1989). The mobilisation of *Tad* in these hybrid is similar to the behaviour of the *I* factor of *Drosophila melanogaster*. However, there is no evidence to explain how transposition of *Tad* is controlled in *Neurospora* so far.

It was shown that *Tad* is able to transpose between nuclei in heterokaryons formed between strain containing *Tad* and strain that lacks the element (Kinsey 1990). Transposition of *Tad* must therefore occur via a cytoplasmic intermediate as no evidence for nuclei fusion in heterokaryons was detected. More recently, an intron inserted in ORF1 of *Tad* has been shown to be removed from the element that transpose to naive nuclei in heterokaryons (Kinsey 1993). This result indicates that *Tad* transpose via an RNA intermediate and therefore is a retrotranposon.

1.9.4 Ribosomal insertion elements

A fraction of ribosomal genes in a large number of insect species are interrupted by specific DNA inserts (Jakubczak, *et al.* 1991). These inserts belong to two types of

elements, *R1* and *R2*, of which those in *D. melanogaster* (Dawid and Rebbert 1981), *Bombyx mori* (Eickbush and Robins 1985, Fujiwara, *et al.* 1984) and *D. virilis* (Rae, *et al.* 1980) have been well characterised. The *R1* and *R2* elements insert at specific sites in a highly conserved region of the 28S rRNA genes. *R1* elements interrupt the 28S gene at a location approximately two-thirds of the distance from the 5' end of the gene and are flanked by 14 bp duplication of rDNA sequence. The insertion of *R2* elements is 74 bp upstream of that of *R1*. No duplication of rDNA sequence is found associated with *R2* insertion. A few insertions outside the rDNA units have also been detected though these elements seem to be non-functional. These insertions, despite their non-rDNA targets, display similar target sequence specificity to those in the 28S gene (Brown, *et al.* 1984, Xiong, *et al.* 1988).

Complete nucleotide sequences of R1 and R2 elements of *D. melanogaster* (*R1Dm* and *R2Dm*) and *B. mori* (*R1Bm* and *R2Bm*) have been reported (Burke, *et al.* 1987, Jakubczak, *et al.* 1990, Xiong and Eickbush 1988b). The elements of the two species are similar in their structures. The *R1* element is about 5 kb long and contains two overlapping ORFs. The first, ORF1, exhibits characteristics of retroviral *gag* genes. The second, ORF2, encodes a protein with similarity to reverse transcriptase-like enzyme. The *R2* elements of both species have a single ORF that occupies almost the entire length of the element. The product of this ORF also shows homology to reverse transcriptase.

Functional expression of the single ORF of *R2Bm* reveals that it encodes a double stranded endonuclease that can specifically cleave the 28S gene at the R2 insertion sites (Xiong and Eickbush 1988a). This sequence specific cleavage by the R2 encoded endonuclease could contribute to the insertion of the elements at the same target sequences outside the rRNA gene as has been proposed by Xiong *et al* 1988.

1.9.5 LINE-like elements in *Drosophila*

Doc

Doc has been found associated with two spontaneous mutations in *D. melanogaster*. In the *white-1* (w^1) mutation, *Doc* is inserted in the promoter of the *white* gene (O'Hare, *et al.* 1991) while in *suppresser of forked* mutation ($Su(f)^{S2}$), the insertion of *Doc* element is within the *Su(f)* transcription unit and disrupts the sequence coding for a *Su(f)* gene product. The elements associated with both mutations have been cloned and analysed (Driver, *et al.* 1989) and their DNA sequences determined (O'Hare, *et al.* 1991). They are about 4.7 kb long and display a structure similar to that of several LINEs. *Doc* is transcribed, with no obvious tissue specificity, as a 5 kb polyadenylated RNA that is presumed to be a full length transcript (Vaury, *et al.* 1994).

F element

F elements are found interspersed with ribosomal insertion-like sequences in *D. melanogaster* (Pardue and Dawid 1981). They are 4.7 kb long with an oligo(A) stretch at one end. The insertion of the element generates target site duplication of 8-13 bp. The genome of *D. melanogaster* contains about 50 copies of *F* elements that could be found at 25 euchromatic sites and in chromocentre (Dawid, *et al.* 1981). Recent evidence for *F* transposition was found in w^{i+A}, a revertant of *white-ivory* mutation (Di Nocera, *et al.* 1983). Transcription of a full length *F* poly(A)$^+$ transcripts has been detected in Schneider line 2 tissue culture cells (Minchiotti and Di Nocera 1991b) providing an evidence to support retrotransposition of the elements.

G element

A *G* element was first identified as a sequence interrupting an *F* element that is inserted in ribosomal insertion-like sequence (Di Nocera and Dawid 1983). All the *G* elements characterised so far are associated with repeated DNA and, unlike other LINE-like elements in *D. melanogaster*, are restricted to the chromocentre.

At least two independent insertions of *G* elements in rDNA region have found at identical sequences. The site of insertion of *G*, *R1* and *R2* elements are in the same region of the 28S rRNA gene (Di Nocera, *et al.* 1986). The homology between the insertion sites might be associated with sequence-specific endonuclease. This suggests a possible role for a sequence-specific endonuclease activity associated with one of the gene products in determining the integration site of the *G* element as has been described for *R2Bm* (Xiong and Eickbush 1988a).

Jockey

The first *jockey* element was cloned from the segment of the locus *cut* in the ct^{MRpN} mutation (Mizrokhi, *et al.* 1985). The nucleotide sequence of *jockey* reveals that it has all the typical features of LINEs including two overlapping ORFs.(Priimagi, *et al.* 1988). The polypeptide encoded by ORF1 contains $CX_2CX_4HX_4C$ motifs characteristic of nucleic acid binding domain of retroviral gag polypeptides. The second ORF codes for reverse transcriptase activity. A 92 kDa *jockey* encoded protein identified as a reverse transcriptase has recently been expressed in E. coli (Ivanov, *et al.* 1991).

Two major classes of *jockey* have been found in the *Drosophila* genome; a 5 kb full-length element and an internally deleted element of approximately 2.5 kb in length (Priimagi, *et al.* 1988). The transcripts corresponding in length to both full-length and deleted copies of *jockey* have been detected in different developmental stages of *D.*

melanogaster as well as in Schneider 2 culture cells (Mizrokhi, *et al.* 1988). *Jockey* is transcribed from an internal RNA polymerase II promoter. The first 13 nucleotides of the element are important for promoter activity as the activity disappears completely when these nucleotides are removed (Mizrokhi, *et al.* 1988).

BS elements

BS was found associated with the mutation Hw^{BS} that is a derivative of Hw^{1} mutation of the achaete-scute (AS-C) complex in *Drosophila melanogaster* (Campuzano *et al.* 1986). The Hw^{1} mutation is associated with an insertion of a copy of *gypsy* within the AS-C transcription unit. Chromosomes carrying the Hw^{BS} mutation have an additional 8 kb of DNA inserted within the Hw^{1} *gypsy* element. Restriction mapping of this insertion suggested that it belongs to a family of elements with long terminal inverted repeats as this element appeared to have inverted repeats of 2.5 kb at both ends.

Udomkit *et al.* (1995) have determined the complete sequence of this insertion and have shown that this is in fact two copies of a novel LINE-like element, called *BS*, inserted in the opposite orientation 337 bp apart. One of these copies, *BS1*, may represent full-length element. It is 5,126 bp long and has two open reading frames that appear to encode a *gag*-like polypeptide and a reverse transcriptase.

HeT-A and *TART* elements

HeT-A is a transposable element that belongs to a family of repeated DNA found only in pericentric and telomeric heterochromatin (Valgeirsdottir, *et al.* 1990). Transposition of *HeT-A* is strictly limited to chromosome ends and is believed to be a mechanism which *Drosophila* employs to balance the loss of chromosome ends that occurs as a result of incomplete terminal DNA replication (Biessmann, *et al.* 1992a).

The structure of the *HeT-A* element derived from the consensus sequence defines an element of about 6 kb long with general characteristics of a class of LINE-like transposable elements (Danilevskaya, *et al.* 1994). The coding region of *HeT-A* has overlapping ORFs with a -1 frameshift. Both ORFs contain motifs similar to those found in ORF1 proteins of many LINEs. No reverse transcriptase domain has been observed in any *HeT-A* element so far however.

The 5' and 3' regions of *HeT-A* have unique features that are not found in other LINEs. The oligo (A) stretches at the 3' end of *HeT-A* is preceded by a non-coding region, more than 2 kb in length made of well-conserved sequence with an 80 bp repeat motif. This sequence could affect the structure of chromatin containing the element. The 5' region has very regularly spaced A-rich clusters that might produce a terminal structure for the chromosome (Biessmann, *et al.* 1992b, Danilevskaya, *et al.* 1994).

Another recently characterised transposable element associated with *Drosophila* chromosome termini is known as *TART* (telomere associated retrotransposon) (Levis, *et al.* 1993). *TART* has a feature more typical of LINEs than *HeT-A* in that it has an ORF with homology to the reverse transcriptase domain of LINEs. *TART* is like *HeT-A* in having an exceptionally long 3' untranslated region suggesting that this is a unique structure of a novel class of LINE-like elements in *Drosophila*. Nevertheless, there is no significant similarity between the 5.1 kb 3' UTR of *TART* and the consensus of *HeT-A* sequences. Direct evidence for *TART* transposition to chromosome ends has been found recently (Sheen and Levis 1994).

1.10 Other systems of hybrid dysgenesis

1.10.1 P-M hybrid dysgenesis

P-M hybrid dysgenesis occurs when males of a P strain are crossed with females of an M strain. The progeny of this cross manifest abnormalities in their germ cells including

high frequencies of mutations, chromosomal rearrangements and male recombinations (Kidwell, *et al.* 1977). The reciprocal cross as well as crosses between P or M strains do not develop dysgenesis. Unlike SF sterility in the I-R system, both male and female progeny of P-M dysgenic crosses suffer from sterility, called gonadal dysgenesis, resulting from a failure of the gonads to develop. Expression of sterility is affected by the temperature at which the flies are raised. A high developmental temperature tends to increase the frequency of sterility whereas a low developmental temperature inhibits the expression of some dysgenic traits (Kidwell, *et al.* 1977).

The *P* factors

The P strains are distinguished from the M strains by the presence of a P strain-specific transposable DNA family called *P* factors that are dispersed over all the major chromosomes (Engels 1979b). The *P* elements were identified during the investigation of mutations at the *white* locus arising during P-M hybrid dysgenesis. The elements associated with these *white* mutation were homologous in sequence but were heterologous in size ranging from 0.5-1.6 kb (Rubin, *et al.* 1982). The fragment from one of these *white* mutant alleles ($w^{\#12}$) containing a 1.2 kb *P* element insertion was used to probe genomic fragments of different P and M strains. The result demonstrated that the $w^{\#12}$ *P* element insertion is homologous to repeated DNA sequences present in all P strains examined but was missing from M strains (Bingham, *et al.* 1982). *In situ* hybridisation reveals about 30-50 copies of *P* elements distributed among all the major chromosome arms.

A complete *P* factor is 2.9 kb long containing 31 bp perfect inverted terminal repeats and produces 8 bp duplication of the sequence found only once at the site of insertion (O'Hare and Rubin 1983). The complete elements have four open reading frames, ORF0, 1, 2 and 3, encoding an 87 kDa *P* transposase required for transposition (Rio, *et al.* 1986).

Incomplete *P* elements can also be found with internal deletions. These elements can not synthesise transposase and therefore are non-autonomous. However, they can be mobilised in *trans* when a complete element is present in the genome (Engels 1984, Rubin and Spradling 1982).

The *P* factor has been demonstrated to play an important role in P-M dysgenic system. The sn^w mutant allele at the *singed* (*sn*) locus produced during P-M hybrid dysgenesis is stable in an M cytotype but is hypermutable under the dysgenesis condition (Engels 1979a). The sn^w allele is caused by two defective *P* elements inserted in inverse orientation. During P-M hybrid dysgenesis one or the other of the two *P* elements excised at high frequencies resulting in two classes of sn^w derivatives, sn^e-the extreme phenotype and sn^+-the allele with almost wild type phenotype (Roiha, *et al.* 1988). The sn^e and sn^+ could also be recovered in the progeny of embryos of an M strain carrying the sn^w allele injected with a recombinant plasmid containing *P* factor DNA (Spradling and Rubin 1982). These events indicate that hybrid dysgenesis is related to *P* factor mobilisation.

Regulation of *P* transposition

Germline specificity

Transposition of *P* element is tissue-specific, restricted only to the germline tissues (Engels 1983). This germline specificity of *P* element transposition is not controlled by *P* promoter as transposition is still limited to the germline cells of the transformants carrying the recombinant in which *P* element coding sequences are placed under control of *hsp70* promoter that functions in both somatic and germline tissues (Laski, *et al.* 1986). Furthermore, *P* element promoter has been shown to be active in somatic tissue producing 2.5 and 3.0 kb transcripts. However, these somatic *P* transcripts would not be related to transposase activity as no transposition has been observed in somatic cells.

Splice site mapping of somatic transcripts showed no splice of the IVS3 between ORF2 and ORF3. The translation would then terminate before ORF3 resulting in translated product that lacks message from ORF3 (Laski, *et al.* 1986). Karess and Rubin (1984) have shown that all of the four open reading frames are essential for transposase function as a *P* element with a frameshift in ORF3 failed to destabilise the sn^w locus.

Laski *et al* (1986) tested whether the IVS3 splicing is the regulatory event that limits *P* element transposition to germline cells by assaying for the ability of a *P* element from which IVS3 has been precisely removed to stimulate somatic excision and transposition. Both activities were detected indicating the requirement of IVS3 splicing for the production of transposase, a process that is restricted to the germ lines. The transposase encoded by this spliced RNA is 87 kDa and is capable of precisely catalysing the excision of *P* element in tissue cultures (Rio, *et al.* 1986).

Siebel and Rio (1990) provided evidence that tissue specific splicing of IVS3 is regulated by inhibitor(s) present in somatic tissues. They showed that an IVS3 pre-mRNA substrate was accurately splice in human Hela cell extract. This splicing was inhibited if the IVS3 pre-mRNA was preincubated with *Drosophila* somatic (Kc cells) extract. Heating the *Drosophila* Kc extract for 5 minutes at 100°C eliminated the inhibitory effect suggesting that a protein component is involved in this process. Indeed, a 97 kDa protein in Kc extract has been found to preferentially interact with IVS3 RNA at the 5' exon region ending 2 nucleotides upstream from the 5' splice site of IVS3 (Siebel and Rio 1990).

A *cis*-acting sequence required for germline specific splicing of IVS3 has also been identified (Chain *et al.* 1991). This sequence is confined to a 20 bp regulatory region located in the exon, 12 to 31 bases from the 5' splice site. It is possible that the

interaction between this *cis*-acting sequence and the regulator protein identified inhibits splicing by blocking one of the early step in spliciosome assembly.

Cytotype regulation

The genetic abnormalities induced during P-M hybrid dysgenesis are due to the mobilisation of *P* element when it is paternally introduced into cytoplasmic state of an M strain. Transposition of *P* element occurs at very low frequency such that no dysgenesis could be detected in the progeny of reciprocal crosses as well as in the progeny of P strains (Preston and Engels 1984). This indicates that *P* transposition is regulated by a maternally transmitted cellular state (cytotype). P cytotype, a characteristic of P strain, represses transposition whereas M cytotype, a characteristic of M strain, allows transcription to occur. Cytotype is determined by both chromosomal and cytoplasmic components. It is maternally inherited over several generations but ultimately the cytoplasmic state is determined by the chromosomes (Engels 1979a). Introduction of a single *P* factor into an M cytotype can switch the cytotype to P. This process takes up to 30 generations after the copy number of *P* elements has reached that of strong P strain (Daniels, *et al.* 1987).

The maternally inherited component that confers P cytotype is thought to be the accumulation of the product encoded by chromosomal *P* elements in the oocytes of the P females (Rio 1991). A possible candidate for a negative regulator of transposition is a 66 kDa protein encoded by IVS3 containing *P* element mRNA (Rio, *et al.* 1986). This protein has been tested for its ability to confer P cytotype (Misra and Rio 1990). Transposase activity was repressed in both somatic and germline tissues of the M strain flies transformed with modified *P* element producing only the 66 kDa protein. The strength of repression varied depending on the sites of integration (Misra and Rio 1990, Robertson and Engels 1989). The position effect of *P* element expression may cause the delay in cytotype switching.

The expression of *P* factor is regulated by P cytotype at the level of transcription. Lemaitre and Coen (1991) have shown that the level of *LacZ* mRNA of the *LacZ* gene expressed from the *P* element promoter was much lower in P strain than in M strain. The mechanism of transposition repression by the 66 kDa protein is yet to be studied. However, as most of the sequence of the 66 kDa protein is contained within the transposase, it is possible that they compete for the same binding site on the *P* element.

Ronsseray *et al* (1993) have investigated the nature of maternal inheritance when chromosomal *P* elements are not present. Zygotes derived from a P cytotype female but not inheriting any *P* elements have nevertheless received an extrachromosomal component. This component is insufficient to specify the P cytotype if the zygote does not carry chromosomal *P* elements but can promote P cytotype determination if *P* elements have been introduced paternally. This extrachromosomally inherited state, so called pre-P cytotype, is not transmitted to the following generation unless the chromosomal *P* elements are present.

P element transposition

P-encoded transposase recognises internal sites on *P* element adjacent to the perfect 31 bp inverted repeat at each end upon transposition. Mullins *et al* (1989) have identified short DNA segments at each end of *P* element that are important for transposition. These segments contain specific binding site for both transposase (Kaufman, *et al.* 1989) and a *Drosophila* protein, IRBP, that may also be required for transposition (Rio and Rubin 1988).

Engels *et al* (1990) provide strong evidence that *P* element transposition occurs by a non-replicative cut and paste reaction. The excision of a *P* element probably results in a gapped donor site that could be repaired by double-strand gap repair mechanism. The outcome at the donor sites after the repair is dependent on the template used. If the template for double strand gap repair contains a *P* element at the same site as the

donor, as would be the case for the sister chromatid, the repair will produce a chromosome that is identical in appearance to the donor. This process would appear to be identical to replicative transposition except the actual DNA strand transfer does not involve DNA synthesis. The precise excision event will be obtained if the wild type homolog is used as a template. According to this model, the internally deleted *P* elements could result from incomplete double-strand gap repair rather than actual deletion (Craig 1990).

1.10.2 H-E hybrid dysgenesis

McGinnis *et al* (1983) reported a variant *Sgs-4* glue protein gene of *Drosophila* that showed a 50-100 fold reduction in expression. The influence on expression was shown to be correlated with the insertion upstream of the *Sgs-4* TATA box of a novel transposable element called *hobo*.

The family of *hobo* transposable element contains both the 3.0 kb fully functional elements and internally deleted, defective elements (Streck, *et al.* 1986). *hobo* displays a close resemblance to the *P* element in several respects. It has short (12 bp) inverted terminal repeats, creates an 8 bp duplication upon insertion and varies in size (O'Hare and Rubin 1983).

A full-length *hobo* element contains a 1.9 kb open reading frame (ORF1) downstream of putative TATA and CAAT sequence. The expected product of ORF1 is 75 kDa containing 644 amino acids. ORF1 displays unusual codon usage that is different from those used in other *D. melanogaster* genes. However, it is similar for many amino acids to the codon usage of the *P* element transposase though no similarity in amino acid sequences has been found between the ORF1 product of *hobo* and the *P* element transposase (Streck, *et al.* 1986). Moreover, the ORF1 product of *hobo* showed amino acid similarity with the transposase of *Activator* element (*Ac*) from *Zea mays* and *Tam3* from *Antirrhinum majus*. Mutation analysis indicated that a c-terminal region of

ORF1 that is conserved among these elements is essential for *hobo* transposase activity. This amino acid sequence similarility between short inverted repeat elements in different kingdoms represents the possibility of horizontal transmission of genetic information between plants and animals (Calvi, *et al.* 1991).

hobo was found associated with several mutations induced by male recombination factor 23.5MRF (Yannopolous, *et al.* 1987). *D. melanogaster* strain bearing 23.5MRF induced hybrid dysgenesis that is similar, though with some particularly different characteristics, to the P-M system. The 23.5MRF could also induce GD sterility when crossed to some P strain females while no effect was observed in the reciprocal crosses (Yannopoulos, *et al.* 1986). The strain carrying 23.5MRF shows a pronounced P cytotype but is unable to destabilise sn^w phenotype when crossed to the M strain. These evidences suggest that the *P* elements in this strain are inactive and that the dysgenic properties could be attributed to the presence of functional *hobo* elements (Yannopolous, *et al.* 1987).

Regarding the *hobo* elements, *Drosophila* strains can be identified as H strain, a strain containing both complete and internally deleted derivatives of the elements, or E strain that lacks the complete *hobo*s. A cross between H strain containing a *hobo* element inserted in the *decapentaplegic* gene complex (DPP-C) and E strain resulted in numerous DPP-C mutations involving chromosomal rearrangements and deletions associated with pre-existing *hobo* elements. This hypermutability occurred premeiotically in the germ lines of the hybrids (Blackman, *et al.* 1987). Unlike P-M or I-R dysgenic systems, H-E hybrid dysgenesis does not seem to be governed by maternal cytotype as high mutation rates were also observed in the reciprocal cross, that is cross between E males and H females. Furthermore, there are some evidences that suggest that there is *hobo* activity in the somatic cells though at very low levels (Kim and Belayaeva 1991, Yannopolous, *et al.* 1987).

Germline specificity of *hobo* elements is due to regulation at different level from that of *P* elements. Whereas intact *hobo* elements are germline specific, a *hsp70* promoter-*hobo* transposase fusion is active in the soma. Moreover, replacement of the *P* promoter of Δ2-3 construct, the construct containing *P* element from which IVS3 has been removed, that is active in both germ lines and soma with the putative *hobo* promoter reduced transposition rate at least 1000-fold in somatic tissues conferring germline bias to Δ2-3 *P* transposase activity (Calvi and Gelbart 1994). Therefore, germline specificity of *hobo* transposition is due to regulation of transposase production at the level of transcription, probably by inhibition of *hobo* transcription by a negatively-acting mechanism in the soma.

Interestingly, despite the evidence for germline specificity of *hobo* transposase, it has been recently shown that *hobo* is capable of somatic excision during early development and its mobility is not restricted in distantly related *Drosophila* species. Furthermore, Drosophila species without *hobo* elements are able to mobilise *hobo* in the absence of helper plasmid. This suggest a hobo-related cross mobilising system in these species (Handler and Gomez 1995).

CHAPTER 2

Materials and methods

2.1 Materials

2.1.1 Chemicals

The chemicals used in this study were either products for molecular biology or analytical grade chemicals. They were supplied from Sigma chemical company, Fisons, BDH, Boehringer Mannheim, Merck and Pierce.

2.1.2 Enzymes

All restriction enzymes, Klenow enzyme, T4 DNA ligase, Taq DNA polymerase and their buffers were purchased from Boehringer Mannheim.

2.1.3 Isotopes

α -^{32}P-dCTP (3,000 Ci/mmol)

α -^{35}S-dATP (400 Ci/mmol)

D-threo-[dichloroacetyl-1-^{14}C] chloramphenicol (53 mCi/mmol)

The isotopes used in this study were supplied from Amersham International.

2.1.4 Bacterial media

Luria broth: (per litre)

Difco Bacto tryptone, 10 g; Difco Bacto yeast extract, 5 g; NaCl, 5 g

adjusted to pH7.2

Luria agar: (per litre)

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

2.1.5 *Drosophila* media

French fly food:

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

Egg collecting medium:

Difco Bacto agar, 1.5 g; pure apple juice, 100 ml

2.1.6 *Drosophila* tissue culture medium

Schneider's *Drosophila* medium (Gibco) supplemented with 5% foetal calf serum (Sera-lab), 100 u/ml penicillin and 100 µg/ml streptomycin (Gibco)

2.1.7 Solutions

Buffers for electrophoresis

50 x TAE: (per litre)

Tris base, 242 g; glacial acetic acid, 57.1 ml; 0.5 M EDTA pH8, 100 ml

20 x TBE: (per litre)

Tris base, 216 g; boric acid, 110 g; EDTA, 18.6 g

Buffer for dissection of *Drosophila* tissues

Ringers solution: (per litre)

NaCl, 6.5 g; KCl, 0.14 g; $NaHCO_3$, 0.2 g; $NaH_2PO_4.2H_2O$, 0.01 g; $CaCl_2,2H_2O$, 0.12 g

Solutions for nuclear extraction

MTBS:

135 mM NaCl; 5 mM KCl; 1 mM $CaCl_2$; 0.5 mM $MgCl_2$; 1.5 mM Na_2HPO_4; 2.5 mM Tris.HCl pH7.5

Solution I:

10 mM HEPES pH7.9; 10 mM KCl; 1.5 mM $MgCl_2$; 0.5 mM DTT; 0.5 mM PMSF

Solution II:

10 mM HEPES pH7.9; 400 mM NaCl; 1.5 mM $MgCl_2$; 0.5 mM DTT; 0.5 mM PMSF; 5% glycerol

2.1.8 Bacterial strains

NAME	GENOTYPE	REFERENCE
RR1 (E. coli)	*supE44, hsdS20($r_B^-m_B^-$). ara-14, proA2, lacY1, galK2, rpsL20(Sm^r), xyl-5, mtl-1, lacZ*$^-$	Bolivar *et al.*, 1977

2.1.9 Plasmids

NAME	DESCRIPTION and USE	REFERENCE
pCAT.1	CAT reporter plasmid constructed by inserting CAT-SV40 fragment of pC4CAT (Thummel *et al.*, 1988) into pUC18 between *Bam*HI/*Sac*I sites.	McLean *et al.*, 1993
phs70Δ2-3wc	Helper plasmid for *P* element mediated transformation containing a *P*-element, lacking the germline specifically spliced intron, transcribed from the *hsp70* promoter	Gift from Jean-Maurice Dura

NAME	DESCRIPTION and USE	REFERENCE
pW8	*P*-element mediated transformation vector containing *white* as a selection marker. Used as vector for *I* promoter-CAT fusions transformation into flies.	Klemenz *et al.*, 1987
p186T.1	Derivative of pCAT.1 in which the *XbaI-Bam*HI PCR fragment, containing nucleotides 1-186 of the *I* factor with a T at the third position, was inserted upstream of the CAT gene. Used as template in PCR to amplify fragments with a deletion in the first 186 nucleotides of the *I* factor.	McLean *et al.*, 1993
p186Δ4.2	Derivative of pCAT.1 in which the *Xba*I-*Bam*HI PCR fragment containing nucleotides 1-186 of the *I* factor, the first four position of which was replaced by the sequence ACCG, was inserted upstream of the CAT gene.	McLean *et al.*, 1993

2.1.10 *Drosophila* strains

NAME	GENOTYPE	I-R/P-M STATE	REFERENCE
Cha	+	R/M	Pélisson, 1981
Cha-RC^{+}	+	I/M	Pélisson and Bregliano, 1987
W^{K}	*White*$^{-}$	R/M	Luning, 1981

2.1.11 Cell culture lines

NAME	COMMENTS	REFERENCE
Schneider Line 2	Derived from Oregon-R late embryos, 60-80% tetraploid and reactive.	Schneider, 1972

2.1.12 Oligonucleotides

NAME	SEQUENCE 5'-3'	COMMENTS
664C	G*TC TAG* **ACA TTA CCA CTT CAA CCT CCG**	+ strand PCR oligonucleotide from base +1 to +20 of the *I* factor sequence
089Y	CG*G GAT CC*G **ATT GTT TAA GGG CTT GAG CCC GGA** T**Δ**G**T TGT TAA GTT**	- strand PCR oligonucleotide from base +186 to +128 of the *I* factor sequence with a deletion of bases +139 to +157
429T	*CTC GAG* **GAT TGT TGG TTA AGG GCT TG**	- strand PCR oligonucleotide from base +186 to +167 of the *I* factor sequence

NAME	SEQUENCE 5'-3'	COMMENTS
430T	*CTC GAG* **CAG TCT AAA GCC TCG TTC GC**	+ strand PCR oligonucleotide from base +41 to +61 of the *I* factor sequence
431T	*CTC GAG* **CAA ATA TCA ACC ACA AAG AG**	+ strand PCR oligonucleotide from base +99 to +118 of the *I* factor sequence
G0762	CG*G GAT CC*G **ATT GTT GGT TAA GGG CTT GAG CCC GGA T/TG TTT ATC AAG**	- strand PCR oligonucleotide from base +186 to +90 of the *I* factor sequence with a deletion of bases +101 to +157
G0763	CG*G GAT CC*G **TTG TTA AGT TTT TTA**	- strand PCR oligonucleotide from base +138 to +123 of the *I* factor sequence
G5325	CG*C TCG AG*G **TTT ATC AAG ATT TTG CTG**	-strand PCR oligonucleotide from base +99 to +81 of the *I* factor sequence
F850	GTT CTT TAC GAT GCC	oligonucleotide starting 61 bases downstream of the first AUG codon of the CAT gene reading towards *I* factor sequence
M13 sequencing primer	GTA AAA CGA CGG CCA GT	-20mM13 17mer sequencing primer

I factor sequences are shown in bold type.

Recognition sequences for restriction enzymes are shown in italic.

Nucleotides flanking the deletion are underlined, Δ represents the deletion.

2.2 Methods

2.2.1 Manipulation of bacteria

2.2.1.1 Growth of bacterial cultures

Liquid culture was used to multiply large amount of bacteria. A single colony of *E. coli* was inoculated using a sterile inoculating loop to Luria broth, supplemented with 100 μg/ml of ampicillin when required, and was grown at 37°C with shaking for 14-16 hours. Cultures were normally grown in conical flasks with a capacity of 5 times of the culture volume. Universal glass bottles were used for growing cultures of small volume (3 - 5 ml).

2.2.1.2 Storage of bacterial cultures

A single colony of *E. coli* was picked with a sterile needle and stabbed through LB agar in a small glass vial. After overnight growing at 37°C, the stabs were stored at room temperature. Alternatively, cultures can be stored for long period by mixing 1 ml of fresh overnight cultures of *E. coli* with 1 ml of glycerol and stored in a sterile vial at -70°C. For recovery, the cultures were streaked out either from stabs or from thawed frozen cultures on Luria agar plates, with antibiotic if required.

2.2.1.3 Preparation of competent cells

Twenty-five millilitres of Luria broth was inoculated with 0.5 ml of an overnight culture of *RRI*. The cells were further grown at 37°C with shaking until an OD_{650} of 0.4 - 0.6 was reached. Cells were placed on ice for 10 minutes and centrifuged at 3,000 rpm for 5 minutes at 4°C. The pellet was resuspended in 2.5 ml of chilled TSB (10% PEG 3,000, 5% DMSO, 10 mM $MgCl_2$ and 10 mM $MgSO_4$ freshly made

up in Luria broth). Cells became competent after 30 minutes incubation on ice and could be used immediately or be stored at -70°C in 0.1 ml aliquot.

2.2.1.4 Transformation

Ten microlitres of ligation reaction (see method 2.2.2.7.) was added to 0.1 ml of competent *RRI* cells and left on ice for 30 minutes. Nine hundred microlitres of cold TSB and 20 µl of 20% glucose were added and incubated with shaking at 37°C for 1 hr. Appropriate volume of aliquots was spread onto Luria agar with 100 mg/ml ampicillin and incubated overnight at 37°C.

2.2.2 Nucleic acid preparation and manipulation techniques

2.2.2.1 Small scale preparation of plasmid DNA (Birnboim and Doly, 1979)

A single colony of plasmid carrying strain was grown overnight at 37°C in 3 ml of Luria broth containing 100 mg/ml of ampicillin. The pellet obtained from 1.5 ml of the overnight culture by centrifugation in a 1.5 ml Eppendorf tube was resuspended in 100 µl of solution containing 25 mM Tris HCl, pH8.0, 10 mM EDTA and 50 mM glucose. The cells were lysed by adding 200 µl of lysis buffer (0.2 N NaOH, 1% SDS) and then left on ice for 5 minutes. The lysate was subsequently neutralised with 150 µl of 3 M NaAcetate pH5.5. After standing on ice for at least 15 minutes genomic DNA and cell debris were precipitated by centrifugation in a microfuge for 10 minutes. The supernatant was then transferred to a new tube and plasmid DNA in the supernatant was precipitated with 1 ml of absolute ethanol at -20°C for 1 hour. The DNA pellet was collected by centrifugation in a microfuge for 10 minutes, washed once with 70 % ethanol, dried under vacuum and resuspended in 50 µl of TE buffer.

2.2.2.2 Large scale preparation of plasmid DNA

Plasmid DNA to be used in sequencing reaction, ligation and injection could be obtained in a form of pure supercoiled DNA by the use of QIAGEN plasmid kit. This rapid purification protocol is based on an alkaline lysis procedure in conjugation with separation of plasmid DNA through the QIAGEN resin, a unique surface modified silica gel that is covalently coated with a hydrophilic substance that prevents non-specific binding.

E. coli harbouring the plasmid of interest was grown at 37°C overnight in 100 ml of Luria broth containing 100 mg/ml ampicillin. Bacterial cells were harvested by centrifugation at 5,000 rpm in a Sorvall GSA rotor for 5 minutes at 4°C. The bacterial pellet was resuspended completely in 4 ml of Buffer P1. The cells were lysed and DNA denatured by addition of 4 ml of buffer P2 after which the solution was mixed gently by inverting the tube 4 - 6 times and incubated at room temperature for 5 minutes. Four millilitres of buffer P3 was added at the end of lysis reaction, gently mixed and incubated on ice for 15 minutes. Centrifugation at 15,000 rpm for 30 minutes at 4°C was performed to precipitated genomic DNA and cell debris. The supernatant was applied to a QIAGEN-tip 100 previously equilibrated with 4 ml of buffer QBT. The QIAGEN-tip was then washed with 2 x 10 ml of buffer QC. Plasmid DNA was finally eluted with 5 ml of buffer QF and precipitated with 0.7 volumes of isopropanol previously incubated at room temperature. The DNA pellet was obtained by centrifugation at 12,000 rpm for 30 minutes at 4°C, washed with 70% ethanol and dissolved in 100 µl of TE.

2.2.2.3 Quantification of nucleic acid concentration

Concentration of DNA could be roughly estimated by running in agarose gel (see method 2.2.2.5.). After ethidium bromide staining, the concentration of the DNA was

estimated by comparing the intensity of the DNA band to that of a standard DNA with known concentration.

More accurate concentration could be measured by Spectrophotometry method. Either lambda 15 UV/VIS spectrophotometer (Perkin Elmer) or Gene Quant (Pharmacia) was used. Approximately 50 μg/ml of double-stranded DNA gives 1 OD_{260}.

2.2.2.4 Restriction enzyme digestion of DNA

DNA was digested by restriction enzyme in 20 μl of appropriate 1 x restriction buffer. At least 1 unit of enzyme was used for digesting 1 μg of DNA at 37°C for 1 hr. To ensure complete digestion, excess amount of enzyme or longer incubation period was normally used.

2.2.2.5 Agarose gel electrophoresis

Electrophoresis was used for separation or identification of DNA fragments following restriction enzyme digestion. DNA sample was mixed with 1/3 volume of loading buffer (20% glycerol, 100 mM EDTA, 0.1% Bromophenol blue) before loading on 0.7 - 1.0% electrophoretic grade agarose gel, depending on the size of DNA fragments, prepared in 1 x TAE (40 mM Tris-acetate, 1 mM EDTA). Electrophoresis was performed horizontally at a potential difference of approximately 10 V/cm for 2 - 3 hours.

Following electrophoresis, the gel was stained by soaking in 0.5 mg/ml ethidium bromide solution for a few minutes. After destaining in distilled water, bands of DNA could be visualised under UV illumination and photographed.

2.2.2.6 Recovery of DNA from agarose gel

Low molecular weight DNA of less than 200 bp was recovered at high yield using MERmaid kit (STRATECH). The desired DNA band was cut from wide range agarose (sigma) after running for an appropriate length of time. The agarose was melted in 3 volumes of high salt binding solution and bound to GLASSFOG by vigorous mixing for 5-15 minutes. Binding capacity is approximately 1 mg of DNA per 1 mg of GLASSFOG, but excess amonut of GLASSFOG was normally added. GLASSFOG was pelletted down by centrifugation for a few seconds. Salts and other compounds were removed by ethanol wash before DNA eluted from GLASSFOG in small volume of TE at 45 - 55°C.

To recover higher molecular weight DNA from agarose gel, Gene Clean II kit (STRATECH) was used. The DNA was purified by the same principle as that of the MERmaid kit. A DNA band was cut from agarose gel that was made up with and run in 1xTAE buffer. The agarose was melted in 3 volumes of NaI solution at 55°C. When the agarose was completely melted, GLASSMILK suspension was added. As when purifying low molecular weight DNA with MERmaid kit, an excess amount of GLASSMLIK was added to ensure efficient binding. The binding was allowed to occur on ice for at least 5 minutes. DNA bound-GLASSMILK was pelletted down by centrifugation for a few seconds and the pellet washed 3 times with NEW WASH. The DNA was eluted from GLASSMILK in small volume of TE at 45-55°C.

For the best yield of both MERmaid and Gene Clean II kit, the manufacturer's instructions were followed.

2.2.2.7 Ligation of DNA fragment to vector DNA

Vector and DNA fragment to be ligated were digested with appropriate restriction enzyme and mixed together at 2 - 4 molar excess of the DNA fragment in 1 x ligation

buffer (10 mM Tris HCl pH7.2, 1 mM EDTA, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP) with 1 unit of T_4 DNA ligase. The reaction was made up to 20 µl and incubated at 14°C for 16-18 hours.

2.2.2.8 Random-primed labelling of DNA (Feinberg and Vogelstein, 1983)

Random-primed labelling is based upon the random annealing of oligodeoxynucleotide mixture that can serve as a primer to single stranded template and the activity of Klenow fragment of *E. coli* DNA polymerase.

A hundred nanogram of linear DNA was heat-denatured by boiling in a total volume of 31.5 µl. After cooling on ice for 1 min, the denatured DNA was mixed with 12.5 µl of 4 x oligo labelling buffer (200 mM Tris.HCl pH8.0, 20 mM $MgCl_2$, 20 mM DTT, 800 mM HEPES pH6.6, 0.027 unit of Upd$(N)_6$, 80 mM dATP, dGTP, dTTP), 10 mg of BSA, 30 µCi of α-^{32}P-dCTP and 5 units of Klenow enzyme in a final volume of 50 µl. The reaction was incubated at room temperature for at least 2 hours or could be left overnight.

Alternatively, the DNA probe was radio-labelled using Oligolabelling kit (Pharmacia). A hundred nanogram of DNA fragment was made up to a volume of 36 µl with water and heat-denatured for 5 minutes. Ten microlitres of reagent mix, 30 µl of α-^{32}P-dCTP and 5 units of Klenow polymerase enzyme were added to the denatured DNA after cooling on ice. The reaction was proceeded at 37°C for 2 hours.

The radio-labelled DNA was separated from unincorporated nucleotides by passing through Nuctrap Push Column (STRATAGENE). The column was prewetted with 70µl of 1 x STE (100 mM NaCl, 20 mM Tris HCl pH7.5 and 10 mM EDTA) before DNA sample (made up to a volume of 70 µl) was applied. The labelled DNA was then eluted from the column by another 70 µl of 1 x STE. For safety, the manufacturer's instruction was strictly followed.

2.2.2.9 Screening of bacteria containing recombinant plasmids by colony hybridisation

Bacterial colonies were transferred to Hybond-N nylon membrane (Amersham) by placing the membrane, cut to fit the size of the petri-dish, onto the surface of agar without any air bubbles trapped. The cells were lysed and the DNA was denatured by soaking the membrane in denaturing solution (0.5 N NaOH, 1.5 M NaCl) for 2 minutes. The membrane was then transferred to neutralising solution (1.5 M NaCl, 0.5 M Tris.HCl pH7.2, 0.001 M EDTA) for 2 minutes and lastly placed in 2 x SSC for 3 minutes. After air-drying, the DNA was fixed by baking the membrane at 80°C for 2 hours. Prehybridisation was performed in hybridising solution (0.5 M $NaPO_4$ pH7.2, 7% SDS, 1 mM EDTA) at 55°C for at least 15 minutes in the HYBRIDISER HB-1 oven (Techne). Heat-denatured ^{32}P-labelled DNA probe was added and hybridisation was allowed to occur at 55°C for 12 - 16 hours. The excess probe was removed and the membranes washed in 40 mM $NaPO_4$, 1 mM EDTA and 1% SDS at 50°C three times, 20 minutes each. Positive colonies were investigated by autoradiography (see 2.2.2.13).

2.2.2.10 DNA detection by Southern blot hybridisation

Restriction enzyme digested DNA fragments were separated on 1% agarose gel (see 2.2.2.5) and double stranded DNA was denatured by soaking the gel in 0.4 N NaOH-0.6 M NaCl with gentle agitation for 30 minutes. The gel was then incubated in 1.5 M NaCl-0.5 M Tris.HCl pH7.5 for 30 minutes. The DNA was transferred to GeneScreen Plus™ hybridzation transfer membrane (Dupont) by capillary action in 10xSSC (1.5M NaCl, 0.15M NaCitrate). Capillary transfer was set up as described in manufacturer's instruction. The transfer was allowed to continue for 16 - 24 hours. After transferring process the membrane was rinsed briefly in 0.4 N NaOH to ensure the denaturation, then soked in 0.2 M Tris.HCl pH7.5-2xSSC for 1 minute and air-dried.

The membrane was prehybridised in prehybridisation solution (1% SDS, 1M NaCl and 10% Dextran sulfate) for at least 15 minutes. The denatured radioactive DNA probe was then added to the solution along with denatured salmon sperm DNA (100 µg/ml). Hybridisation was done at 65°C for 12-16 hours. The solution and excess probe were removed and the membrane washed in 2x100ml of 2xSSC at room temperature for 5 minutes then in 2x200 ml of a solution containing 2xSSC and 1% SDS at 65°C for 30 minutes. Finally the membrane was washed in 2x100 ml of 0.1xSSC at room temperatur for 30 minutes. The membrane was placed, with the DNA face up, on a sheet of filter paper and allowed to dry at room temperature. Autoradiography was done as deecribed in 2.2.2.13.

2.2.2.11 Polymerase chain reaction

To amplify specific DNA fragments, primers were designed such that they shared at least 20 nucleotides of homology to the template and had appropriate restriction enzyme recognition sites at their 5' ends. The reactions were set up in 0.5 ml Eppendorf tubes. One hundred nanograms of template DNA was mixed with 10 µl of 10xTaq polymerase buffer, 20 pmole of each primer, 50 mM of each dNTP and 5 unit of Taq polymerase (Boehringer Mannheim). The final volume of 100 µl was made up with water. The reaction mixture was submerged under 50 µl of mineral oil (Sigma) to prevent evaporation during the reaction.

The first three cycles of the reaction started with denaturation at 93°C for 5 minutes, annealing at 40°C for 2 minutes and polymerisation at 70°C for 1 minute. A further 27 cycles of 93°C for 1.5 minutes, 40°C for 1.5 minutes and 70°C for 1 minute were continued to complete the amplification. The PCR product was checked on 2-3% low melting temperature agarose gel electrophoresis. To clean up the PCR product, the band containing the product was cut from the gel and purified using Gene Clean II or MERmaid kit (STRATECH).

2.2.2.12 DNA sequencing (Dideoxynucleotide method)

Sequencing of double stranded plasmid DNA was carried out using Sequenase™ version 2.0 kit (United States Biochemicals). Three micrograms of QIAGEN purified plasmid DNA was denatured in 0.2 N NaOH, 0.2 mM EDTA at 37°C for 30 minutes. The DNA was then ethanol precipitated and dissolved in 7 µl of dH_2O before 2µl of 5 x reaction buffer and 1 µl of primer (20 ng/µl) were added. Annealing of primer to template was allowed to occur at 65°C for 2 minutes followed by slow cooling to 37°C over 30 minutes. The labelling reaction was performed by adding to the annealed DNA mixture 1 µl of 0.1 M DTT, 2 µl of 1:4 dilution of labelling mix (0.75 mM of dGTP, dCTP and dTTP), 0.5 µl of ^{35}S-dATP (10 mCi/ml) and 2 µl of diluted sequenase polymerase (1:8 dilution in 10 mM Tris.HCl pH7.5, 5 mM DTT, 0.5 mg/ml BSA). The reaction was allowed to proceed at room temperature for 5 minutes. The reaction was terminated by transferring 3.5 µl of labelling reaction to four Eppendorf tubes, each containing 2.5 µl of termination mixture of ddGTP, ddATP, ddTTP and ddCTP that had been preheated at 37°C. The mixture was incubated at 37°C for 5 minutes after which 4 µl of stop solution (95% formamide, 20 mM EDTA, 0.05% Bromophenol blue, 0.05% Xylene cyanol FF) was added to each of the termination reaction.

Before loading sequencing gel, samples were boiled up for 2 minutes. Then, 2 - 3 µl of each sample was immediately loaded in each lane. A 6% denaturing polyacrylamide gel of 0.5 - 2.5% TBE gradient was made up by mixing 7 ml of 6% acrylamide in 2.5% TBE (15 ml of 40% ACCUGEL, 46 g urea and 5 ml of 10xTBE for 100 ml solution) to 35 ml of 6% acrylamide in 0.5% TBE (15 ml of 40% ACCUGEL, 46 g urea and 25 ml of 10xTBE for 100 ml solution).

Electrophoresis was carried out in 1xTBE buffer at 35 watts for 2 - 3 hours using SEQUI-GEN® nucleic acid sequencing gel (BIO-RAD). After electrophoresis the gel was fixed in 10% methanol (v/v) and 10% acetic acid (v/v) for 20 minutes before

drying under vacuum at 80°C for 60 - 90 minutes in gel dryer model 583 (Biorad). Autoradiography of the dried gel was done as described in 2.2.2.13.

2.2.2.13 Autoradiography

Autoradiography was carried out by exposing hybridised membrane to a pre-flashed X-Ray film (Kodak) in an autoradiograph cassette. The film was exposed at -70°C for overnight or longer and was developed with Automatic X-Ray film developer (X-OGRAPH COMPACT X-2).

The exposure was carried out at room temperature without pre-flashing for sequencing gel.

2.2.2.14 Preparation of *Drosophila* genomic DNA

Twenty flies were homogenised on ice in 0.4 ml of homogenisation buffer (8 M urea, 0.35 M NaCl, 10 mM Tris.HCl pH8.3, 10 mM EDTA) using a motor homogeniser (Kontes). The homogenate was extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) by gently mixing for 5 minutes. After centrifugation to remove cell debris, the aqueous phase was transferred to a new tube and extracted one more time with phenol:chloroform:isoamyl alcohol followed by another extraction with chloroform: isoamyl alcohol. The nucleic acid was ethanol precipitated and dissolved in 50 µl of TE buffer containing 20 µg/ml of DNase-free RNase and stored at 4°C.

2.2.3 Protein preparation and manipulation techniques

2.2.3.1 Preparation of CAT extracts from *Drosophila* tissues

A hundred 4 -5 day old male and female flies, a hundred pairs of ovary and a hundred female carcasses were used for preparation of CAT extracts. Flies, ovaries and

carcasses could be kept frozen at -70°C and allowed to thaw before use. To the thawed tissues, 500 µl of 0.25 M Tris.HCl pH7.8 was added and tissues homogenised with a motor homogeniser. The homogenate was frozen in liquid nitrogen and thawed in a 37°C water bath five times to disrupt cell membranes. The homogenate was then heated to 65°C for 5 minutes to inactivate proteins that would interfere with the CAT assay. The extract was then spun in a microfuge to pellet cell debris and some denatured proteins. The supernatant was kept as CAT extract in small aliquot at -70°C.

2.2.3.2 Determination of protein concentration in *Drosophila* extracts

Protein concentration was estimated using Bradford method. The method utilises the protein binding property of Coomassie blue G-250. Once the binding occurs, an absorbance shifts from 465 to 595 nm. BSA (1 - 15 mg/ml in 0.00125 M Tris.HCl pH7.8) was used to set a standard graph that was prepared each time the protein concentration was measured as the Bradford assay is sensitive to fluctuations in temperature. Five microlitres of protein extract was made up to 1 ml with water and 1 ml of Coomassie protein assay reagent (Pierce) added and mixed. The OD_{595} was measured and the concentration of protein estimated from standard graph of BSA.

2.2.3.3 Assay for CAT activity in *Drosophila* extracts

One microgram of extract was mixed with 2 µl of 25 mM acetyl coenzyme A , lithium salt (Sigma), 0.5 µl of D-threo-[dichloroacetyl-1-^{14}C] chloramphenicol (227.27 pmole). A final volume of 40 µl was made up with water and the reaction incubated at 37°C for 1 hour. The reaction was stopped and radioactivity extracted by adding 200 µl of ethyl acetate and vortexing for 1 minute. Organic phase was separated by centrifugation in a microfuge for 2 minutes and was transferred to a new tube. Ethyl acetate was allowed to evaporate overnight on bench. Twenty microlitres of fresh ethyl acetate was then added to the tube and vortexed. The acetylated (1-acetate- and 3-

acetate-) and unacetylated forms of chloramphenicol were separated by silica gel TLC on Merck's TLC silica matrix. TLC was performed in a sealed tank using 95% chloroform (v/v), 5% methanol (v/v) as the solvent (Cohen *et al.*, 1980) until the solvent front reached within 1 cm of the top edge of the matrix. The matrix was removed from the tank and dried before exposing to Phosphor screen overnight for quantitation using a Molecular Dynamic PhosphorImager. The CAT activity was calculated by converting the proportion of acetylated bands into a unit of pmoles of ^{14}C-chloramphenicol acetylated /min/mg of protein.

2.2.3.4 β-galactosidase activity assay

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

2.2.3.5 Preparation of nuclear extracts from ovaries and carcass

Ovaries were dissected from 3-4 day old flies in Ringers solution. About 500 - 600 flies were needed for a total volume of 600 µl of ovaries. If this amount of ovaries could not be obtained within one day the ovaries were rinsed twice with MTBS and stored overnight on ice in MTBS. Dissection was continued on the next day until the required amount obtained. Egg chambers were separated by gently pipetting up and down using a cut off 1 ml pipette tip so that the oocytes should not be broken. Egg chambers were collected by centrifugation at 2,200 rpm for 7 minutes at 4°C in BIOFUGE 22R (HERAEUS Sepatech) using HFA 17.1 rotor. Supernatant was

removed and egg chambers washed in MTBS. The centrifugation was repeated to collect mostly intact egg chambers that were next resuspended in 2 volumes of solution I. The egg chambers were homogenised thoroughly with motorised homogeniser. The homogenate was centrifuged at 4,000 rpm for 4 minutes at 4°C. The pellet was resuspended in an equal volume of solution II. Then 5M NaCl was added to a final concentration of 0.35 M and incubated on ice for 30 minutes. Nuclear extract was obtained by high speed centrifugation at 17,000 rpm overnight at 4°C. To the supernatant, glycerol was added to a final concentration of 20% and the extract stored in 20 µl aliquots at -70°C. The protein concentration of the extract was determined by Bradford method (see 2.2.3.2).

Extracts of the whole flies and carcass tissues were prepared in the same way as described for ovaries.

2.2.3.6 End-labelling of DNA fragment to be used as a probe in gel retardation assay

The 3' ends of purified DNA fragment were labelled with α-^{32}P-dCTP by the activity of Klenow polymerase enzyme. About 200 ng of the purified DNA fragment was mixed with 6 µl of 5xbuffer (0.25 M Tris pH7.2, 50 mM $MgSO_4$, 0.5 mM DTT made up in TE), 3 µl of 20 mM dGTP, dATP and dTTP, 2 µl (20µCi) of α-^{32}P-dCTP, 5 units of Klenow enzyme and made up to a final volume of 30 µl with water. The reaction was incubated at room temperature for 30 minutes before 5 µl of 5 mM dCTP was added and the incubation continued for another 30 minutes at room temperature. The labelled DNA was then separated from unincorporated DNA by Nuctrap Push Column as described in 2.2.2.8.

2.2.3.7 Gel retardation assay

Three micrograms of *Drosophila* nuclear extract was incubated with 1 µg of non-specific competitor DNA (poly[dI-dC], Boehringer Mannheim) on ice in a 10 µl

reaction volume containing 25 mM HEPES pH7.6, 40 mM KCl, 2 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT and 10% glycerol for 10 minutes. Two nanograms of end-labelled DNA fragment was then added to the reaction and the incubation continued on ice for a further 20 minutes. In competition assays, a 10-fold molar excess of unlabelled DNA fragment was added to the reaction 10 minutes before the labelled probe. The DNA-protein complexes were resolved in 5% polyacrylamide gel (8.4 ml of 30% acrylamide, 2.5 ml of 20xTBE, 38.9 ml of water, 200 µl of 25% ammonium persulfate and 50 µl of TEMED) in 1xTBE at 200 V until the blue dye reaches about 2 cm from the bottom of the gel. The gel was then dried under vacuum and autoradiographed.

2.2.3.8 DNase I footprinting

The method described by Leblance and Moss (1994) was followed. The binding reaction was performed as described in 2.2.3.7. After the incubation, binding buffer was added to make a total volume of 50 µl. The reaction tube was transferred to room temperature and 50 µl of cofactor solution (10 mM $MgCl_2$, 5 mM $CaCl_2$) was added. Five microlitres of the appropriate dilution of DNase I (0.003 U/µl for free DNA and 0.04 U/µl for DNA-protein complex) was then added. After 2 minutes of digestion, the reaction was stopped by the addition of 100 µl of stop solution (1% SDS, 200 mM NaCl, 20 mM EDTA, pH8.0, 40 µg/ml tRNA). The reaction was then extracted with an equal volume of phenol:chloroform (1:1) and centrifuged in a microcentrifuge for 10 minutes. The upper phase was recovered and precipitated with 2 volumes of ethanol at -70°C for 20 minutes. The DNA was pelleted down by centrifugation for 15 minutes and the DNA pellet washed with 70% ethanol. After centrifugation for 5 minutes and the supernatant removed, the pellet was dried in a vacuum desicator and resuspended in 5 µl of sequencing loading buffer. The DNA was denatured by boiling for 2 minutes and the digestion fragments separated on 6% acrylamide sequencing gel. The G+A sequencing ladder was run in parallel with the sample as a marker to localise the position of the footprints.

2.2.3.9 Preparation of the G+A sequencing ladder

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

About 5,000 cpm of the end labelled DNA fragment was mixed with 1.5 µl of 1 mg/ml salmon sperm DNA in a total volume of 10 µl. The mixture was chilled on ice and 1.5 µl of formic acid added. The reaction was incubated at 37°C for 14 minutes and chilled again on ice prior to the addition of 150 µl of freshly prepared 1 M piperidine. The cap of the tube was wrapped tighly with a parafilm then the reaction was heated at 90°C for 30 minutes. After incubation, the tube was cooled on ice and spun for a few second in a microcentrifuge. The reaction was transferred to a new tube and 1 ml of 1-butanol added. The tube was vortexed until only one phase was obtained and then centrifuged at 12,000 rpm for 2 minutes. The supernatant was carefully removed and 150 ml of 1% SDS added to the pellet then vortexed. The DNA was precipitate again in 1-butanol. After the supernatant has been removed, the pellet was dried in a vacuum desicator for 30 minutes and resuspend in 5 µl of sequencing loading buffer.

2.2.4 Manipulations of *Drosophila* stocks

2.2.4.1 Collection of eggs for microinjection

Egg collection chambers with a sliding food drawer were used for collecting eggs. The food drawer was filled with 1.5% apple agar and placed in the chamber to which hundreds of 4 - 5 day old W^K flies (Luning 1981) were transferred. The flies were kept and allowed to lay eggs in a chamber at 18°C.

2.2.4.2 *P* element mediated transformation

The DNA used for transformation was prepared by QIAGEN (2.2.2.2). The plasmid carrying the transformation vector with the gene of interest was mixed with helper plasmid (phs70Δ2-3wc) at a ratio of 5:1. The DNA mixture was ethanol precipitated and dissolved in injection buffer (5 mM KCl, 0.1 mM $NaPO_4$ pH6.8) to a final concentration of 500 mg/ml of the vector plasmid and 100 mg/ml of the helper plasmid. The DNA was spun for 2 minutes in a microfuge to pellet any debris, and sucked into injection needle. It was then stored under halocarbon oil to prevent drying.

Eggs were collected every 30 minutes from egg collection chambers and were dechorionated manually by gently rolling across double sided sticky tape (Scotch brand 3M No. 066) on a microscope slide. The naked eggs were then lined up on a narrow strip of sticky tape with their posterior ends pointing in the same direction. Embryos were desiccated in a petri dish containing silica gel for 5 minutes, or other appropriate length of time depending on humidity and temperature of the room, after which the embryos were covered with halocarbon oil. DNA was injected gently into the posterior pole of preblastoderm eggs. Older eggs were killed by tearing the vitelline membrane. Injected eggs were incubated at 18°C for 24 - 36 hours in a petri dish with pieces of wet filter paper. Hatched larvae were removed from the oil with the tip of forceps and transferred to fresh food. The vial containing larvae was incubated at 25°C until pupae developed. Each pupa was then separated into an individual vial and allowed to further develop at 25°C.

2.2.4.3 Establishment of homozygous transgenic lines

Indivual eclosed adult was crossed with W^K virgins of the opposite sex and their progeny were examine for a red eye phenotype. Transformants were then crossed again with W^K virgins. Homozygous lines were established by sibling mating.

2.2.5 Tissue culture techniques

2.2.5.1 Growth and maintenance of *Drosophila* Schneider Line 2 cells

Cells were maintained in 1xSchneider's *Droaophila* medium supplemented with 5% foetal calf serum, 100 u/ml penicillin and 100 mg/ml streptomycin at a density around 2-4x10^6 cells/ml. Cells were incubated at 23°C and 50% of cell volume was replaced by fresh medium every 7 days.

2.2.5.2 Preparation of nuclear extract from Schneider Line 2 cells

Ten millilitres of cells (2x10^7-8x10^7 cells) were collected by spining at 1,000 rpm for 5 minutes at 4°C. The cells were then washed in 5 ml of MTBS and re-spun. Two volumes of solution I were added to the cells and the next steps were followed as described in 2.2.3.5.

CHAPTER 3

The role of the first four nucleotides of the *I* factor promoter in its expression in *Drosophila*

3.1 INTRODUCTION

RNA polymerase II promoter

RNA polymerase II is one of the three forms of nuclear DNA dependent RNA polymerases that transcribe different sets of genes in eukaryotic system. RNA polymerase II is required for transcription of protein-coding genes and some small nuclear RNAs (SnRNAs). Transcription initiation by RNA polymerase II requires several *cis* and *trans* acting elements.

The most well characterised class of protein-coding genes that are transcribed by RNA polymerase II contains two transcription control elements that form a minimal promoter. The first element, TATA box, is located 20-30 nucleotides upstream of the mRNA start point (position +1) with the consensus sequence TATA(T/A)A(T/A) (Corden, *et al.* 1980). The second motif stretching across the transcription initiation site has been identified in several mammalian TATA-containing genes. The transcription begins at the A in a weak consensus 5'-PyPyCAPyPyPyPy-3' (Corden, *et al.* 1980).

These two transcription control elements appear to be recognised by a complex set of transcription factors. Among these factors, TFIID seems to be the only one that has specific DNA binding activity to the TATA motif (Nakajima, *et al.* 1988, VanDyke, *et al.* 1988). TFIID itself is a complex formed between TATA-binding protein (TBP) and TBP associated factors (TAFs). TBP alone is sufficient for TATA element recognition and subsequent association of the other basal factors for basal transcription such as TFIIB, -IIE, -IIF, and -IIH. TAFs are required as co-activators in the transcriptions that are regulated by upstream factors (Buratowski 1994). The binding of TFIID to the

TATA motif is stimulated by TFIIA (Maldonado, *et al.* 1990). The other transcription factors as well as RNA polymerase II associate with the promoter sequence and the TATA-TFIID complex by protein-protein interactions. A model for the assembly of the RNA polymerase II transcription initiation complex has been recently proposed based on kinetic assays, native gel electrophoresis and nuclease protection assays (Buratowski 1994).

RNA polymerase II also initiates transcription of the genes that do not contain TATA sequences. The promoters of these genes can be divided into two classes. One is known as GC-rich promoters primarily found in housekeeping genes (Sehgal, *et al.* 1988). The other includes the promoters that have neither TATA element nor GC-rich sequences.

The GC-rich promoters initiate transcription at heterogeneous sites but usually have one or two major start sites (Crouse, *et al.* 1985, Melton, *et al.* 1984). The promoters in this class also have several binding sites for the transcription factor Sp1 which was identified as a mammalian host protein required for SV40 gene expression. The binding of Sp1 to the promoter and its role in transcription activation has been shown *in vitro* for mouse housekeeping gene dihydrofolate reductase (Dynan, *et al.* 1986). Basal transcription factors that are involved in transcription initiation of TATA-containing promoters such as TFIID, -IIB, -IIE and -IIF are also required for transcription activation of TATA-less, GC rich promoters by Sp1 (Pugh and Tjian 1991, Smale, *et al.* 1990).

Many TATA-less promoters that are not GC rich are regulated during differentiation or development. Examples of the genes that are expressed from the promoter in this class are *Drosophila* homeotic genes such as *Ultrabithorax* (Biggin and Tjian 1988) and *Antennapedia* (Perkins, *et al.* 1988) and genes that are regulated during mammalian immunodifferentiation such as *terminal deoxynucleotidyl transferase* (TdT) gene (Smale and Baltimore 1989). Transcription from TATA-less, non GC rich promoters

is initiated at one or a few clustered sites. Accurate basal transcription initiation *in vitro* requires sequences at and/or downstream of the start site. Smale and Baltimore (1989) have identified a 17 bp motif of TdT gene that is sufficient for accurate basal transcription. This motif, so called an initiator (Inr), contains within itself the transcription start site. The sequence CTCATTCT included in this 17 mer is essential for functional initiator and the transcription starts at the A (Smale, *et al.* 1990). The sequence of this motif matches a consensus TCA(G/T)T present at or near the transcription start site of many *Drosophila* retrotransposon and some TATA-less genes (Arkhipova and Ilyin 1991, Arkhipova and Ilyin 1992) and a weak consensus 5'-PyPyCAPyPyPyPyPy-3' surrounding the start site of several mammalian TATA-containing genes (Corden, *et al.* 1980). Mutagenesis of the initiator altered both the efficiency and accuracy of initiation (Smale and Baltimore 1989).

The TATA-less, non GC-rich promoters have been identified in some retrovirus-like transposons (Arkhipova and Ilyin 1991, Jarrel and Meselson 1991) and LINE-like transposons (Minchiotti and Di Nocera 1991a, Mizrokhi, *et al.* 1988). The *I* factor promoter also falls into this class (McLean, *et al.* 1993). The four nucleotides at the transcription start site of the *I* factor, CA(G/T)T, are similar to the core consensus TCANT (where N is G or T) found in many other TATA-less promoters (Arkhipova and Ilyin 1991). The CA(G/T)T motif of the *I* factor has been shown to be important for transcription process in *Drosophila* SL_2 tissue culture cells as mutations of some nucleotides in this motif reduced the activity several fold and replacement of this motif with the sequence ACCG abolished the activity (McLean, *et al.* 1993).

The aim of this chapter is to investigate whether the CA(G/T)T motif at the beginning of the *I* factor sequence is required for promoter activity in *Drosophila melanogaster* flies. This would provide an information on the role of this motif in controlling transposition of the *I* factor during hybrid dysgenesis. The 186 bp *I* promoter carrying ACCG at the first four positions fused to the CAT gene was introduced into reactive

flies by *P*-mediated transformation and the expression of the CAT gene was determined.

3.2 RESULTS

3.2.1 Construction of plasmid p186Δ4.2W8

Plasmid p186Δ4.2 was constructed such that the expression of the CAT gene is under control of a 186 bp fragment of the *I* factor promoter in which the first four nucleotides CATT were replaced by ACCG (McLean, *et al.* 1993). The *PstI* fragment of p186Δ4.2 containing the mutated *I* promoter, CAT gene and SV40 polyadenylation sequences was inserted into the *PstI* site of *P*-mediated transformation vector, pW8 (Thummel, *et al.* 1988) in an orientation such that the direction of the transcription of *I* is opposite to that of the white (*w*) selectable marker gene and *P* promoter (see figure 3.1 for the diagram of the construction). The orientation of the *I* promoter in the construct was checked by restriction mapping. The resulting construct, named p186Δ4.2W8 was co-injected with the helper plasmid phs70Δ2-3wc into embryos of the reactive strain W^K. The helper plasmid contains a *P* element that lacks IVS3 and thus provides transposase enzyme that is required for transposition of the *P*-mediated transformation vector into the host chromosomes. The reactive strain was chosen to be the host for transformation as it should provide the incoming *I* promoter with a permissive environment similar to that found in dysgenic flies.

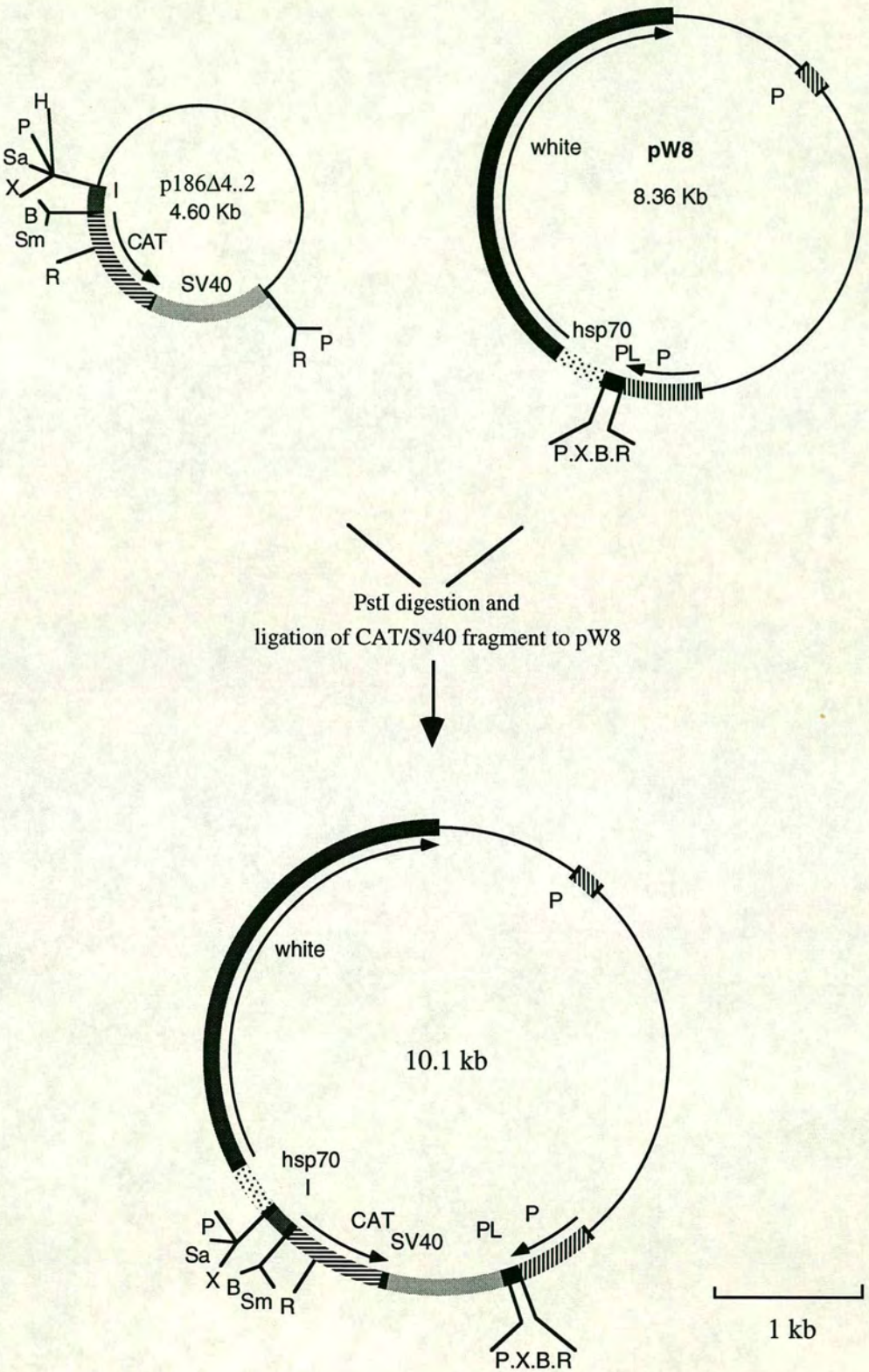
3.2.2 Establishment of lines transformed with p186Δ4.2W8

Only two homozygous lines, 227 and 228, were established by transformation of W^K flies with p186Δ4.2W8. Genomic DNA of flies from these two transformed lines was prepared and was digested with *Hind*III that does not cut within the inserted fragment. The DNA fragments were separated on a 1% agarose gel, Southern transferred to nylon membrane and then probed with the 250 bp *Bam*HI-*Eco*RI fragment of the CAT

Figure 3.1 Construction of plasmid p186Δ4.2W8 for *P*-mediated transformation

The *Pst*I fragment of plasmid p186Δ4.2 containing the *I* promoter that begins with ACCG fused to CAT/SV40 sequence was inserted to pW8 at *Pst*I site. The arrows indicate the direction of transcription. Only some restriction sties in the pw8 polylinker are shown - see Klemenz *et al* ,1987 for complete restriction map of the polylinker.

Abbreviations for restriction sites: B - *Bam*HI, H - *Hind*III, P - *Pst*I, R - *Eco*RI, Sa - *Sal*I, Sm - *Sma*I, X - *Xba*I


H
P
Sa
X
B
Sm
R
I
p186Δ4..2
4.60 Kb
CAT
SV40
P
R
white
pW8
8.36 Kb
P
hsp70
PL
P
P.X.B.R
PstI digestion and
ligation of CAT/Sv40 fragment to pW8
white
P
10.1 kb
hsp70
I
CAT
SV40
PL
P
P
Sa
X
B
Sm
R
P.X.B.R
1 kb

p186Δ4.2W8

gene from pCAT.1 (McLean, *et al.* 1993). This experiment was performed in order to confirm that each line actually contains the fragment of interest and to establish the copy numbers of the transgene in each line. Figure 3.2 illustrates Southern blot hybridisation showing that each line contains one copy of the transgene. The transgenes present in each line are the result of independent insertion events as they were detected as the bands of different sizes.

3.2.3 CAT assays in transformed lines

The activity of the *I* promoter fragment in p186Δ4.2W8 construct was determined by its ability to direct the expression of the CAT gene. The CAT extracts were prepared from tissues of each transformed line as described in 2.2.3.1. Flies from line 186.137 containing the full length 186 bp of *I* promoter fragment immediately upstream of the CAT gene were used as a control. The activity of CAT was measured as the amount of D-thero-[dichloroacetyl-1-^{14}C] chloramphenicol that has been acetylated by one milligram of CAT extract in one minute.

Figure 3.3 a and b show CAT assays of males, females, ovaries and female carcass of a line transformed with p186W8 (McLean, *et al.* 1993) carrying the full-length 186 bp *I* promoter fragment linked to the CAT gene, line 186.137, and lines transformed with p186Δ4.2W8, lines 227 and 228. Line 186.137 was chosen as a control for the activity of the full-length promoter because it gave CAT values that are closest to the mean values of the five lines transformed with p186W8 (McLean, *et al.* 1993). The CAT activities in all tissues tested of lines 227 and 228 were reduced 10-40 fold relative to that of line 186.137 (table 3.1 a). This result shows that the *I* factor promoter in which the first four nucleotides CA(G/T)T were replaced with ACCG lost the ability to express the CAT gene and suggests that the CA(G/T)T motif at the first position of the *I* factor is required for promoter activity of the element in flies.

Figure 3.2 Southern blot analysis of lines transformed with p186Δ4.2W8

Approximately 3 μg of genomic DNA from the two independent transformed lines were digested with *Hind*III. After electrophoresis and Southern transfer, the DNA was hybridised to 250 bp *Bam*HI/*Eco*RI fragment of the CAT gene in pCAT.1. Genomic DNA of W^{K} (Luning 1981) and line 186.137 (McLean, *et al.* 1993) were used as negative and positive controls respectively.

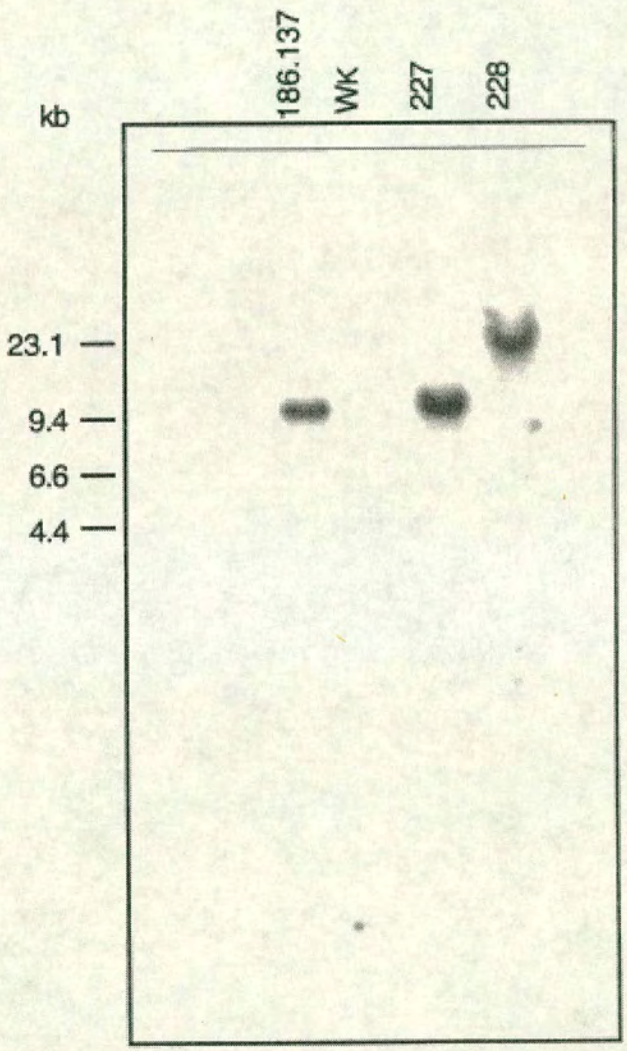

kb
186.137
WK
227
228
23.1
9.4
6.6
4.4

Figure 3.3 a CAT assays of the tissues from line transformed with p186W8

Transformant line 186.137 was used as a control for CAT assays. The assays were carried out with one microgram of each extract.

M - 3-4 day old males, F - 3-4 day old females, ov - ovaries, car - carcass

Figure 3.3 b CAT assays of the tissues from lines transformed with p186Δ4.2W8

CAT assays for lines 227 and 228 are illustrated. One microgram of each extract was used in the assays.

M - 3-4 day old males, F - 3-4 day old females, ov - ovaries, car - carcass

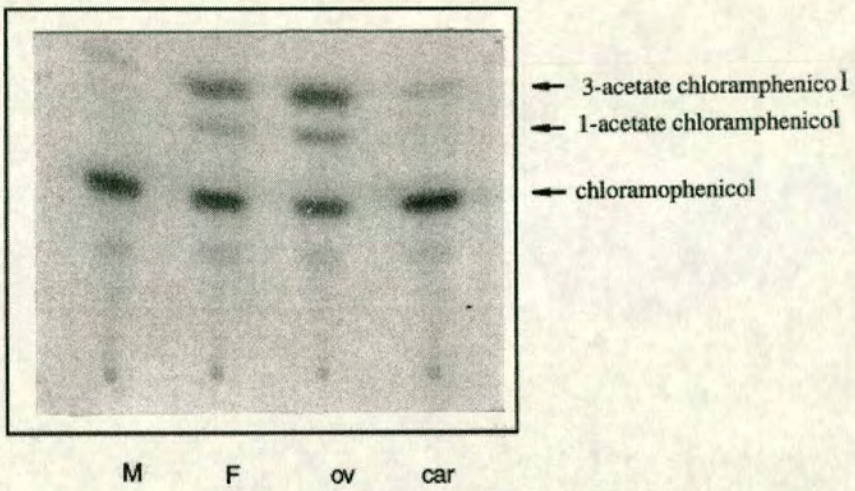


Figure 3.3 a

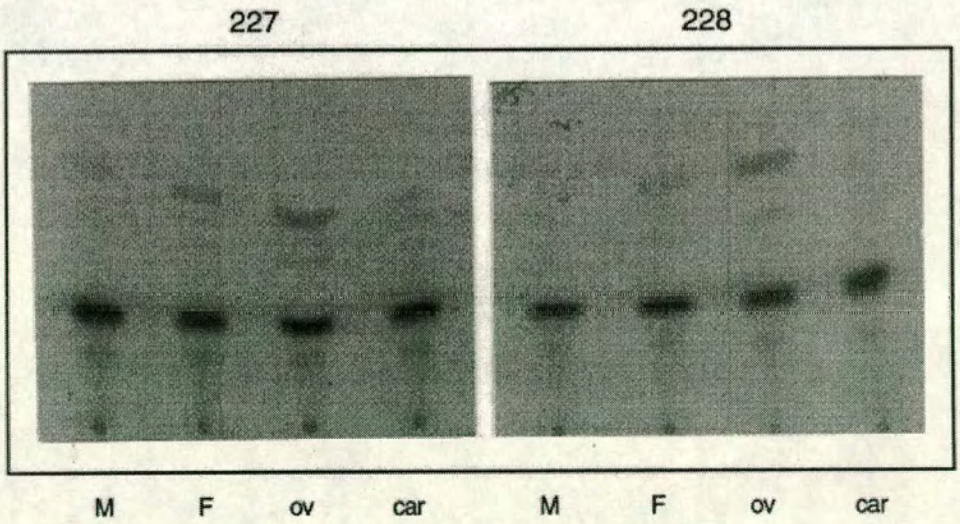


Figure 3.3 b

Table 3.1 a The ratio of CAT activities of line transformed with p186W8 (186.137) and lines transformed with p186Δ4.2W8 (227 and 228)

	MALE	FEMALE	OVARY	CARCASS
186.137/227	13.3	28.2	43	10.7
186.137/228	13.3	22.9	20	15

Table 3.1 b CAT activities of the tissues from line transformed with p186W8 (186.137) and lines transformed with p186Δ4.2W8 (227 and 228)

LINE	MALE	FEMALE	OVARY	CARCASS	O/C
186.137	24±2.6	158±5.4	374±11.2	15±1.2	25±3.4
227	1.8±0.23	5.6±0.35	8.6±0.46	1.4±0.38	6.1±0.28
228	1.7±0.20	6.9±0.52	18.6±2.3	1.0±0.22	18.6±2.3

Values of CAT activity are expressed in pmoles ^{14}C-chloramphenicol acetylated/min/mg protein. The mean values and standard errors of three individual assays on same extract of each line are shown.

The wild type *I* promoter was shown to enhance expression of the CAT gene in the ovaries at the level around 20 fold higher than the activity expressed in the female carcass (McLean, *et al.* 1993). Despite the low levels of activity directed from the promoter with ACCG as the first four nucleotides, the enhancement of the activity in the ovaries is still notable in the transformed lines 227 and 228 containing such a promoter (table 3.1 b). This indicates another sequence in the first 186 nucleotides of the *I* factor that is responsible for the enhanced activity in the ovaries.

3.3 DISCUSSION

The CA(G/T)T motif at the transcription initiation site of the *I* factor has been demonstrated to be important for transcription from the *I* promoter in *Drosophila melanogaster* flies. Similar motifs, so called initiators, encompassing the transcription start site have been found in many promoters, both TATA-containing and TATA-less (Corden, *et al.* 1980, Weis and Reinberg 1992). The initiator is a discrete promoter element that, on its own, can direct accurate transcription of a TATA-less promoter (Smale and Baltimore 1989).

TATA-mediated and initiator-mediated transcription are different in some aspects of regulation. For example, TATA dependent transcription initiation is inhibited by a tumour suppressor p53 whereas transcription directed by an initiator element is resistant to the effect of p53 (Mack, *et al.* 1993). The mechanism of preinitiation complex formation is similar for both TATA and TATA-less promoters although a difference in an initial recognition step has been demonstrated (Zenzie-Gregory, *et al.* 1992). In the case of TATA-containing promoter, TBP firstly binds the TATA motif providing the nucleation site for RNA Pol II and the remaining general transcription factors. In the absence of a TATA element, the initiator plays an essential role in complex formation. RNA Pol II, though with low affinity, recognises the initiator and forms a specific transcription complex. The binding of RNA Pol II to the initiator is

stabilised, via protein-protein interactions, by general transcription factors (Weis and Reinberg 1992).

Factors other than TBP and those necessary for transcription from a TATA-containing promoter are required for transcription from a TATA-less promoter. Transcriptions from the TATA-less promoters of some genes such as human porphobilinogen deaminase (PBGD) and adeno-associated virus p5 have been shown to be independent of TFIID but may require a protein that can function like TBP in the recruitment of an initiation complex (Beaupain, *et al.* 1990, Usheva and Shenk 1994). Proteins that bind the initiator element and play an essential role in conducting accurate transcription have recently been identified in many TATA-less promoters (reviewed in Weis and Reinberg, 1992). Although recognition of the initiator by a protein factor has been suggested to be the first step in a proposed mechanism for transcription initiation of the *I* factor promoter (C. McLean, personal communication), no attempt has been made so far to identify such a protein.

In addition to the similarity in having an initiator with the consensus TCA(G/T)T, many TATA-less promoters, particularly those of LINE-like retrotransposons and some developmentally regulated genes, also share another conserved sequence around position +30 (Arkhipova and Ilyin 1991, McLean, *et al.* 1993). A sequence AGTCGTGCCT that matches the consensus AGTCGTGPyPy of the +30 motif is also present within the *I* factor promoter. Despite the well conserved sequence that suggests it to be functionally important, the +30 motif does not seem to be essential for the activity of the *I* promoter. Removal of the sequence downstream of the position +28 that also remove this motif from the *I* promoter only reduced the activity about 2 to 3 fold in transient expression assay in *Drosophila* SL_2 cells and did not have a marked influence on the activity in flies comparing with the activity directed from nucleotides 1-40 that contain this motif (C. McLean, personal communication).

As transcription from nucleotides 1-40 of the *I* factor has been shown to initiate accurately at position +1 (McLean, *et al.* 1993), the role of the +30 motif could be to mediate start site selection though no such a function has been reported for this conserved motif in the other promoters. Mapping the 5' end of the transcript that is initiated from nucleotides 1-28 which do not contain the +30 motif would be able to address whether or not the +30 motif plays a role in confining the transcription initiation site of the *I* factor.

CHAPTER 4

Quantitative assays of enhacer activity of nucleotides 41-186 of the *I* factor

4.1 INTRODUCTION

Enhancers

The expression of protein-coding genes in eukaryotes is controlled by two types of DNA elements, promoters and enhancers. The promoter is required for accurate and efficient basal transcription as discussed in some detail in the previous chapter. The rate of transcription from the promoter can be increased by the activity of the second type of the elements called enhancers.

Enhancers function in an orientation independent manner and can act over long distances, either upstream or downstream from the transcription unit (Serfling, *et al.* 1985). Enhancer activity can be exerted by the synergistic effects of discrete DNA elements that specifically interact with cellular proteins. A number of experiments have illustrated that some promoters and enhancers are functionally interchangeable. For instance, the constitutive promoter elements of the inducible metallothionein-I gene could behave like an enhancer when detached from the TATA box and linked to a heterologous gene (Serfling, *et al.* 1985). Another example is the SV40 enhancer that was found to stimulate transcription of a linked β-globin gene over distances of more than 3,000 bp, even from a downstream position (Banerji, *et al.* 1981). This enhancer element, when placed adjacent to the β-globin gene from which the upstream promoter elements were deleted, could restore a high level of transcription (Treisman and Maniatis 1985).

Bienz and Pelham, 1986 showed that a single heat-shock regulatory element could not function at distance but the duplication of this sequence was able to create an element that showed enhancer properties. This result suggests that multiple protein-protein

interactions are needed for the formation of transcription complexes between separated DNA recognition elements. As enhancers and promoters are usually distantly apart, a possible model for remote control by an enhancer has been proposed in which transcription is stimulated by interactions between regulatory proteins that bind the promoter and the remote enhancer, with the intervening DNA looped out (Ptashne 1986, Serfling, *et al.* 1985).

Enhancers can be divided into two categories depending on the parameters to which they respond, inducible enhancers and tissue-specific enhancers (Maniatis, *et al.* 1987).

Inducible enhancers

Inducible enhancers include the elements that respond to changes in environment such as heat shock, viral infection, exposure to heavy metals as well as those that are activated by growth factors or steroids (reviewed in Maniatis *et al*, 1987). Very short DNA sequences, such as a 12 bp sequence from the human metallothionein gene (Searle, *et al.* 1985), are necessary for regulated gene expression. This short regulatory element is present in multiple copies in each gene and the level of induction depends on the copy numbers of the element. Similar effects have been reported for heat-shock (Dudler and Travers 1984), and β-interferon (Goodbourn, *et al.* 1985) genes. Therefore, the additive effect of tandemly repeated regulatory elements seems to be general phenomenon for this type of enhancer.

Activation of inducible enhancer elements could be achieved by different mechanisms (Maniatis, *et al.* 1987). The simplest mechanism is the activation by positive regulatory factor. Treatment with inducer modifies the regulatory factor that is present in an inactive configuration to the active configuration that is able to bind the regulatory element. The best example for this simple mechanism is the temperature induction of the *Drosophila* heat-shock *hsp70* gene (Parker and Topol 1984). An

alternative mechanism is the inactivation of a negative regulatory factor that binds the regulatory element in uninduced cells. Induction leads to the displacement of the repressor and the binding of regulatory factor to regulatory element. This could be the result of modification of the repressor so that it loses affinity for the template or the target of modification could be the regulatory factor. In the latter case, the modification increases template affinity of the regulatory factor so that it competes with the repressor for the regulatory DNA element. The regulation of the human β-interferon gene is a well characterised example of this type of induction (Goodbourn, *et al.* 1986, Zinn and Maniatis 1986). Enhancer activation could also be the consequence of both the activation of a regulatory factor and the inactivation of the repressor.

Tissue-specific enhancers

The best example of tissue-specific gene expression is the expression of immunoglobulin (Ig) genes (Gerster, *et al.* 1987) that is limited to B cells and is activated at a specific stage during lymphoid cell differentiation. The Ig heavy-chain enhancer, situated in the intron between the joining (j) and the constant (c) segments of the gene, was the first genetic element identified to confer cell-type specificity (Gillies, *et al.* 1983).

Unlike immunoglobulin genes, a large number of genes are transcribed in several tissues or at different stages during development. One of the possible models for the regulation of gene expression in multiple cell types is that a gene is controlled by different *trans*-acting factors present in different tissues. These tissue-specific factors could interact with the same enhancer, where different factors recognise different sequences of a single enhancer element, or with separate enhancers (Maniatis, *et al.* 1987). Examples of regulation that provide evidences for this model are those of the Drosophila alcohol dehydrogenase gene (Fischer and Maniatis 1986) and yolk protein genes (Logan and Wensink 1990).

The transcripts of two *Drosophila* yolk protein genes, *yp1 and yp2*, have been found only in ovarian follicle cells and in female fat body cells. The cell-specific expression of both yolk protein genes is controlled by two distinct *cis*-regulatory elements, ovarian enhancer 1 and 2, that act on the promoters of both genes (Logan and Wensink 1990). Another yolk protein gene, *yp3*, has similar developmental pattern to those of *yp1* and *yp2* but is controlled by different regulatory sequences. Two independent enhancer elements have recently been identified to regulate tissue specific expression of *yp3* (Ronaldson and Bownes 1995).

I factor enhancer

The expression of the *I* factor of *Drosophila melanogaster* is also regulated in a tissue-specific manner as high levels of transposition could be detected only in the ovaries of SF females. Previous studies on the *I* promoter suggested that two elements, including the sequences of nucleotides +41 to +100 and nucleotides +101 to +186 of the *I* factor, are capable of enhancing the promoter activity in ovaries of reactive flies (C. McLean, unpublished data). More recently, the DNA element containing both sequences, that is nucleotides 41-186, has been shown to have enhancer properties. This element , when placed adjacent upstream of the *hsp70* promoter fused to the *lacZ* gene, was able to confer the expression of the *lacZ* gene in the absence of heat-shock treatment. This enhancer activity was observed regardless of the orientation of the +41 to +186 element (S. Forbes, unpublished data). The aim of this chapter is to quantitatively assay the enhancer activity of the region +41 to +186 of the *I* factor. Flies transformed with different *I-lacZ* constructs were homogenised and the homogenate was used in the assays for β-galactosidase activity by colour reaction with O-nitrophenyl-β-galactopyranoside (ONPG) as substrate.

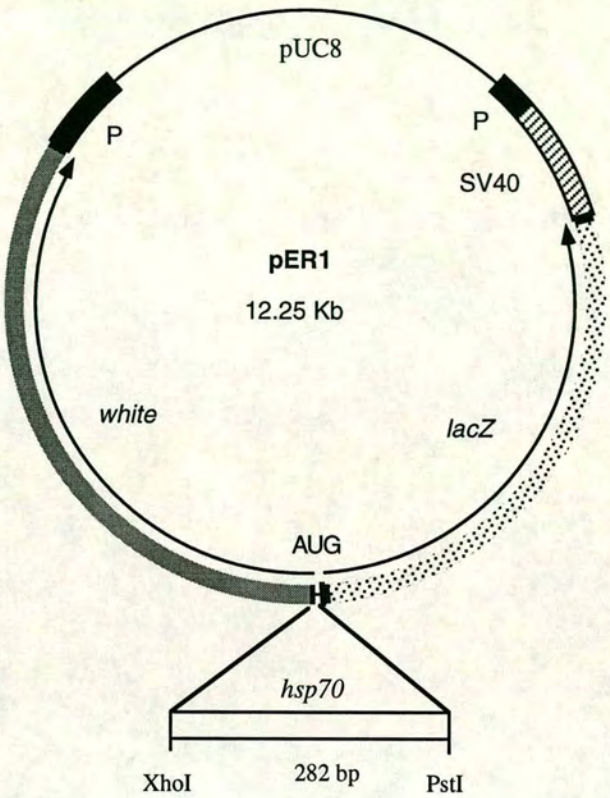


Figure 4.1 pER1 transformation vector (Ronaldson and Bownes 1995)

The *P* element transformation vector, pER1, contains the *white* gene as a selection marker and the *lacZ* reporter gene that is under control of 282 bp *Xho*I-*Pst*I fragment from the *hsp70* promoter region. This vector allows insertion of the DNA fragment of interest at a unique *Xho*I site adjacent to and upstream of the *hsp70* promoter. For details of its construction see Ronaldson and Bownes, 1995.

4.2 RESULTS

β-galactosidase activity assay of flies transformed with *I-lacZ* constructs

Four different *I-lacZ* constructs have been made in order to detect the enhancer activity lying within nucleotides 41 and 186 of the *I* factor. The fragments 41-186 and 99-186 were obtained by PCR from pI954 template that contains a functional *I* factor sequence (Pritchard, *et al.* 1988). The primers used in these cases contain an *Xho*I site at the 5' end to make it possible for the insertion of the PCR fragments to the *Xho*I site in the vector. The pER1 vector is illustrated in figure 4.1 (Ingolia, *et al.* 1980, Ronaldson and Bownes 1995).

The constructs pI821, pI822 were obtained by insertion of the 41-186 fragment to the *Xho*I sites of the pER1 vector immediately upstream of the *hsp70* promoter in the 5' to 3' and 3' to 5' directions respectively (figures 4.2 a, b). The constructs pI823 and pI824 have the 99-186 fragment inserted upstream of the *hsp70* promoter in the 5' to 3' and 3' to 5' directions respectively (figures 4.2 c, d). The constructs were used for injecting embryos of the W^K strain and homozygous lines of the transformants for each construct were established. The construction of the *I-lacZ* constructs and the transformation of flies were carried out by S. Forbes.

Figure 4.3 shows the β-galactosidase activity assays of W^K and flies transformed with pER1 as controls. The assays were done using extracts from the ovaries of the flies that had been heat-shocked as well as those that had not. The ovaries were homogenised thoroughly in cold homogenisation buffer and samples were centrifuged at 12,000 rpm for 5 minutes. Four hundred microlitres of the supernatant was pre-incubated at 37°C for 5 minutes before 600 µl of O-Nitrophenyl-β-galactopyranoside (ONPG) was added and the incubation continued at 37°C. The enzyme reactions were followed by measuring OD_{420} at 10 minute time intervals. The activities were

measured as a slope of the graphs plotted between OD_{420} and time in the unit of $OD_{420}x10^{-2}$/min/mg. Low levels of activity were observed in either non heat-shocked or heat-shocked W^K flies whereas the activity of heat-shocked flies transformed with pER1 was enhanced more than sevenfold when compared to the activity of non heat-shocked flies.

Figures 4.4 a-d illustrate the β-galactosidase activity assays of lines transformed with pI821, pI822, pI823 and pI824. The activities of flies subjected to heat-shock and flies that were not treated with heat-shock were compared in each case. Table 4.1 shows the values of β-galactosidase activities in extract of each transformed line. In non heat-shock condition, lines 272 and 274 obtained from transformation with pI821, and lines 279 and 280 from pI822 expressed β-galactosidase activity at the levels of about 14 and 6 fold greater than the level of activity in control flies transformed with pER1 vector. The levels of activity in these results suggest that sequence 41-186 of the *I* factor can induce a high level of expression of the *lacZ* gene independently of its orientation. In contrast, most of the lines obtained from pI823 and pI824 exhibited more or less similar levels of β-galactosidase to that of the line obtained from pER1 vector in both heat-shock and non heat-shock conditions.

4.3 DISCUSSION

The results in this chapter show that nucleotides 41-186 of the *I* factor contain the sequences that are capable of inducing expression of the *lacZ* gene from the *hsp70* promoter in the absence of heat shock treatment.

The patterns of expression of the *lacZ* gene have been observed by staining for β-galactosidase in the ovaries of reactive flies transformed with the *I-lacZ* constructs (S. Forbes, unpublished results). Without heat-shock, no β-galactosidase staining was seen in the ovaries of control flies carrying the sequences of *hsp70-lacZ* alone

Figure 4.2 Construction of the *I-lacZ* constructs

The 41-186 bp and the 99-186 bp fragments from the *I* factor promoter were obtained by PCR using plasmid pI954 that carries functional *I* factor as a template . The primers for both fragments contain the *Xho*I site at their 5' ends to make the insertion at the *Xho*I site possible. The fragments were then inserted in the pER1 vector at the *Xho*I site and the orientation of the inserts was checked by DNA sequencing.

a. pI821 The 41-186 bp of the *I* factor was inserted at the *Xho*I site in the pER1 vector in a 5'-3' direction.

b. pI822 The 41-186 bp of the *I* factor was inserted at the *Xho*I site in the pER1 vector in a 3'-5' direction.

c. pI823 The 99-186 bp of the *I* factor was inserted at the *Xho*I site in the pER1 vector in a 5'-3' direction.

d. pI824 The 99-186 bp of the *I* factor was inserted at the *Xho*I site in the pER1 vector in a 3'-5' direction.

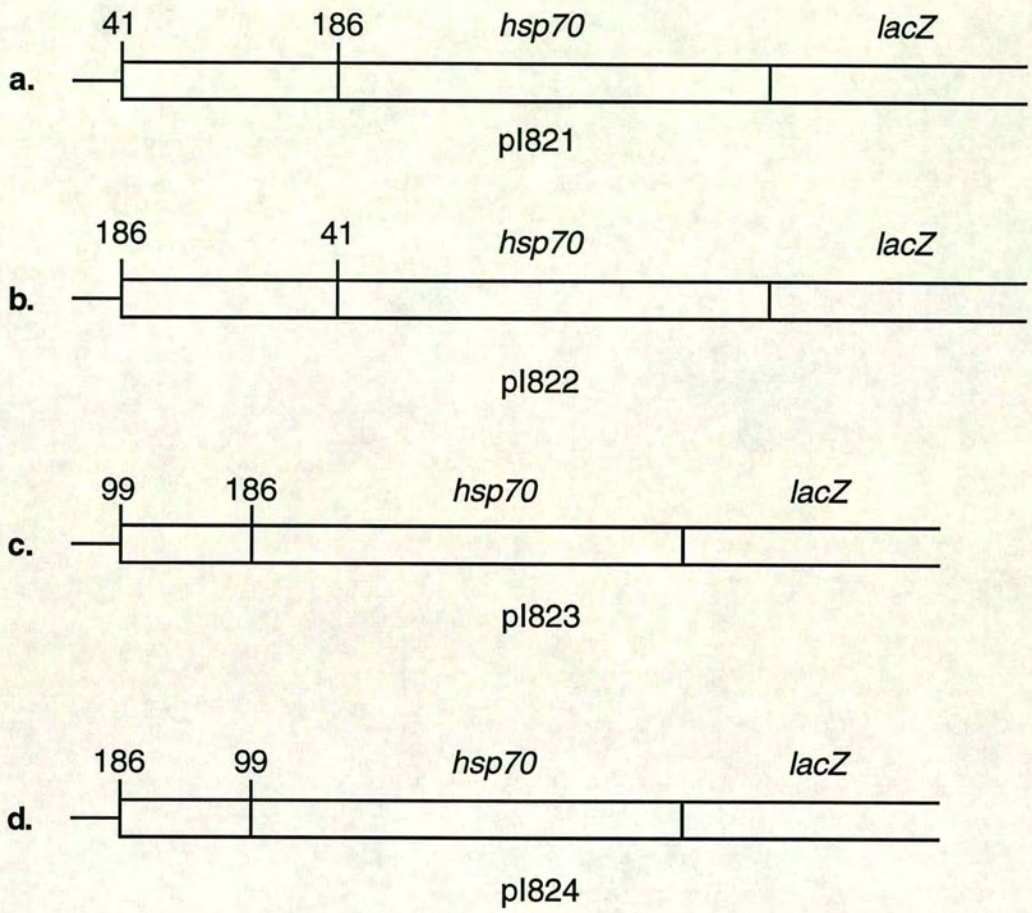
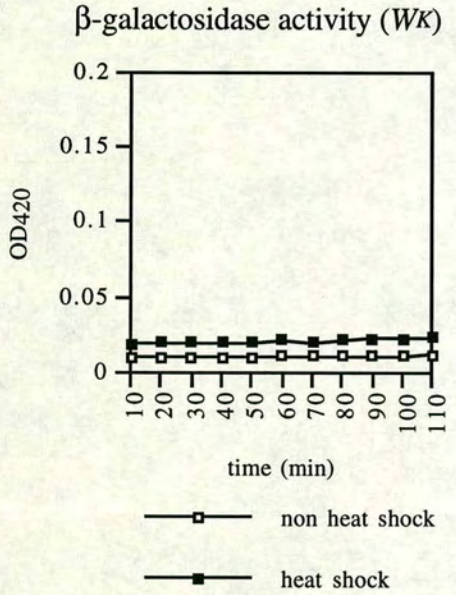

a.
41
186
hsp70
lacZ
pl821
b.
186
41
hsp70
lacZ
pl822
c.
99
186
hsp70
lacZ
pl823
d.
186
99
hsp70
lacZ
pl824

Figure 4.3 β-galactosidase activity assays in the ovaries of W^K flies and flies transformed with pER1 vector

The reactions were allowed to occur at 37°C for 2 hours. The absorbance at 420 nm were measured every 10 minutes. The values of the activity are expressed as $OD_{420}x10^{-2}$/min/mg protein.


β-galactosidase activity (WK)
OD420
0.2
0.15
0.1
0.05
0
10
20
30
40
50
60
70
80
90
100
110
time (min)
non heat shock
heat shock

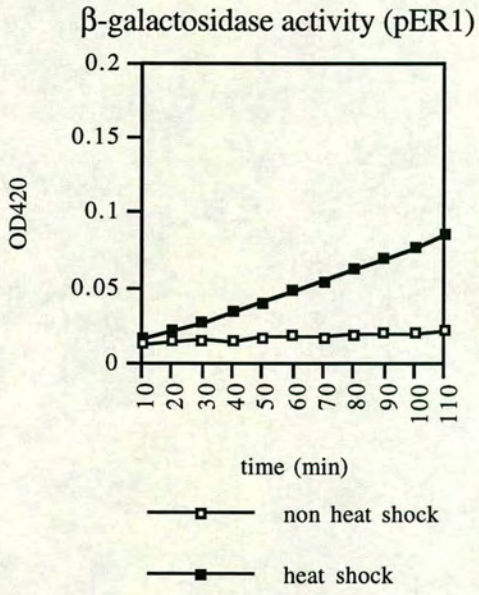

β-galactosidase activity (pER1)
OD420
0.2
0.15
0.1
0.05
0
10
20
30
40
50
60
70
80
90
100
110
time (min)
non heat shock
heat shock

Figure 4.4 β-galactosidase activity assays in the ovaries of flies transformed with *I-lacZ* constructs

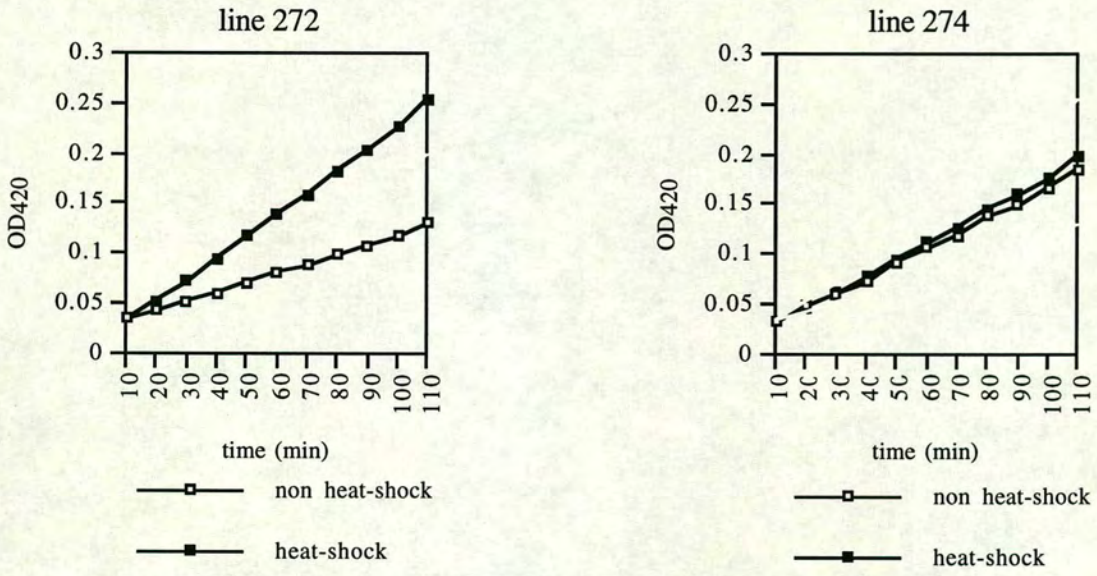
The reactions were allowed to occur at 37°C for 2 hours. The absorbance at 420 nm were measured every 10 minutes. The values of the activity were expressed as $OD_{420}x10^{-2}$/min/mg protein.

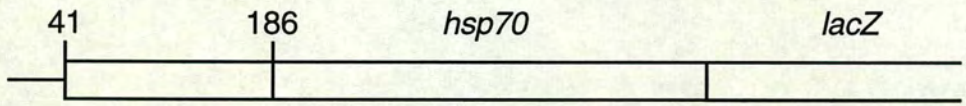
a. activity of transformed lines from pI821

b. activity of transformed lines from pI822

c. activity of transformed lines from pI823

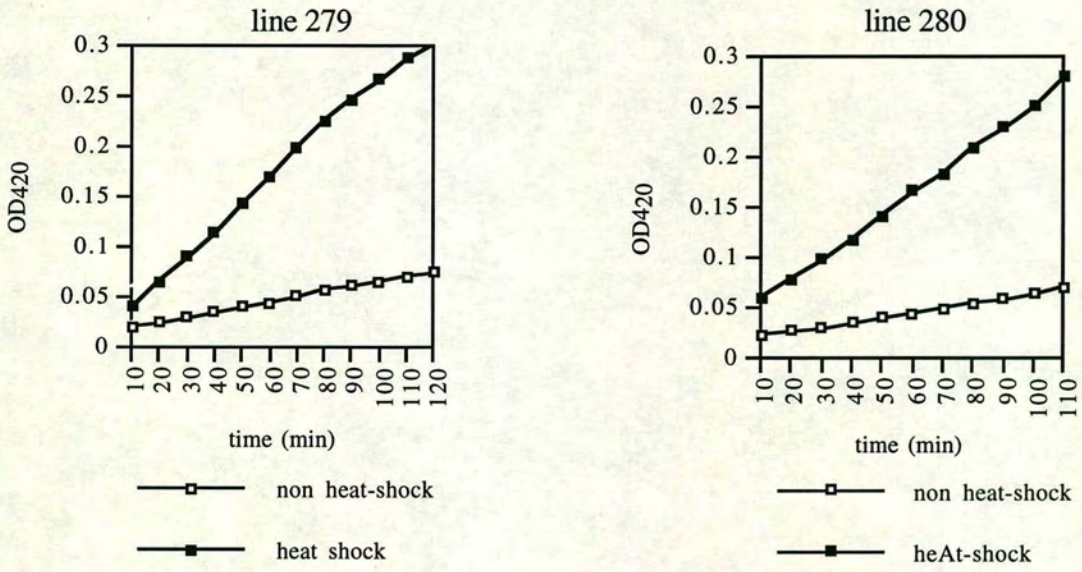
d. activity of transformed lines from pI824

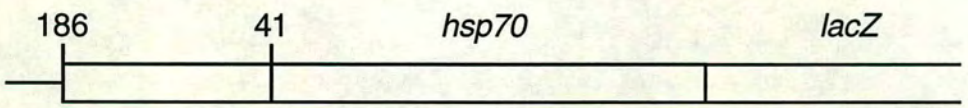




pl821

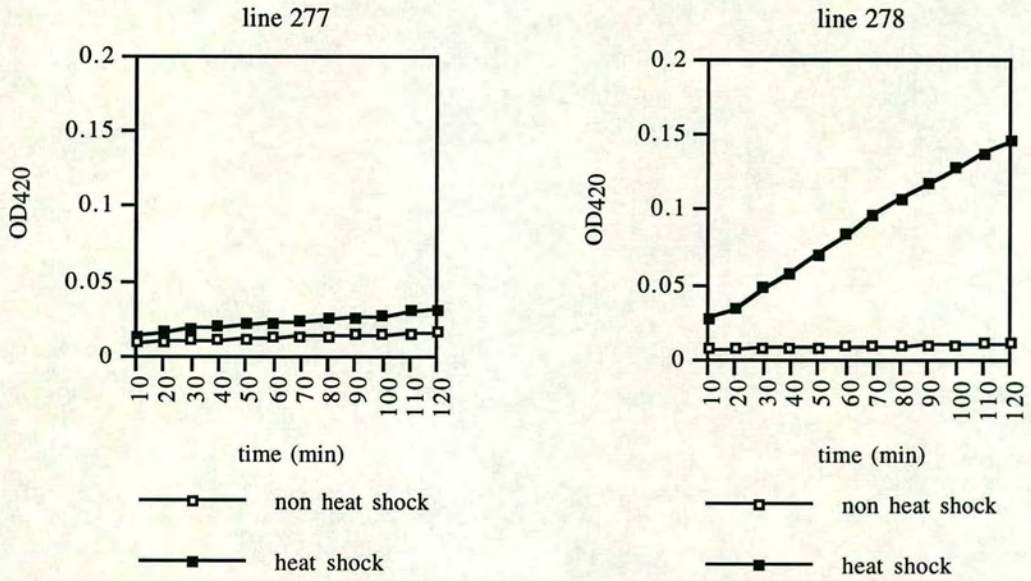
figure 4.4 a

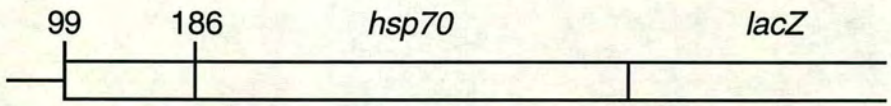




pl822

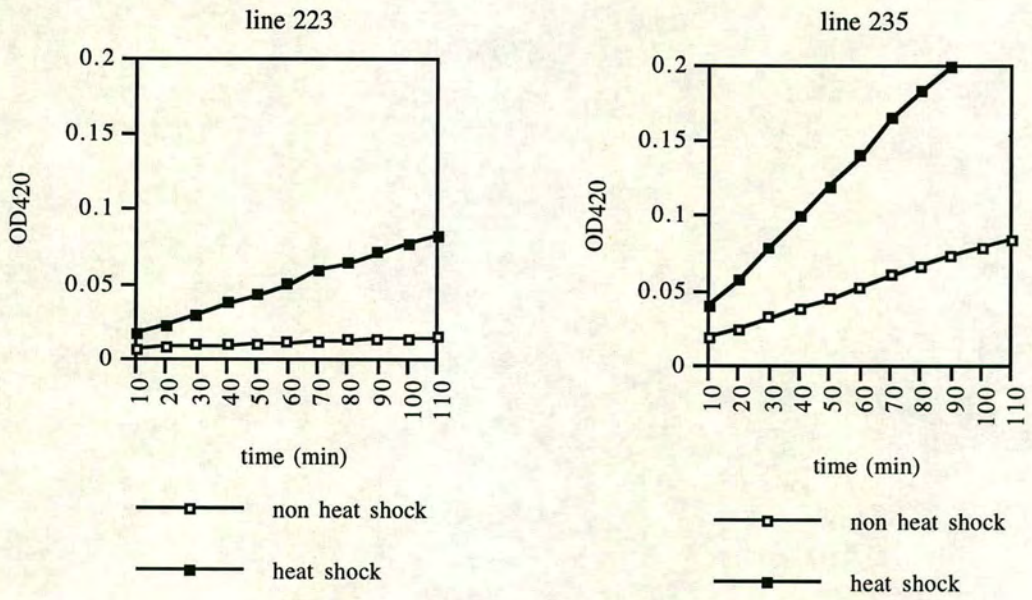
figure 4.4 b

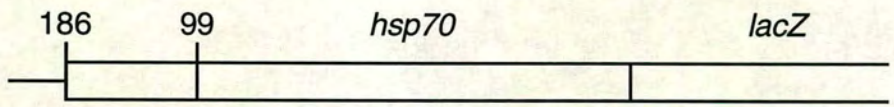




pI823

figure 4.4 c





pI824

figure 4.4 d

Table 4.1 β-galactosidase activity of control and transformed lines in non heat shock and heat shock conditions

LINE	Activity ($OD_{420}x10^{-2}$/min/mg)	
	Non heat shock	Heat shock
w^{κ}	1.8±0.4	3.4±1.1
pER1	8.9±0.7	69±2.3
272 (pI821)	95±11.0	218±14
274 (pI821)	150±18.0	164±13
279 (pI822)	51±7.5	265±11
280 (pI822)	48±5.5	220±18
275 (pI823)	11±2.0	41±3.4
277 (pI823)	5.5±1.9	15±0.8
278 (pI823)	3.4±0.2	110±12
223 (pI824)	7.4±0.8	67±3.1
235 (pI824)	67±12	195±8.0

whereas flies carrying 41-186 fragment of the *I* factor, in either orientation, adjacent to and upstream of the *hsp70-lacZ* showed strong staining in the nurse cells of stages 8 to 10 of oogenesis. On the other hand, the staining in oocytes at these stages was very weak, if any, but increased to some degree at stage 13. No staining was observed in egg chambers at earlier stages. These staining patterns were different from those obtained when flies were heat-shocked prior to staining in which case β–galactosidase activity was detected in the somatic follicle cells at all stages. This supports the conclusion that nucleotides 41-186 of the *I* factor contain an enhancer that stimulates germ line specific expression of the *I* promoter in the ovaries of reactive flies.

The strength of this enhancer was estimated by comparing the β-galactosidase activity in flies carrying *hsp70-lacZ* fragment preceded by 41-186 bp of the *I* factor with the activity in control flies carrying *hsp70-lacZ* fragment alone. The levels of β-galactosidase activity in flies carrying the 41-186 bp in 5'-3' and 3'-5' orientations were about 14 fold and 6 fold greater than the activity in control flies respectively.

On the other hand, extracts of ovaries from most of the transformed lines carrying the smaller fragment, 99-186 bp, of the *I* factor in either orientations gave levels of β-galactosidase activity very similar to those from flies carrying *hsp70-lacZ* alone. The difference in the levels of β-galactosidase activity between the two lines carrying 99-186 bp fragment in 3'-5' orientation (lines 223 and 235) could be due to the effect of the flanking chromosomal sequences that allowed the expression of the *lacZ* gene to occur more efficiently in line 235. This effect could also contribute to the lower levels of β-galactosidase activity observed in some lines carrying 99-186 bp fragment as compared to that in flies carrying sequences from pER1 alone. The quantitative data of β-galactosidase activity assays are consistent with the β-galactosidase staining pattern reported by S. Forbes.

Although the 99-186 bp fragment showed very weak enhancer activity, the deletion of this sequence from the full-length 186 bp *I* promoter resulted in a dramatic drop in the

activity of the promoter in the ovaries but not in non ovarian tissues (carcass) as measured by the activity of CAT expressed from the full-length (1-186) and the deleted (1-100) *I* promoter fragments. This brought the ratio of the level of expression in the ovaries to the level in carcass down about 5 fold from the ratio of the activities expressed from the full-length promoter (C. McLean, unpublished data and table 1.2). Further deletion of the 3' end of the *I* promoter to nucleotide 40 caused another 6 fold decrease in this ratio.

Taken together, these results suggested that the 5' UTR of the *I* factor contains separate enhancer elements, element 1 within the 41-99 bp and element 2 within the 100-186 bp, that synergistically enhance expression from the *I* factor promoter in the ovaries of reactive flies. The synergistic effect of two enhancer elements that leads to a high level of activity has also been reported for the SV40 enhancer (Zenke, *et al.* 1986).

Element 2 is unable to function as a strong enhancer unless it is cooperated with the element 1. Whether or not element 1 itself is capable of enhancing the level of expression from the promoter is yet to be determined. Derivative of pER1 with nucleotides 41-99 immediate upstream of *hsp70-lacZ* are being constructed. Once the construct has been obtained, it will be used to transform reactive flies of strain W^K. The staining pattern of β-galactosidase and the activity of β-galactosidase will be assayed in the ovaries of transformed lines to determine the enhancer activity of 41-99 bp fragment of the *I* factor.

CHAPTER 5

Identification of DNA-protein interactions in nucleotides 1-186 of the *I* factor

5.1 INTRODUCTION

The analysis of the *I* factor promoter region described in previous chapters suggests that there are at least two elements located within nucleotides 41-186 of the *I* factor that play a role in enhancing its expression in the ovaries of reactive flies. The first element, element 1, is located between nucleotides 41-100 and element 2 lies within nucleotides 101-186. A model for the regulation of *I* factor expression has been proposed by C. McLean that involves recognition of these two possible enhancer sequences by regulatory proteins. However, no evidence on these DNA-protein interactions in the promoter region of the *I* factor has been published so far.

The experiments described in this chapter were designed to detect such a regulatory factor in nuclear extracts from tissues of *Drosophila melanogaster.* Gel retardation experiments were performed to detect any DNA binding activities in nuclear extracts prepared from various tissues of reactive (*Cha*) and inducer (*Cha-RC*$^{+}$) flies that recognise nucleotides 1-186 of the *I* factor. The DNA sequence within nucleotides 1-186 recognised by this factor were identified by DNA footprinitng experiment using nuclear extracts from the ovaries of reactive females.

5.2 RESULTS

5.2.1 Determination of a specific DNA binding activity recognising a sequence within nucleotides 1-186 of the *I* factor

Nuclear extracts from ovaries of reactive females were chosen for the detection of sequence specific DNA binding factors that recognise the regulatory sequences in the *I*

factor promoter. These extracts should be a good source for these transcription factors as expression of the *I* factor is restricted to the germ line of reactive females.

The 1-186 bp fragment of the *I* factor was obtained by digestion of p186T.1 (McLean, *et al.* 1993) with *Xba*I and *Bam*HI followed by purification from agarose gel by MERmaid kit (2.2.2.6). The 3' ends of the fragment were radiolabelled with α-^{32}P-dCTP by the activity of Klenow polymerase enzyme as described in 2.2.3.6. The binding reactions were performed in the presence of 1 μg of poly[dI-dC] as a non-specific competitor and, in particular cases, a 100 fold molar excess of unlabelled 1-186 bp fragment as a sequence specific competitor.

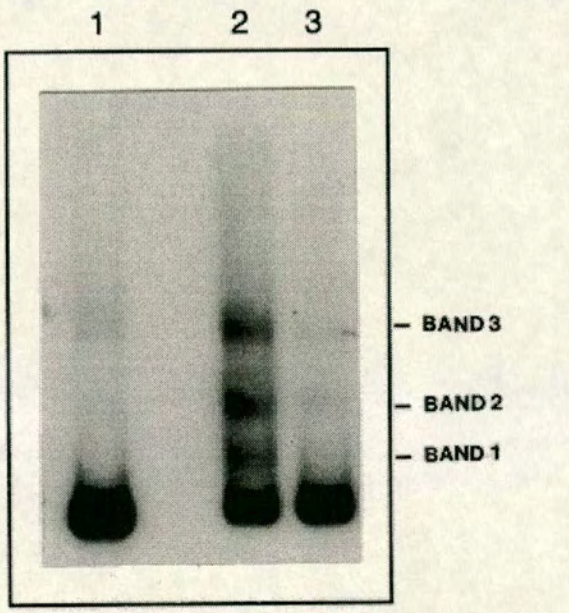
Figure 5.1 shows that the radioactive probe, nucleotides 1-186, was bound by a factor or factors present in nuclear extract from the ovaries of reactive females (lane 2). Three retarded bands, bands 1-3, were seen representing the binding of proteins to the probe. The binding was sequence specific as the retarded bands were not observed in the presence of a 100 fold molar excess of unlabelled 1-186 fragment (lane 3).

5.2.2 Increasing amounts of ovarian nuclear extract do not affect the pattern of DNA-protein complexes in gel retardation assays

The sequence-specific DNA-protein complexes formed between nucleotides 1-186 of the *I* factor DNA and the binding factors in nuclear extract from ovaries of reactive females differ in their molecular weights. The simplest explanation of this is that the complexes are the consequence of the binding of different proteins to the same DNA molecule. Alternatively, the presence of more than one complexes could reflect a difference in the stoichiometries of the complexes for instance, the DNA that has 1, 2, 3 or more molecules bound to it (Fried 1989). In order to determine which could be the case for the formation of the DNA-protein complexes detected in figure 5.1, the amounts of nuclear extract from the ovaries of reactive females added to the binding

Figure 5.1 Gel retardation assays for DNA binding activity in ovarian nuclear extract from reactive females to the 1-186 bp of the *I* promoter

Three micrograms of nuclear extract prepared from ovaries of reactive females were mixed with about 1 ng of end-labelled 1-186 bp fragment of the *I* factor in the presence of 1 μg of poly[dI-dC] as a non-specific competitor. Lane 1 contains only the 1-186 probe, lane 2 contains the probe and nuclear extract, lane 3 is the same as lane 2 but also contains 100 ng of unlabelled 1-186 bp fragment as a specific competitor.


1
2
3
BAND 3
BAND 2
BAND 1
-
+
+
Extract
-
-
186
Competitor

reactions were gradually increased and the patterns of complex formation were observed by gel retardation assays.

Figure 5.2 shows that the ratio of the intensity of the three retarded bands did not change when more extract was supplied to the reaction. This suggests that the presence of the three complexes is independent of each other as increasing amount of the extract in the binding reaction did not favour the formation of the higher molecular weight complexes. Therefore, different complexes is rather due to the binding of different proteins or different forms of one protein to the DNA.

5.2.3 Determination of a binding site in the promoter of the *I* factor that is recognised by a sequence specific factor in reactive ovarian nuclear extract

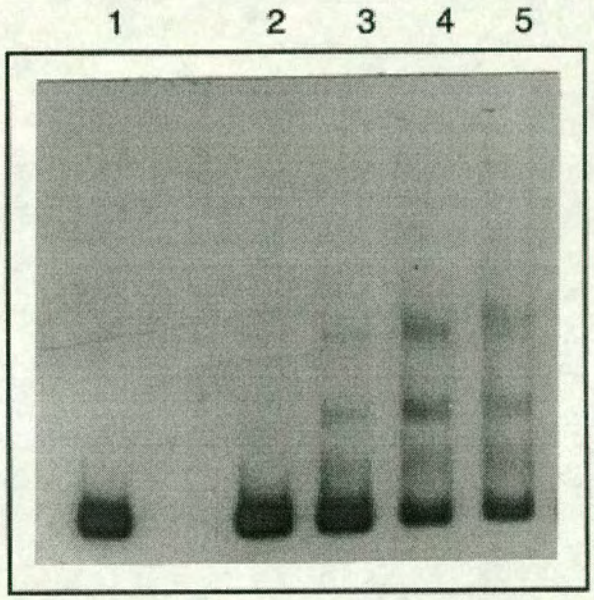
The sequence recognised by sequence specific binding proteins has been determined by DNase I footprinting. DNase I footprinting was developed by Schmitz and Galas in 1978 as a method to study specific binding between DNA and proteins. The technique allows end-labelled DNA fragments to interact with a DNA-binding protein that can be either purified or within a cellular or nuclear extract. The complex is then partially digested with DNase I. The bound protein protects the region of the DNA with which it interacts from the digestion by the DNase I enzyme. The digestion products are denatured and separated by electrophoresis and the protected region is determined as a gap in the ladder produced by nuclease digestion of unprotected DNA.

5.2.3.1 DNase I footprinting on the plus strand of the 1-186 bp fragment of the *I* factor

The DNA fragment to be used in DNase I footprinting must be labelled at one end of one strand. The 1-186 bp fragment of the *I* factor was obtained by digestion of p186T.1 with *Bam*HI and *Pst*I and purified by MERmaid kit. The *Pst*I enzyme left 3'

Figure 5.2 Gel retardation assay with increasing amounts of the extract.

About 1 ng of the 1-186 bp probe was mixed with increasing amounts of ovarian nuclear extract from a reactive strain. No extract was added in lane 1 and the amounts of the extract were 0.5, 1, 2 and 3 μg in lanes 2-4 respectively.


1
2
3
4
5
-
0.5
1
2
3
Extract (μg)

overhang at one end of the fragment such that this end cannot be labelled by the activity of Klenow polymerase. The 3' end of the plus strand of the 1-186 fragment was labelled by the incorporation of α-^{32}P-dCTP using the 5' overhang from the *Bam*HI site as a template (figure 5.3 a). The DNase I footprinting was performed as described in 2.2.3.8.

Figure 5.4 shows the footprint on the plus strand of 1-186 bp of the *I* factor. Digestions with increasing amounts of DNase I were performed to determine the appropriate amount of enzyme that give a clear footprint. The appropriate DNase I dilutions are 0.003 Kunitz U/µl for free DNA and 0.04 Kunitz U/µl for protein-bound DNA. The prominent footprint was seen between nucleotides 137 and 150 of *I* promoter fragment. No any other obvious protected region was detected along the fragment.

5.2.3.2 DNase I footprinting on the minus strand of the 1-186 fragment of the *I* factor

The footprint seen on the plus strand of the *I* promoter was confirmed by determining the footprint on the other strand. The *Bam*HI-*Pst*I fragment purified from p186rp (McLean, *et al.* 1993) allows 3' end labelling by Klenow polymerase on its minus strand (figure 5.3 b). Figure 5.5 shows that the same region, nucleotides 137-150, was protected by the binding factor present in ovarian nuclear extract of reactive flies. This 137-150 bp, ACAAAAACAACAAT, was the only protected region that could be detected in this experiment.

5.2.4 The binding factors in ovarian nuclear extract from reactive strain recognise the footprint region

Preliminary experiments by I. Arkhipova (unpublished data) revealed that a synthetic double-stranded oligonucleotide comprising 19 bp of the *I* factor sequence,

Figure 5.3 3' end labelling of the 1-186 bp fragment by the activity of Klenow polymerase

a. Labelling of the plus strand The fragment was obtained from p186T.1 (McLean, *et al.* 1993) as a *Bam*HI-*Pst*I fragment. The *Bam*HI enzyme creates a 5' overhang on the minus strand such that it could be used as a template for Klenow polymerase to fill-in the 3' end of the plus strand.

b. Labelling of the minus strand The fragment was obtained from p186rp (McLean, *et al.* 1993) as a *Bam*HI-*Pst*I fragment. The *Bam*HI enzyme creates a 5' overhang on the plus strand such that it could be used as a template for Klenow polymerase to fill-in the 3' end of the minus strand.

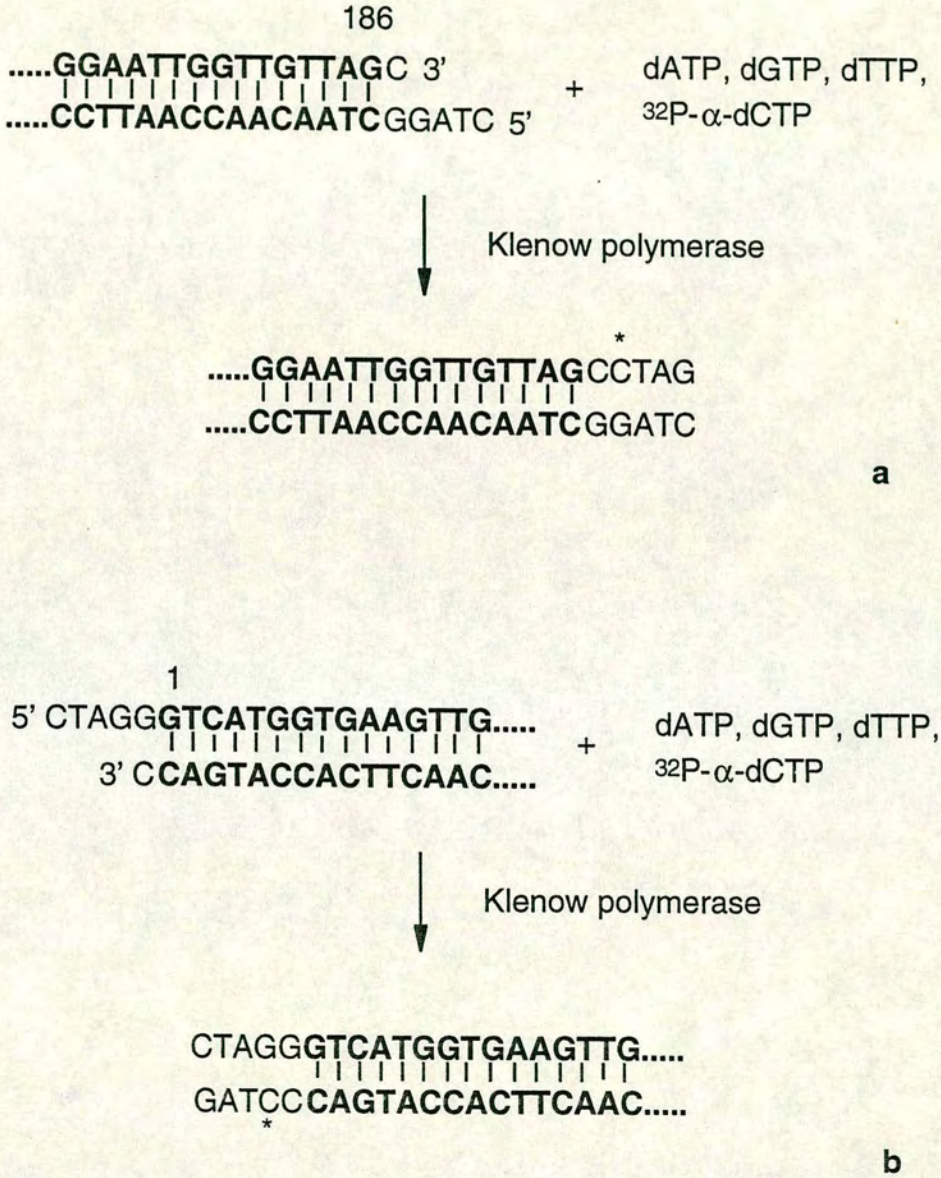

186
.....GGAATTGGTTGTTAGC 3'
.....CCTTAACCAACAATCGGATC 5'
+
dATP, dGTP, dTTP,
32P-α-dCTP
Klenow polymerase
*
.....GGAATTGGTTGTTAGCCTAG
.....CCTTAACCAACAATCGGATC
a
1
5' CTAGGGTCATGGTGAAGTTG.....
3' CCAGTACCACTTCAAC.....
+
dATP, dGTP, dTTP,
32P-α-dCTP
Klenow polymerase
CTAGGGTCATGGTGAAGTTG.....
GATCCCAGTACCACTTCAAC.....
*
b

Figure 5.4 DNase I footprinting on the plus strand of the *I* factor promoter

The binding reactions between the 1-186 bp fragment and reactive ovarian nuclear extract were performed as described for gel retardation. After the binding reactions, the DNA-protein complexes were partially digested with appropriated amount of DNase I and the digested products separated by 6% acrylamide gel electrophoresis.

Lane 1 is the G+A marker obtained by Maxam-Gilbert reactions used to localised the footprint region. In lanes 2-4 the unbound probe (U) was digested with increasing amounts of DNase I (0.005, 0.015 and 0.025 U respectively). In lanes 5-7 the bound DNA (B) was digested with increasing amounts of DNase I (0.1, 0.2 and 0.4 U respectively). The numbers on the right indicate the protected region on the probe.

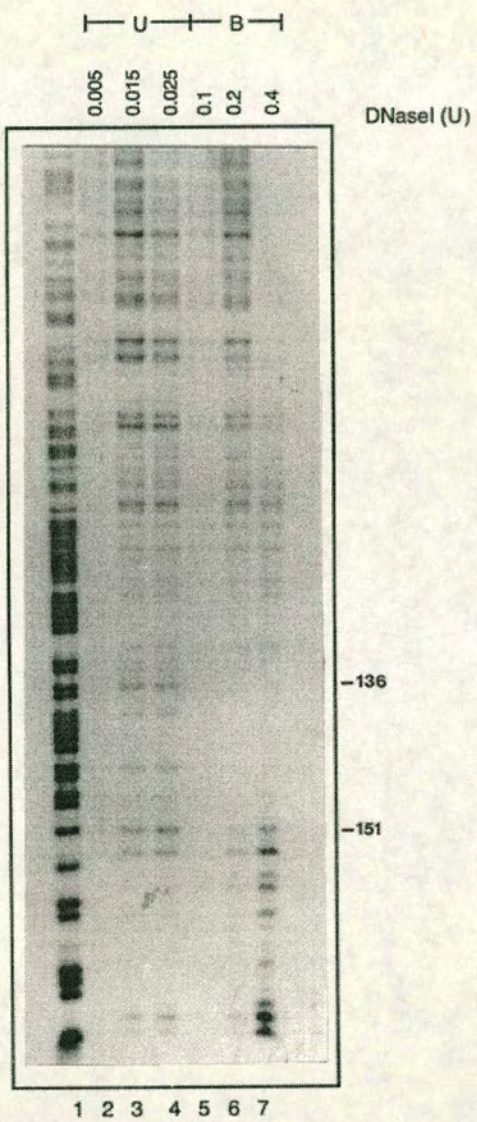
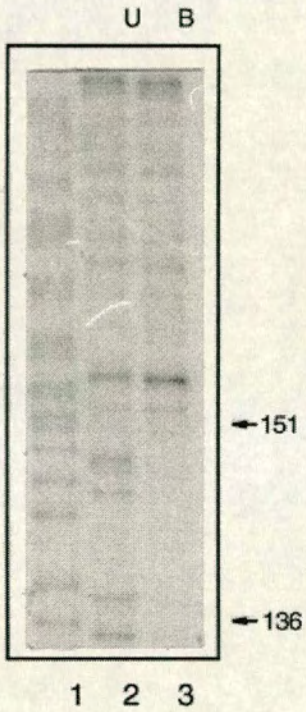

U
B
0.005
0.015
0.025
0.1
0.2
0.4
DNaseI (U)
–136
–151
1 2 3 4 5 6 7

Figure 5.5 DNase I footprinting on the minus strand of the *I* factor promoter

The reactions were the same as described for the plus strand except that the probe was labelled on the minus strand.

Lane 1 is G+A marker, lane 2 is unbound probe (U) digested with 0.015 U of DNase I and lane 3 is bound DNA (B) digested with 0.2 U of DNase I.


U
B
151
136
1
2
3

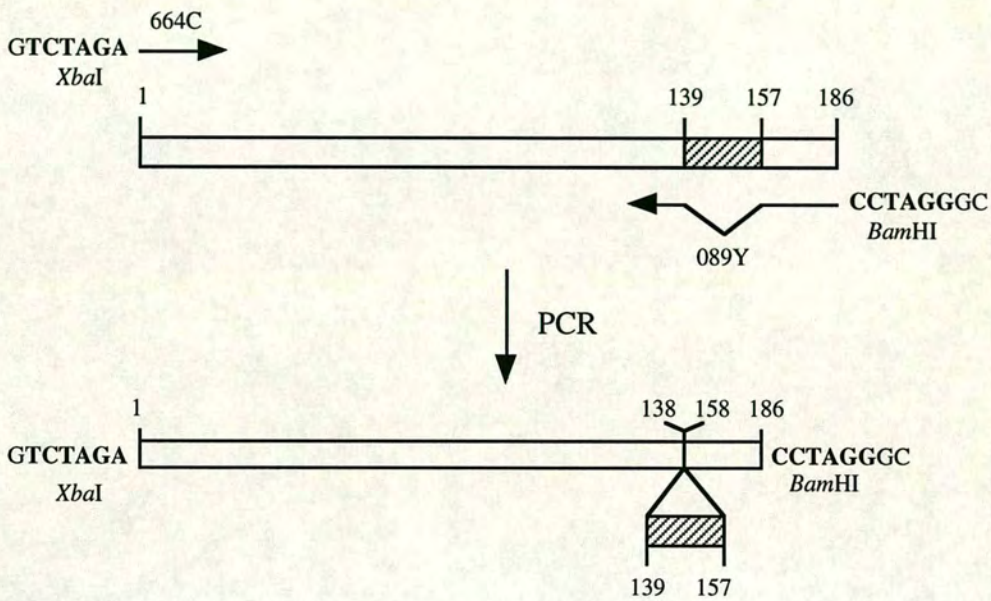
nucleotides 139-157, that cover most of the footprint region shown in figure 5.4 and 5.5 completely abolished the formation of the DNA-protein complexes between the 1-186 bp probe and the factors present in ovarian nuclear extracts of reactive females. In this thesis this 139-157 bp will be referred to as site 1. The complementary experiment has been done by looking at the effect on the formation of DNA-protein complexes of the 1-186 bp fragment from which site 1 has been deleted.

A DNA fragment called 1-Δ186 containing nucleotides 1-186 of the *I* factor but without site 1 was obtained by PCR as illustrated in figure 5.6. The oligonucleotide 089Y was used as a minus strand primer for PCR. This primer contains nucleotides 186 to 128 with site 1 deleted such that nucleotide 158 is next to nucleotide 138. The PCR product was purified by MERmaid kit and was used as a competitor in gel retardation experiment.

Figure 5.7 shows that the DNA-protein complexes still occurred in the presence of unlabelled 1-Δ186 fragment (lane 3). This reveals that the 1-Δ186 fragment failed to compete for the formation of DNA-protein complexes and that nucleotides 139-157 of the *I* factor was the only binding site for the binding factors that were present in the ovarian nuclear extract from reactive flies.

5.2.5 Determination of sequence-specific DNA binding activity in various tissues of *Drosophila* from a reactive strain

The previous results in this chapter suggest that site 1 is recognised by a sequence specific binding activity from the ovaries and thus is the most likely candidate for the enhancer element that plays a role in specific enhancement of the activity of the *I* factor promoter in the ovaries of reactive females. In order to determine whether or not the binding factors that recognise site 1 were present only in the ovaries, the binding activity to the 1-186 bp probe were investigated in nuclear extracts from various tissues of reactive flies.



1-Δ186 fragment

Figure 5.6 Synthesis of the 1-Δ186 fragment by PCR

Plasmid p186T.1 was used as a template for PCR. Oligonucleotide 664C comprising bases 1-20 of the *I* factor with the *Xba*I site attached at the 5' end was used as a + strand primer. Oligonucleotide 089Y, used as a - strand PCR primer, contains bases 186 to 128 of the *I* factor from which bases 139-157 were deleted and has the *BamH*I site attached to base 186 at its 5' end.

Figure 5.7 Gel retardation assays with 1-Δ186 fragment as competitor

Three micrograms of nuclear extract prepared from the ovaries of reactive females were mixed with about 1 ng of end-labelled 1-186 bp fragment of the *I* factor in the presence of 1 μg of poly[dI-dC] as a non-specific competitor. Lane 1 contains only the 1-186 probe, lane 2 contains the probe and the nuclear extract, lane 3 is the same as lane 2 but also contains 100 ng of unlabelled 1-Δ186 bp fragment as a specific competitor.

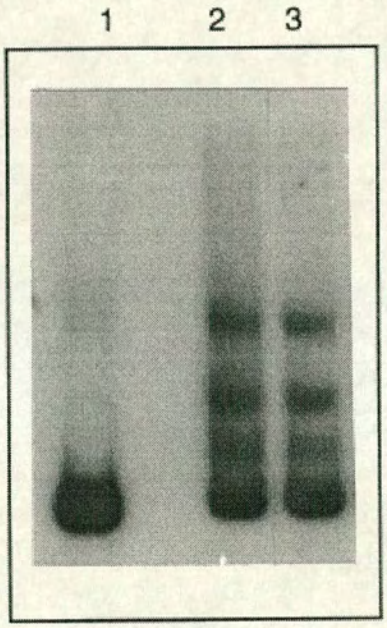

1
2
3
- + +
Extract
- - Δ
Competitor

Figure 5.8 shows gel retardation experiments in which the 1-186 bp probe was mixed with nuclear extracts from males, females, ovaries and female carcass of reactive strain. Similar patterns of binding comprising three major retarded bands, bands 1-3, were observed in the extracts from every tissues (lanes 5, 8, 11, 14) except that band 3 was absent in nuclear extract from females. These bands are due to sequence-specific binding as they were abolished in the presence of a 100 fold molar excess of the unlabelled 1-186 bp fragment (lanes 6, 9, 12, 15) but were unaffected by the 1-Δ186 fragment (lanes 7, 10, 13, 16). These results suggest that the binding factors specifically recognise site 1 of the *I* factor and that these factors are present in nuclear extracts from females, both in ovaries and carcass, as well as in nuclear extract from males.

Amongst the three retarded bands observed with nuclear extract from males, female carcass or ovaries, band 3 is the only band that could be seen with nuclear extract from *Drosophila* tissue culture, SL_2. This extract from SL_2 nuclei also produced a broad band of higher molecular weight that was not seen with the other extracts (figure 5.8, lane 2). Addition of the 1-Δ186 fragment abolished the formation of this high molecular weight band but not the band equivalent to band 3 (lane 4).

5.2.6 Determination of sequence-specific DNA binding activity in tissues of *Drosophila* from an inducer strain

As transposition of the full-length *I* factors is very rare in an inducer strain, there must be some form of regulation that represses expression of these elements. This regulation may be due to the interaction between regulatory proteins and the promoter sequence. In an attempt to determine such an interaction, nuclear extracts from inducer flies were used in gel retardation experiments. Figures 5.9 shows that the 1-186 bp probe was recognised by factors in the nuclear extracts from ovaries and female carcass of inducer flies (lane 2 and 6). These binding showed similar patterns to those

Figure 5.8 Gel retardation assays for DNA binding activity in nuclear extracts from different tissues from a reactive strain to the 1-186 bp of the *I* promoter

Three micrograms of nuclear extracts prepared from SL_2 tissue culture cells, males, females, ovaries and female carcass of a reactive strain were mixed with about 1 ng of end-labelled 1-186 bp fragment of the *I* factor in the presence of 1 μg of poly[dI-dC] as a non-specific competitor.

The lanes labelled with (-) contain no specific competitor. The lanes labelled with 186 and Δ contain unlabelled 186 and 1-Δ186 fragments as specific competitors respectively.

Abbreviations

SL_2 - SL_2 tissue culture cells

$OV_{(R)}$ - nuclear extract from ovaries of a reactive strain

$CAR_{(R)}$ - nuclear extract from female carcass of a reactive strain

$M_{(R)}$ - nuclear extract from males of a reactive strain

$F_{(R)}$ - nuclear extract from females of a reactive strain

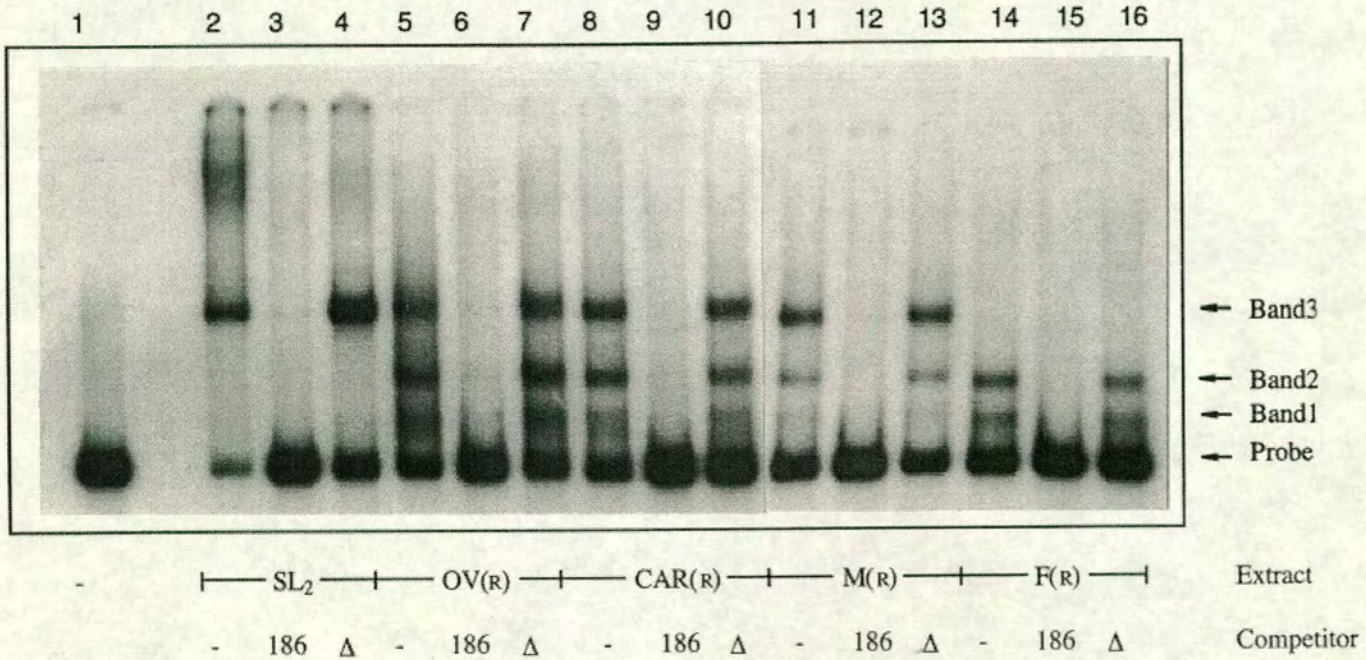
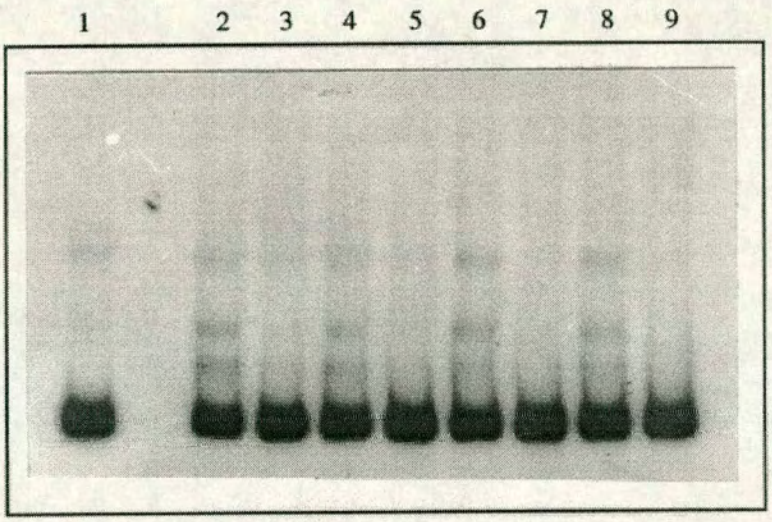

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Band3
Band2
Band1
Probe
- SL2 OV(R) CAR(R) M(R) F(R) Extract
- 186 Δ - 186 Δ - 186 Δ - 186 Δ - 186 Δ Competitor

Figure 5.9 Gel retardation assays for DNA binding activity in nuclear extracts from ovaries and non-ovarian tissues from inducer strain to the 1-186 bp of the *I* promoter

Three micrograms of nuclear extracts prepared from ovaries and female carcass of an inducer strain were mixed with about 1 ng of end-labelled 1-186 bp fragment of the *I* factor in the presence of 1 μg of poly[dI-dC] as a non-specific competitor.

The lanes labelled with (-) contain no specific competitor. The lanes labelled with 186, Δ and O contain unlabelled 1-186, 1-Δ186 and synthetic oligonucleotides comprising bases 139-157 of the *I* factor as competitors respectively.


1 2 3 4 5 6 7 8 9
- OV(1) CAR(1) Extract
- 186 Δ O - 186 Δ O Competitor

seen with extracts from reactive strain and also appear to be sequence specific as the binding was diminished by the addition of unlabelled 1-186 bp fragment.

5.2.7 Nuclear extracts from individual preparation could give different patterns of DNA-protein complex formation

More than one preparations of nuclear extract were made from ovaries of reactive flies during the course of this study. The extracts from some preparations showed different patterns of the retarded bands when used in gel retardation assay from the others. Figure 5.10 shows the results of gel retardation assays in which nucleotides 1-186 probe was mixed with ovarian nuclear extracts of reactive flies from two individual preparations. The differnce observed is that band 3 was absent from the extract of the second preparation (lane 6).

5.3 DISCUSSION

The results of CAT and β-galactosidase activity assays using reporter constructs with various regions of the *I* factor promoter (C. McLean, 1993 and results from the previous chapter) suggest the possibility that two enhancer elements lying within 41-186 bp of the *I* factor work co-operatively for germ line specific expression in ovaries of reactive flies. The sequences that make up the enhancers should be recognised by regulatory factors in order to activate transcritption.

The DNA-protein complexes formed between nucleotides 1-186 of the *I* factor DNA and the protein factors in nuclear extract from ovaries of reactive females were detected in gel retardation assays. These complexes may represent the binding of the proteins in ovarian nuclear extract to one or more sequences within the 1-186 bp of the *I* factor. DNA footprinting experiments confirmed that these binding activities recognised nucleotides 137-150 of the *I* factor. If different proteins in the extract are involved in the formation of different DNA-protein complexes then the proteins must bind the *I*

figure 5.10 Gel retardation assays with reactive ovarian nuclear extracts from different preparations

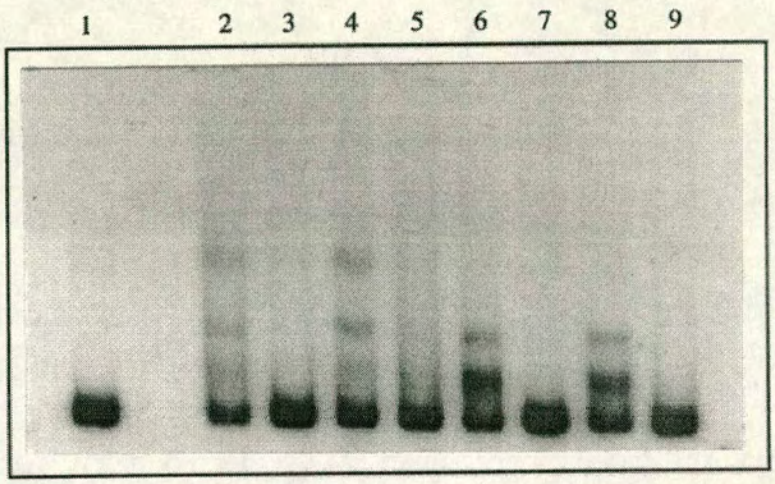
Three micrograms of nuclear extracts prepared from ovaries of two different sets of a reactive strain were mixed with about 1 ng of end-labelled 1-186 bp fragment of the *I* factor in the presence of 1 μg of poly[dI-dC] as a non-specific competitor.

The lanes labelled with (-) contain no specific competitor. The lanes labelled with 186, Δ and O contain unlabelled 1-186, 1-Δ186 and synthetic oligonucleotides comprising bases 139-157 of the *I* factor as competitors respectively.

Abbreviations

1st Ext. - ovarian nuclear extract from the first preparation

2nd Ext. - ovarian nuclear extract from the second preparation


1 2 3 4 5 6 7 8 9
- 1st Ext. 2ndExt.
- 186 Δ O - 186 Δ O
Competitor

promoter at the same site that is nucleotides 137-150. However, it is not necessary that the specific recognition sequences of these proteins are the same. The ability of different proteins to recognise overlapping binding sites present within an enhancer has been recently demonstrated for an enhancer that regulates sex- and tissue-specific transcription of *Drosophila yolk protein* genes (An and Wensink 1995).

Similar patterns of binding were observed when the binding activity that recognised the 1-186 probe was determined in nuclear extracts from different tissues of reactive strain. The relative amount of the smallest complex, band 1, when compared with the higher molecular weight complexes appeared to be less in nuclear extracts from males, females and female carcass than in ovarian nuclear extract. This band was hardly detected at all in the nuclear extract from males as well as in SL_2 nuclear extract. Therefore, this complex may represent binding activity of an ovarian specific binding factor. This tissue specificity may be due to the presence of a larger amount of the protein in the ovaries than in other tissues. Alternatively, this factor could be present in an inactive form in other cell types but is modified in some way before it specifically binds to an enhancer and allows high level of expression in ovaries. Similar mechanism has been observed in the case of NF-κB factor that only induces kappa-chain immunoglobulin gene expression in the B cells where it is present in an active form (Sen and Baltimore 1986).

Band 3 was absent from female nuclear extract whereas it was detected in the extracts from other tissues examined including ovarian and female carcass. This band was also not present in ovarian nuclear extract from some different preparations. Therefore it could represent the binding of a protein that is unstable under some circumstances probably during preparation of the extract. This protein could be necessary but is not sufficient for ovarian-specific expression of the *I* factor as similar amounts of band 3 were present in nuclear extracts from both ovaries and female carcass and this retarded band was the only prominent band present in nuclear extract from males.

Similar patterns of binding were observed in the extracts from males and females of an inducer strain. This result supports the assumption that the factor responsible for enhancing the expression from the *I* promoter is present in all tissues but is only active in ovaries of reactive females. Moreover, the repression of the *I* factor activity in an inducer strain may involve binding of a factor that inhibits the expression to nucleotides 137-150. This inhibitory factor is possibly an unmodified form of a sequence specific binding factor that produced band 3 in gel retardation experiments.

CHAPTER 6

The role of site 1 in controlling expression of the *I* factor promoter

6.1 INTRODUCTION

Site 1, defined as nucleotides 139-157 of the *I* factor promoter, including a footprint region, nucleotides 137-150, detected with nuclear extracts from reactive ovaries is recognised by a sequence-specific binding factor that may induces level of *I* factor expression in reactive ovaries. Factors binding to site 1 are also present in other tissues but may be inactive or function differently. This chapter aims at determining the importance of site 1 to the promoter activity of the *I* factor. Site 1 was deleted from the 1-186 bp promoter and the deleted fragment (1-Δ186) linked to the CAT gene. The fusion was inserted to *P*-transformation vector, pW8 and introduced into embryos of the W^K reactive strain. The activity of the *I* promoter was determined by assaying for a chloramphenicol acetyltransferase activity.

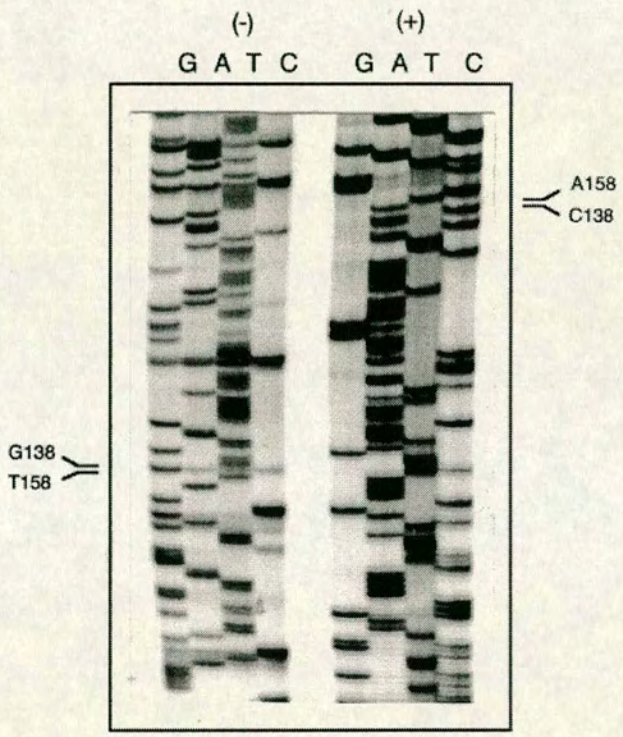
6.2 RESULTS

6.2.1 Construction of p 1-Δ186W8 for *P*-mediated transformation

Plasmid p1-Δ186 was constructed by inserting the 1-Δ186 fragment, obtained by PCR as illustrated in figure 5.5 at the *Bam*HI and *Xba*I sites of the pCAT.1 vector. The sequence of the 1-Δ186 fragment was checked to confirm that nucleotides 139-157 had been exactly removed by sequencing as shown in figure 6.1. The *PstI* fragment from p1-Δ186 carrying the *I*-CAT fusion was inserted into the *P*-transformation vector, pW8, so that expression of the CAT gene is in the opposite direction to that of the *P* promoter and of the *white* gene as illustrated for the construction of p186Δ4.2W8 in figure 3.1. The resulting plasmid, p1-Δ186w8, was introduced into the embryos of W^K reactive strain by *P*-mediated transformation.

Figure 6.1 Sequencing of the 1-Δ186 bp fragment in p1-Δ186

The M13 sequencing primer was used to determine the nucleotide sequence on the plus strand of 1-Δ186 region in p1-Δ186 (+). Oligonucleotide F850 was used as a sequencing primer for the complementary strand (-).


(-)
(+)
G A T C
G A T C
A158
C138
G138
T158

6.2.2 CAT assays for the activity of 1-Δ186 fragment of the *I* promoter

Five homozygous lines, lines 230-234, containing independent insertion of the 1-Δ186-CAT fragment were obtained. The presence of the 1-Δ186*I*-CAT fragment in the genome of each line was determined by genomic Southern analysis by probing *Hind*III digested genomic DNA from each line with the 250 bp *Bam*HI-*Eco*RI DNA fragment containing part of the CAT gene. Figure 6.2 shows that only one band hybridised to the probe in all lines except for line 230 where three hybridising bands were observed. These bands given by each line were of different sizes indicating that they are due to independent single insertion events. Line 230 was the only line that contains three independent insertions.

Protein extracts were prepared from 100 males, 100 females, 100 pairs of ovaries and 100 female carcass from each line and were assayed for the activity of chloramphenicol acetyltransferase as described in 2.2.3.3. Figure 6.3 shows the CAT assays of the transformed lines carrying 1-Δ186*I*-CAT fragment. The values of CAT activities are summarised in table 6.1. The levels of CAT activity expressed in lines 230-234 were several fold lower than the level of the activities expressed in line 186.137 (table 3.1b) that contains the full-length *I* promoter fragment. Moreover, line 186.137 showed the ratio of the activity in the ovaries to the activity in female carcass (O/C)) of about 20, whereas all the lines carrying the 1-Δ186 fragment showed the O/C ratio of less than one ranging from about 0.4 to 0.85.

6.2.3 CAT assays for the activity of 1-138 and 186Δ101-157 fragments of the *I* promoter

The above results suggest that site 1 of the *I* factor is essential for both normal levels of expression in both ovary and carcass and the enhanced expression in the ovaries of reactive flies. Previous data have shown that nucleotides 1-100 of the *I* factor that also

Figure 6.2 Southern blot analysis of lines transformed with p1-Δ186W8

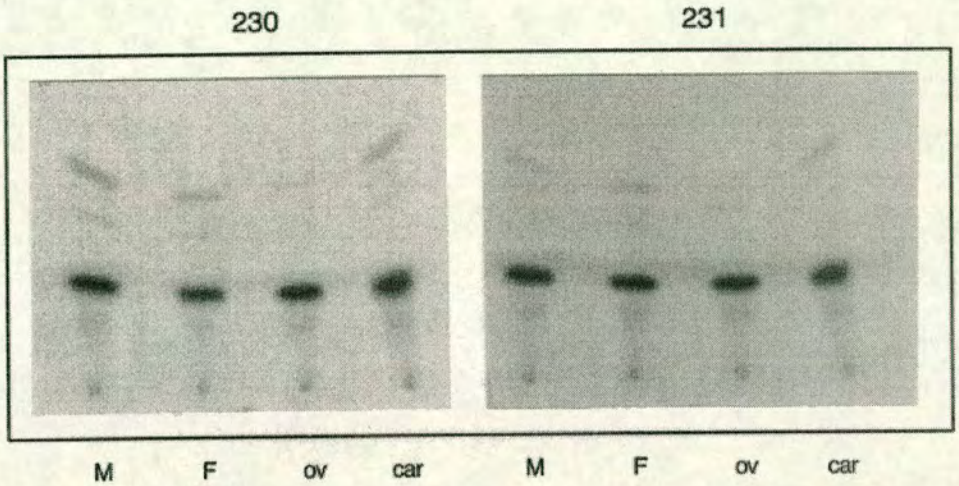
Approximately 3 μg of genomic DNA from five independent transformed lines were digested with *Hind*III. After electrophoresis and Southern transfer, the DNA was hybridised to the 250 bp *Bam*HI-*Eco*RI fragment of the CAT gene in pCAT.1. Genomic DNA of W^K (Luning 1981) and line 186.137 (McLean, *et al.* 1993) were used as negative and positive controls respectively.

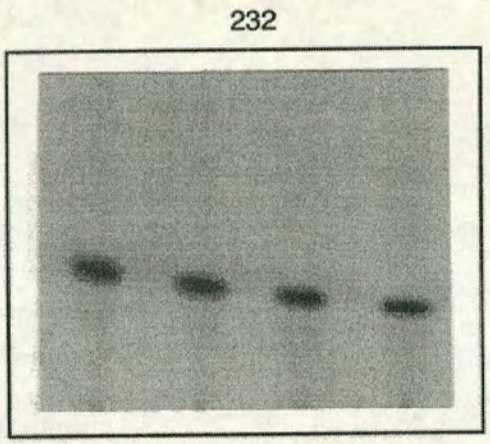
kb
23.1
9.4
6.6
4.4
186.137
WK
230
231
232
233
234

Figure 6.3 CAT assays of the tissues from lines transformed with p1-Δ186W8

Lines 230-234 containing the 1-Δ186 *I* promoter fused to the CAT gene were assayed for CAT activity. The assays were carried out with one microgram of each extract.

M - 3-4 day old males, F - 3-4 day old females, ov - ovaries, car - carcass


230
231
M
F
ov
car
M
F
ov
car


232
M
F
ov
car

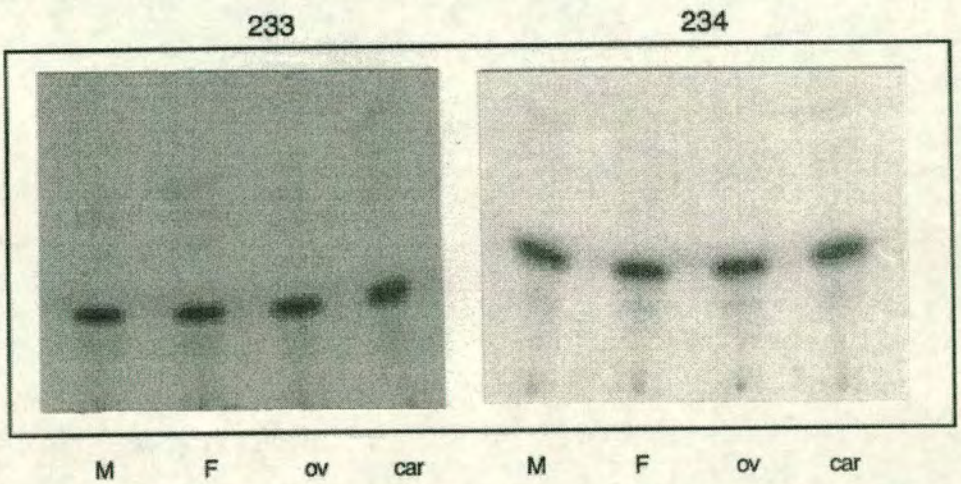

233
234
M
F
ov
car
M
F
ov
car

Table 6.1 CAT activity (pmoles/min/mg protein) of lines transformed with p1-Δ186W8

LINE	MALE	FEMALE	OVARY	CARCASS	O/C
230	12.7±1.10	8.95±0.75	3.15±0.25	9.35±0.35	0.34±0.02
231	5.30±1.04	6.83±1.01	6.33±1.08	6.95±0.77	0.90±0.08
232	2.70±0.07	1.32±0.23	1.81±0.04	2.02±0.23	0.93±0.08
233	3.08±0.31	2.85±0.13	2.20±0.32	3.38±0.30	0.65±0.08
234	6.58±0.99	4.43±0.33	4.70±0.53	5.18±0.29	0.91±0.06

The mean values and standard errors of 2 assays on one extract are shown for line 230.

The mean values and standard errors of 4 assays on 2 extracts (2 assays on each extract) are shown for lines 231-234.

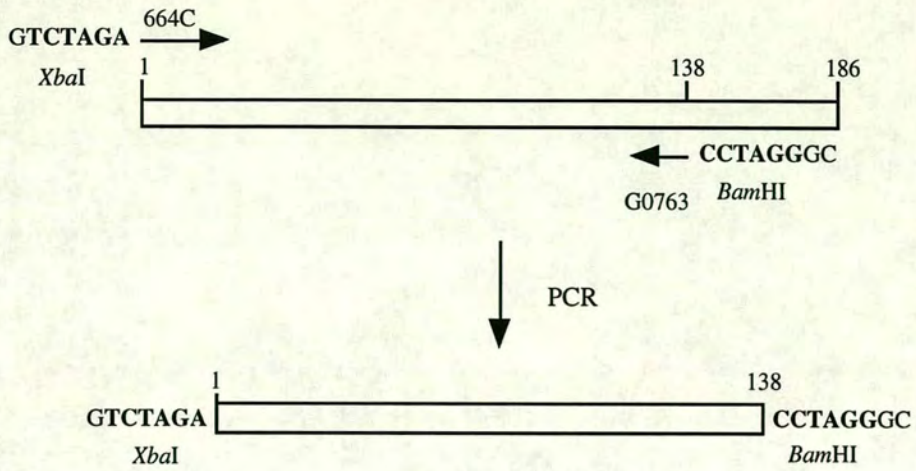
lack site 1 were able to expressed CAT activity at high levels and retained some enhancement of the CAT activities in the ovaries with the O/C ratio of about 4 (C. McLean-personal communication). The difference between the fragments 1-100 and 1-Δ186 is the presence of nucleotides 101-137 and nucleotides 158-186 in the latter. In order to determined whether or not these two sequences have any effects on the level of expression from the *I* promoter, two *P*-transformation vectors were constructed (figure 6.4 a and b) and were tested for their abilities to express the CAT gene in reactive flies. The first construct, p138w8, contains nucleotides 1-138 of the *I* factor fused to the CAT gene. The second construct, p186Δ101−157W8 has a fusion between 1-186 bp of the *I* factor from which nucleotides 101-157 were deleted and the CAT gene.

Six homozygous lines, lines 283-288 were obtained from the transformation of the W^K embryos with p138W8, each line resulted from independent insertion events. Four homozygous lines, lines 289-292, were established by transformation with p186Δ101−157W8, each line was also the result of independent insertion events. The CAT assays of lines carrying the 1-138 and 186Δ101−157 fragments were shown in figures 6.5 a and b and the values of CAT activity were summarised in table 6.2 a and b respectively. Most of the lines carrying either the 1-138 or 186Δ101−157 fragments showed relatively high levels of CAT activity in both ovarian and non-ovarian tissues as compared to those of lines carrying 1-Δ186 fragment. The average O/C ratios of lines carrying 1-138 and 186Δ101-157 fragments are 2.2 and 2.8 respectively.

The O/C ratio of line 287 is exceedingly high compared with those of the other lines carrying the same promoter fragment (1-138 bp) as a result of increased expression in ovaries and less expression in carcass. The exceptional high O/C ratio is also observed in line 290 carrying the 186Δ101-157 fragment. This indicates that the transgenes in lines 287 and 290 may be under position effect. Infact, statistical analysis shows that the O/C ratio of these two lines are significantly different from those of the other lines

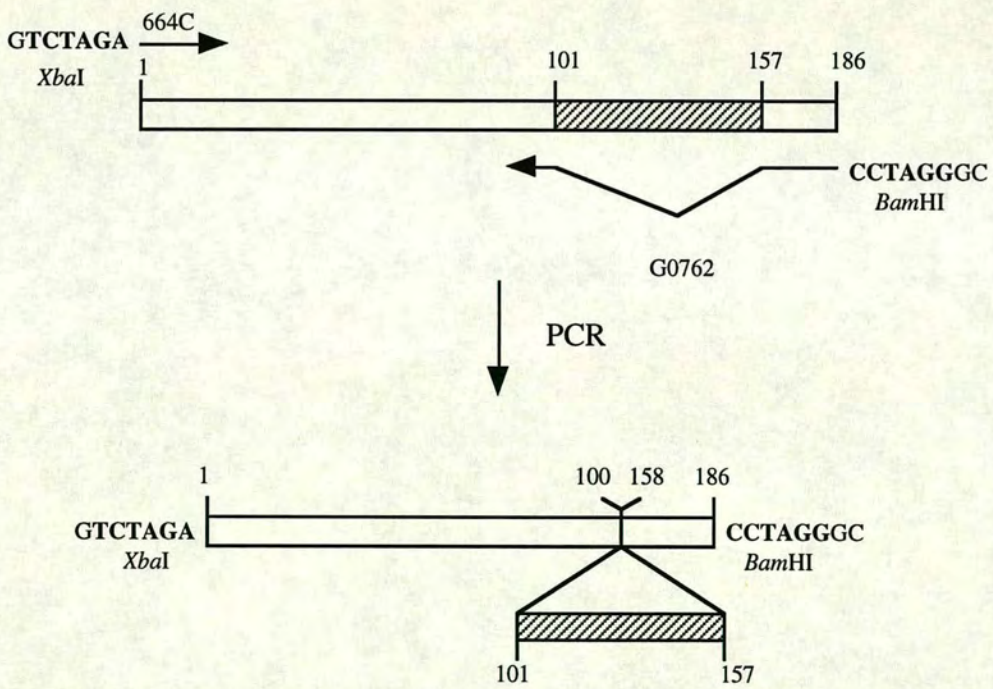
Figure 6.4 Synthesis of the 1-138 and 186Δ101-157 fragments by PCR

Plasmid p186T.1 was used as a template for PCR. Oligonucleotide 664C comprising bases 1-20 of the *I* factor with the *Xba*I site attached at the 5' end was used as a + strand primer. Oligonucleotide G0763, used as a - strand PCR primer for the 1-138 fragment, contains bases 138 to 123 of the *I* factor and has the *BamH*I site attached to base 138 at its 5' end (a). Oligonucleotide G0762 contains bases 186-90 of the *I* factor from which bases 101-157 were deleted was used as a - strand PCR primer for 186Δ101-157 fragment (b).



1-138 fragment

a



186Δ101-157
fragment

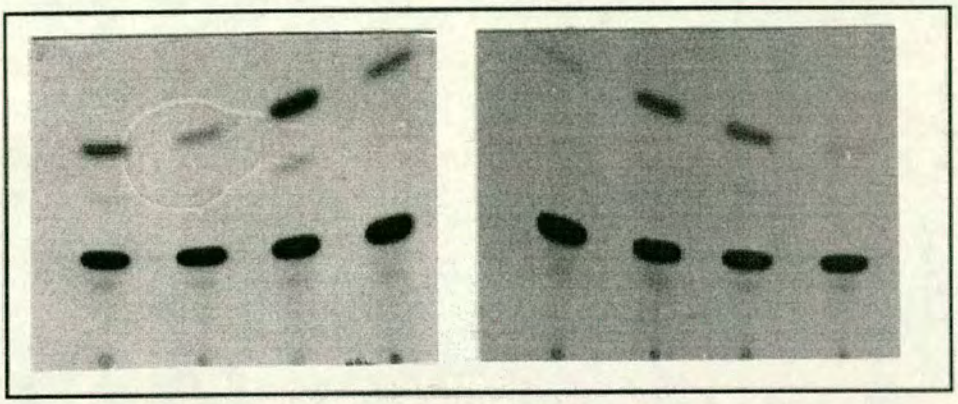
b

Figure 6.5 a CAT assays of the tissues from lines transformed with p138W8

Lines 283-288 containing the 1-138 *I* promoter fused to the CAT gene were assayed for CAT activity. The assays were carried out with one microgram of each extract.

M - 3-4 day old males, F - 3-4 day old females, ov - ovaries, car - carcass

283

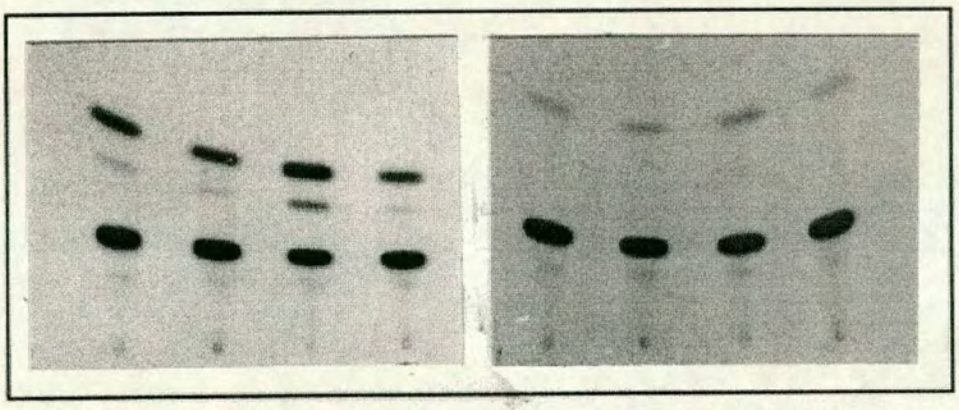


M F ov car

284

M F ov car

285

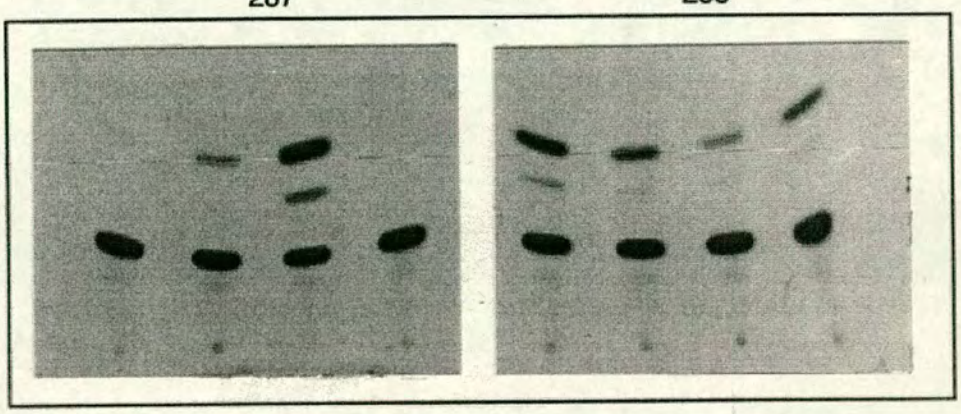


M F ov car

286

M F ov car

287



M F ov car

288

M F ov car

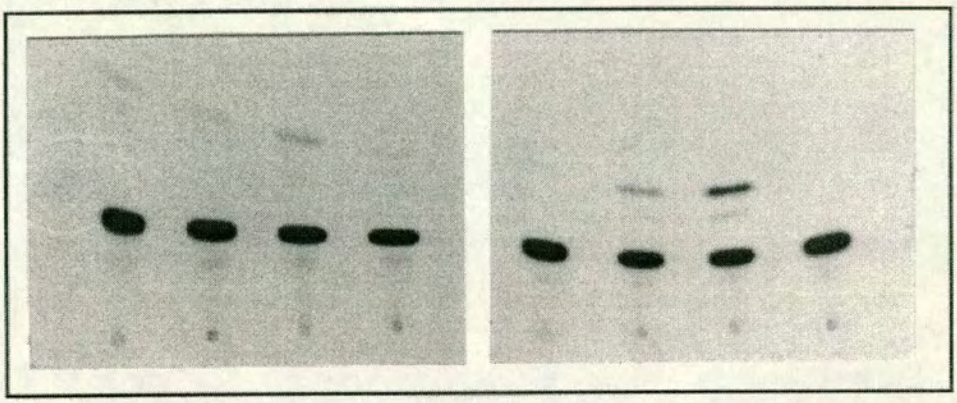
Figure 6.5 b CAT assays of the tissues from lines transformed with p186Δ101-157W8

Lines 289-292 containing the 186Δ101-157 *I* promoter fused to the CAT gene were assayed for CAT activity. The assays were carried out with one microgram of each extract.

M - 3-4 day old males, F - 3-4 day old females, ov - ovaries, car - carcass

289

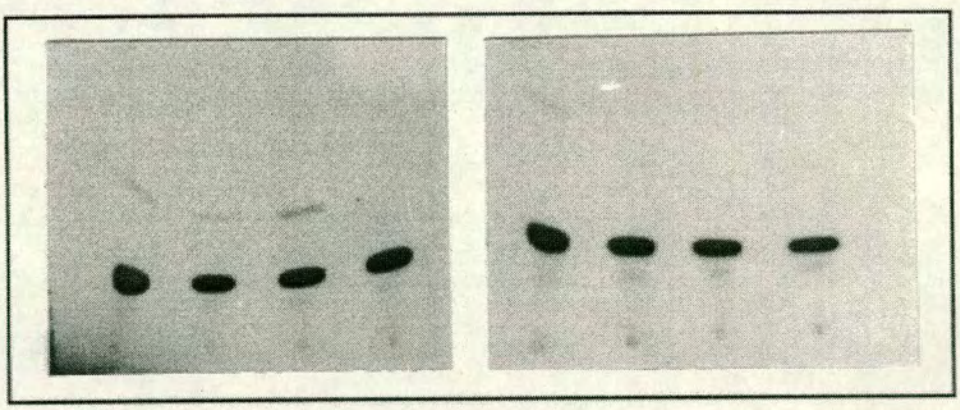
290



M F ov car M F ov car

291

292



M F ov car M F ov car

Table 6.2 a CAT activity (pmoles/min/mg protein) of lines transformed with p138W8

LINE	MALE	FEMALE	OVARY	CARCASS	O/C
283	1404±70.8	1568±333	2551±282	1371±223	2.13±0.54
284	448±52.2	950±189	1298±181	311±56.2	4.65±0.76
285	2798±710	1675±534	4443±818	1228±235	3.95±1.06
286	499±31.3	457±28.2	668±46.8	434±64.4	1.63±0.13
287	133±34.8	585±65.3	5115±189	10.3±20.8	55.5±10.3
288	1883±261	1670±377	337±131	934±292	0.39±0.03

The mean values and standard errors of 4 assays on 2 extracts (2 assays on each extract) are shown for every line.

Table 6.2 b CAT activity (pmoles/min/mg protein) of lines transformed with p186Δ101-157W8

LINE	MALE	FEMALE	OVARY	CARCASS	O/C
289	257±47.6	209±39.7	472±34.9	171±33.6	3.20±0.98
290	159±21.4	482±86.1	3302±954	110±15.5	28.8±6.05
291	159±28.7	146±21.1	298±62.9	92.8±15.7	3.50±0.83
292	187±87.5	124±28.7	85.5±17.9	127±30.9	0.70±0.09

The mean values and standard errors of 4 assays on 2 extracts (2 assays on each extract) are shown for every line.

as indicated by the P values of less than 0.05 (table A.3, appendix A). Therefore the results from these two lines could be excluded from consideration.

6.3 DISCUSSION

The results in this chapter suggest the importance of site 1 to both the normal level of expression from the *I* promoter in ovaries and other tissues of reactive flies and the enhanced expression that is specific to the ovaries. However, these results were somewhat unexpected as removal of site 1 from the *I* promoter had more severe effects on its activity than removal of the longer sequence, nucleotides 101-186. The activities from the 186 bp full-length *I* promoter dropped by a factor of about 100 when site 1 was deleted (fragment 1-Δ186) whereas the activity remained at high level when the 3' end of the 186 bp was deleted to nucleotide 100 (fragment 1-100). The ratio of the CAT activity in the ovaries to the activity in female carcass also remarkably decreased from 4 when expressed from fragment 1-100 to about 0.6 when expressed from 1-Δ186 promoter.

The low level of activity from 1-Δ186 promoter could be the consequence of repression by inhibitory factors that are able to bind sequences downstream of nucleotide 100 of the 1-Δ186 promoter, that is when site 1 is not present. The binding sites of these inhibitory factors may be masked when site 1 is occupied by an ovary specific enhancer.

However, the CAT activities of transformed lines carrying 1-138 and 186Δ101–157 fragments are significanlty greater than those of lines carrying 1-Δ186 fragment. Therefore, there was no evidence for the binding of any inhibitors to the sequences comprising nucleotides 101-138 and nucleotides 158-186 of the *I* factor.

Removal of site 1 from the promoter has placed nucleotides 138 and 158 next to each other. This might create an artificial binding site for an inhibitor that represses

expression from the *I* promoter. To investigate this possibility, a transformed line carrying the *I* promoter with mutations or base substitutions within site 1 could be assayed for CAT activity. This will provide evidence as to whether mutations or substitutions of site 1 have any effects on the activity of the *I* promoter. DNA footprinting on fragment 1-Δ186 may also determine whether or not this fragment contains an artificial binding site for proteins that could act as an inhibitor.

Another possible explanation is that deletion of site 1 from the *I* promoter affects the conformation of the DNA such that transcription factors cannot assemble or RNA Pol II cannot read through the transcription unit.

CHAPTER 7

7.1 GENERAL DISCUSSION

The work present in this thesis has identified a region that is required for specific expression of the *I* factor promoter in ovaries of reactive strain of *Drosophila melanogaster*. This region, so called site 1, encompasses nucleotides 139-157 of the *I* factor. Removal of site 1 from the *I* factor promoter had two effects on its expression, a dramatic decrease in the absolute level of activity in both ovarian and non ovarian tissues and a reduction in expression in ovaries relative to carcass. However, it has been suggested that both the fragment carrying site 1 and another element that is located within 41-99 bp of the *I* factor are required for a maximum enhancer activity (S. Forbes, personal communication). Site 1 is recognised by a sequence specific binding factor that is present, probably in different forms, in males, ovaries and female carcass of both reactive and inducer strains as well as *Drosophila* tissue culture cells.

Chaboissier *et al.*, 1990 have shown that the presence of a full length *I* factor transcript correlates with *I* factor transposition. This was the first evidence to indicate that transposition of the *I* factor is regulated at the transcription level. However, this work as well as determination of the *I* factor activity by CAT assays carried out by McLean *et al.*, 1993 has left another possibility that *I* factor activity might also be regulated post-transcriptionally at the translation level. Recent work by S. Forbes in which the *lacZ* gene was expressed from the *hsp70* promoter that was regulated by the upstream *I* factor enhancer sequences showed that the expression of *lacZ* was restricted to the germ line in the ovaries. This result together with the data from β-galactosidase activity assays to determine the strength of *I* factor enhancers present in chapter 4 of this thesis confirms that the effects on CAT expression are due to changes in the level of transcription. The RNA stability cannot account for these effects as no *I* factor sequence were present in the transcripts.

Transposition of *I* factors occurs at very low frequencies in inducer strains that contain complete *I* factors. The only condition in which high levels of *I* factor expression can be detected is in the ovaries of female progeny of a dysgenic cross. Therefore, *I* factor expression is regulated in two ways. Firstly expression is reduced in females of an inducer strain. Secondly, in reactive strains, high levels of transcription are confined to the ovaries as compared to those in female carcass and males.

In an attempt to identify sequences that are responsible for the reduced activity of the *I* factor promoter, reactive strains carrying nucleotides 1-186 and 1-100 of the *I* factor linked to the CAT gene have been converted to the inducer state and the levels of CAT activity in ovaries and carcass in the reactive and inducer backgrounds compared (C. McLean, unpublished results). The expression from nucleotides 1-186 decreased, about thirty fold, in the ovaries of inducer flies as compared to that in the ovaries of reactive flies whereas only four fold difference in the activity was observed when the CAT gene was expressed from nucleotides 1-100. This indicates that nucleotides 101-186 contain a sequence that is responsible, at least in part, for the inhibition of expression from the *I* factor promoter in an inducer strain.

When a complete *I* factor is introduced into a reactive strain there is an increase in its copy number as a result of transposition and eventually the reactive strain becomes inducer perhaps when the copy number of the *I* factor reaches a critical level (Pélisson and Bregliano 1987). This may suggest some sort of autoregulation that is controlled by products of the *I* factors themselves.

Jensen *et al.*, 1995 have recently shown that a transposition-defective *I* factor with a deletion within ORF2 is able to regulate reactivity when introduced into a highly reactive strain. The levels of reactivity of the transgenic lines were greatly reduced after 15 generations as measured by the sterility levels of female progeny of crosses between w^{1118} inducer males and transgenic females. The fertility of these females ultimately reached a level close to that of females from a non-dysgenic cross. This

result suggests that the product of ORF2 cannot account for the auto-regulation process.

The predicted amino acid sequence of ORF1 of the *I* factor contains CCHC motifs similar to those found in nucleic acid binding regions of retroviral gag polypeptides (Fawcett, *et al.* 1986) and therefore suggests its possible role as a regulatory factor that controls *I* factor activity. However, no sequence-specific binding of ORF1 protein to *I* factor DNA has been detected so far (A. Dawson-personal communication). When expressed in insect cells ORF1 protein has been found as a high molecular weight form that probably represent a virus like particle (A. Dawson-personal communication). These results suggest that ORF1 protein plays a role in the assembly of virus like particles rather than in controlling the expression of *I* factor activity in inducer strains.

As the products from neither ORF1 nor ORF2 seem to be responsible for reduced expression of the *I* factor in inducer strains, it is possible that *I* factor DNA itself may play a role in this process. Titration of a factor that is necessary for *I* factor expression may be the mechanism that explains this phenomenon (D. Finnegan-personal communication) assuming that more than one monomer of the factor is required for its function. Nucleotides 101-186 of the *I* factor contain a sequence that inhibits expression from the *I* factor promoter in inducer strains (C. McLean-personal communication). This could be site 1 that has been shown to be recognised by a sequence specific binding factor. If this is the case, the factor that regulates *I* factor expression in inducer strains could therefore be either site 1 binding protein or other proteins that interact with it.

In reactive strains site 1 binding protein is present at the level that is enough to allows transcription and transposition of the incoming complete *I* factor. Transposition results in increasing numbers of site 1 that could then gradually titrate this protein until it falls below the point that allows transcription to occur at the level that is sufficient for

transposition. By this stage, the copy numbers of complete *I* factors remains stable and a reactive strain is converted to inducer state.

The hybrid dysgenesis phenomenon that stimulates transposition of the *I* factor in the dysgenic females could be explained by a similar mechanism. The female progeny of the dysgenic cross contain about half the number of complete *I* factors that are present in the inducer parent. Therefore the ratio of site 1 to its binding protein is reduced probably to the level that is sufficient to allow transposition.

It seems that *I* factors exploit host factors to confine their activity to the germ line and also to maintain their copy number without doing any harm to the host. This is similar to the regulation of the expression of *P* elements that are responsible for P-M hybrid dysgenesis in the way that they utilise host factors for both aspects of their regulation. However, unlike *I* factors that employ a tissue specific transcription factor in restricting germ-line specific expression, tissue-specific expression of *P* elements is regulated post-transcriptionally by an alternative splicing of the third intron (IVS3) (Laski, *et al.* 1986). Removal of IVS3 allows production of transposase in the germ line while IVS3 splicing is inhibited in somatic cells. This is regulated by a negative splicing factor that is present in different amounts in somatic and germ cells (Siebel, *et al.* 1995).

The regulation of transcription in P cytotype involves a combination of transcriptional repression and reduction of third intron splicing (Roche, *et al.* 1995). Transcription repression in P cytotype results in low level of *P* element pre-mRNA. A limiting amounts of a negative splicing factor in the cell is therefore sufficient to reduce IVS3 splicing. This results in more unspliced-mRNA and thus increases production of the 66 kDa mRNA that leads to an autoregulatory loop for synthesis of the 66kDa repressor protein. This provides a mechanism to maintain repression during maternal inheritance of P cytotype.

In M cytotype high levels of *P* element pre-mRNA would titrate the negative splicing factor allowing IVS3 splicing to occur and transposase produced (Roche, *et al.* 1995). Therefore, the regulation of *P* element expression is dependent on the ratio of *P* element mRNA to the negative splicing factor. This is similar to a proposal for *I* factor regulation where the ratio of site 1 and its binding protein determines the level of its expression.

The activity of *hobo* transposable element that regulates another system of hybrid dysgenesis in *Drosophila malanogaster* has been shown to be restricted to the germ line by transcriptional control (Calvi and Gelbart 1994). Not much is known about the mechanism of *hobo* regulation so far. However, this may be different from the mechanism that regulates expression of *I* factors as high mutation rates that could be the result of *hobo* transposition were observed in both dysgenic and the reciprocal crosses (Blackman, *et al.* 1987).

Transposition of transposable elements can affect gene structure and expression in many ways indicating that they may have contributed to genetic instability and genome evolution (Finnegan 1989a). Processed pseudogene formation that could be involved in genesis of new function of the genes has been recently demonstrated to be an ongoing process that utilise the same mechanism as described for retrotransposition (Maestre, *et al.* 1995). Many questions of the effects that transposable elements can bring about and how they are maintained over the long period of evolution remain to be answered. Transposition frequencies for most transposable elements are only about 10^{-4} per generation suggesting that these elements must be highly regulated so that they do not cause a burst of evolutionary consequences. Therefore, understanding regulation of transposable elements may be one of the most important steps that contributes to the knowledge of genome evolution.

7.2 FUTURE WORK

Since expression of the *lac Z* gene from nucleotides 41-186 of the *I* factor is restricted to nurse cells in which the synthesis of proteins that are required for embryogenesis occurs, the binding factor that recognises the sequence within nucleotides 41-186, presumably site 1 binding protein, may function to stimulate expression of the genes coding for some of these products and therefore be essential for female fertility. Purification of such a factor could provide a possibility to test this idea.

As nuclear extracts from *Drosophila* tissue culture cells and whole flies showed similar patterns of *I* factor DNA binding to that seen with ovarian nuclear extract, either tissue culture or flies may be used as a source for this binding factor in order to avoid too much work on ovary dissection. Specific binding factors would then be purified by affinity chromatography. Another possibility is that site 1 DNA could be used to screen an ovary specific expression library that would lead to the gene encoding such a factor. Being able to identify this sequence specific binding factor will provide an opportunity to determine its function in controlling expression of the *I* factor and thus to propose a model for this mechanism with more details.

REFERENCES

Aksoy, S., Lalor, T. M., Martin, J., Van der Ploeg, L. and Richards, F. F. (1987) Multiple copies of a retroposon interupt spliced leader RNA genes of the Afericantrypanosome *Trypanosoma gambiense*. *EMBO J.* 6: 3819-3826

Aksoy, S., Williams, S., Chang, S. and Richards, F. F. (1990) *SLACS* retrotransposon from *Trypanosoma brucei gambiense* is similar to mammalian LINEs. *Nucleic Acids Res.* 18: 785-792

An, W. and Wensink, P. C. (1995) Three protein bindings form an enhancer that regulates sex- and fat body-specific transcription of D*rosophila yolk protein* genes. *EMBO J.* 14: 1221-1230

Arkhipova, I. R. and Ilyin, Y. V. (1991) Properties of promoter region of mdg1 *Drosophila* retrotransposon indicate that it belongs to a specific class of promoter. *EMBO J.* 10: 1169-1177

Arkhipova, I. R. and Ilyin, Y. V. (1992) Control of transcription of *Drosophila* retrotransposons. *BioEssays* 14: 161-168

Banerji, J., Rusconi, S. and Schaffner, W. (1981) Expression of a globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27: 299-308

Beaupain, D., Eleouet, J. F. and Romeo, P. H. (1990) Initiation of transcription of the erythroid promoter of the porphobilinogen deaminase gene is regulated by a *cis*-acting sequence around the cap site. *Nucleic Acids Res.* 18: 6509-6515

Biessmann, H., Champion, L. E., O'Hair, M., Ikenaga, K., Kasravi, B. and Mason, J. M. (1992a) Frequent transposition of *Drosophila melanogaster* HeT-A elements to receding chromosome ends. *EMBO J.* 11: 4459-4469

Biessmann, H., Valgeirsdottir, K., Lofssky, A., Chin, C., Ginther, B., Lewis, R. W. and Pardue, M. L. (1992b) HeT-A, a transposable element specifically involved in "healing" broken chromosome ends in *Drosophila melanogaster*. *Mol. Cell. Biol.* 12: 3910-3918

Biggin, M. D. and Tjian, R. (1988) Transcription factors that activate the *Ultrabithorax* promoter in developmentally staged extracts. *Cell* 53: 699-711

Bingham, P. M., Kidwell, M. G. and Rubin, G. M. (1982) The molecular basis of P-M hybrid dysgenesis: the role of the *P* element, a P strain-specific transposon family. *Cell* 29: 995-1004

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

Blackman, R. K., Grimaila, R., Koehler, M. M. D. and Gelbart, W. M. (1987) Mobilization of *hobo* elements residing within the decapentaplegic gene complex: suggestion of a new hybrid dysgenesis system in *Drosophila melanogaster. Cell* 49: 497-505

Bogenhagen, D. F. and Brown, D. D. (1981) Nucleotide sequence in Xenopus 5S DNA required for transcription termination . *Cell* 24: 261-270

Boliva, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H. and Falkow, S. (1977) Construction and characterisation of new cloning vehicles. II a multipurpose cloning system. *Gene* 2: 95-113

Bregliano, J. C., Picard, G., Bucheton, A., Pélisson, A., Lavige, J. M. and L'Heirtier, P. (1980) Hybrid dysgenesis in *Drosophila melanogaster. Science* 207: 606-611

Brown, M. J., Read, C. A., Roiha, H. and Glover, D. M. (1984) Site specific insertion of a type I rDNA element into a unique sequence in the *Drosophila melanogaster* genome. *Nucleic Acid Res.* 12: 9111-9122

Bucheton, A. (1978) Non-mendelian female sterility in *Drosophila melanogaster*: influence of aging and thermic treatments. I. Evidence for a partly inheritable effect of these two factors. *Heredity* 41: 357-369

Bucheton, A. (1979) Non-mendelian female sterility in *Drosophila melanogaster*: influence of age and thermic treatments. II. Action of thermic treatments on the sterility of SF females and on the reactivity of reactive females. *Biol. Cell.* 34: 43-49

Bucheton, A. (1990) *I* transposable elements and I-R hybrid dysgenesis in *Drosophila TIG* 6: 16-21

Bucheton, A., Lavige, J. M., Picard, G. and L'Heritier, P. (1976) Non-mendelian female sterility in *Drosophila melanogaster*: quantitative variations in the efficiency of inducer and reactive strains. *Heredity* 36: 305-314

Bucheton, A., Paro, R., Sang, H. M., Pélisson, A. and Finnegan, D. J. (1984) The molecular basis of I-R hybrid dysgenesis in *Drosophila melanogaster*: identification, cloning and properties of the *I* factor. *Cell* 38: 153-163

Bucheton, A. and Picard, G. (1978) Non-mendelian female sterility in *Drosophila melanogaster*: hereditary transmission of reactivity levels. *Heredity* 40: 207-233

Buratowski, S. (1994) The basics of basal transcription by RNA polymerase II. *Cell* 77: 1-3

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

Burton, F. H., Loeb, D. D., Voliva, C. F., Martin, S. L., Edgell, M. H. and Hutchison III, C. A. (1986) Conservation throughout mammalia and extensive protein-encoding capacity of the highly repeated DNA long interspersed sequence one. *J. Mol. Biol.* 187: 291-304

Busseau, I., Pélisson, A. and Bucheton, A. (1989) Characterization of 5' truncated transposed copies of the *I* factor in *Drosophila melanogaster. Nucleic Acids Res.* 17: 6939-6945

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

Calvi, B. R., Hong, T. J., Findley, S. D. and Gelbart, W. M. (1991) Evidence for a common evolutionary origin of inverted repeat transposons in Drosophila and plants: hobo, Activator and Tam3. *Cell* 66: 465-471

Campuzano S., Balcells, L., Vallares, R., Carramolino, L., Garcia-Alanso, L. and Modelell, J. (1986) Excess function *Hairy-wing* mutations caused by *gypsy* and *copia* transposable elements inserted within structural genes of the achaete-scute locus of *Drosophila. Cell* 44: 303-312

Chaboissier, M. C., Busseau, I., Prosser, J., Finnegan, D. J. and Bucheton, A. (1990) Identification of a potential RNA intermediate for transposition of the LINE-like element *I* factor in *Drosophila melanogsaster. EMBO J.* 9: 3557-3563

Chain, A. C., Zollman, S., Tseng, J. C. and Laski, F. A. (1991) Identification of a *cis*-acting sequence required for germline specific splicing of the P element ORF2-ORF3 intron. *Mol. Cell. Biol.* 11: 1538-1546

Charlesworth, B., Lapid, A. (1989) A study of ten families of transposable elements on X chromosomes from a population of *Drosophila melanogaster. Genet. Res. Camb.* 54: 113-125

Cohen, J., Eccleshall, T., Needleman, H., Federoff, H., Buchferer, B. and Marmur, J. (1980) Functional expression in yeast of *Escherichia coli* plasmid gene coding for chloramphenicol acetyltransferase. *Proc. Natl. Acad. Sci. USA* 77: 1078-1082

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) Promoter sequences of eukaryotic protein coding genes. *Science* 209: 1406-1414

Covey, S. N. (1986) Amino acid sequence homology in *gag* region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus. *Nucleic Acids Res.* 14: 623-633

Craig, N. L. (1990) *P* element transposition. *Cell* 62: 399-402

Crouse, G. F., Leys, E. J., McEwan, R. N., Frayme, E. G. and Kellems, R. E. (1985) Analysis of the mouse DHFR promoter region: existance of a divergently transcribed gene. *Mol. Cell. Biol.* 5: 1847-1858

Crozatier, M., Vauey, C., Busseau, I., Pélisson, A. and Bucheton, A. (1988) Structure and genomic organization of *I* Elements involved in I-R hybrid dysgenesis in *Drosophila melanogaster. Nucleic Acids Res.* 16: 9199-9212

D'Ambrosio, E., Waitzkin, F. R., Witney, F. R., Salemme, A. and Furano, A. V. (1986) Structure of the highly repeated, long interspersed DNA family (LINE or L1Rn) of the rat. *Mol. Cell. Biol.* 6: 411-424

Daniels, S. B., Clark, S. H., Kidwell, M. G. and Chovnick, A. (1987) Genetic transformation of *Drosophila melanogaster* with an autonomous P element:

phenotypic and molecular analyses of long-established transformed lines. *Genetics* 115: 711-723

Danilevskaya, O., Slot, F., Pavlova, M. and Pardue, M. L. (1994) Structure of the *Drosophila* HeT-A transposon: a retrotransposon-like element forming telomeres. *Chromosoma* 103: 215-224

Dawid, I. B., Long, E. O. and Di nocera, P. P. (1981) Ribosomal insertion-like elements in *Drosophila melanogaster* are interspersed with mobile sequence. *Cell* 25: 399-408

Dawid, I. B. and Rebbert, M. L. (1981) Nucleotide sequence at the boundaries between gene and insertion regions in the rDNA of *Drosophila melanogaster. Nucleic Acids Res.* 9: 5011-5020

Demers, G. W., Brech, K. and Hardison, R. C. (1986) Long interspersed L1 repeats in rabbit DNA are homologouse to L1 repeats of rodents and primates in an open reading frame region. *Mol. Biol. Evol.* 3: 179-190

Deragon, J-M., Sinnett, D. and Labuda, D. (1990) Reverse transcriptase activity from human embryonal carcinoma cells NTera2D1. *EMBO J.* 9: 3363-3368

Di Nocera, P. P. (1988) Close relationship between non-viral retroposons in *Drosophila melanogaster. Nucleic Acids Res.* 16: 4041-4052

Di Nocera, P. P. and Dawid, I. B. (1983) Interdigitated arrangements of two oligo-A terminated DNA sequences in *Drosophila. Nucleic Acids Res.* 11: 5475-5482

Di Nocera, P. P., Digar, M. E. and Dawid, I. B. (1983) A family of oligo-adenylated terminated transposable sequences in *Drosophila melanogaster. J. Mol. Biol.* 168: 715-727

Di Nocera, P. P., Graziani, F. and Lavorgna, G. (1986) Genomic and structural organisation of *Drosophila melanogaster* G elements. *Nucleic Acids Res.* 14: 675-691

Dombroski, B. A., Mathias, S. L., Nanthakumar, E., Scott, A. F. and Kazazian, J., H. H. (1991) Isolation of an active human transposable element. *Science* 254: 1805-1808

Driver, A., Lacey, S. F., Cullingford, T. E., Mitchelson, A. and O'Hare, K. (1989) Structure analysis of *Doc* transposable elements associated with

mutations at the w*hite* and *suppressor of forked* loci of *Drosophila melanogaster*. *Mol. Gen. Genet.* 220: 49-52

Dudler, R. and Travers, A. A. (1984) Upstream elements necessary for optimal function of the hsp70 promoter in transformed flies. *Cell* 38: 391-398

Dynan, W. S., Sazer, S., Tjian, R. and Schimke, R. T. (1986) Transcription factor Sp1 recognizes a DNA sequence in the mouse dihydrofolate reductase promoter. *Nature (London)* 319: 246-248

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

Engels, W. R. (1979a) Extrachromosomal control of mutability in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 76: 4011-4015

Engels, W. R. (1979b) Hybrid dysgenesis in *Drosophila melanogaster*: rules of inheritance of female sterility. *Genet. Res. Camb.* 33: 219-236

Engels, W. R. (1983) The P family of transposable elements in *Drosophila*. *Ann. Rev. Genet.* 17: 315-344

Engels, W. R. (1984) A *trans*-acting product needed for P factor transposition in *Drosophila*. *Science* 226: 1194-1196

Engels, W. R. (1988) P elements in Drosophila. in *Mobile DNA*. eds. D. E. Breg and M. M. Howe. American Society for Microbiology, Washington, DC:

Engels, W. R., Johnson-Schlitz, D. M., Eggleston, W. B and Sved, J. (1990) High frequency P element loss in Drosophila is homolog dependent. *Cell* 62: 515-525

Evans, J. P. and Palmiter, R. D. (1991) Retrotransposition of a mouse L1 element. *Proc. Natl. Acad. Sci. USA* 88: 8792-8795

Fanning, T. and Singer, M. F. (1987a) The LINE-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retrovirus proteins. *Nucleic Acids Res.* 15: 2251-2260

Fanning, T. and Singer, M. F. (1987b) LINE-1: a mammalian transposable element. *Biochim. Biophys. Acta* 910: 203-212

Fanning, T. G. (1983) Size and structure of the highly repetitive BAM HI element in mice. *Nucleic Acids Res.* 11: 5073-5091

Fawcett, D. H., Lister, C. K., Kellett, E. and Finnegean, D. J. (1986) Transposable elements controlling I-R hybrid dysgenesis in *D. melanogaster* are similar to mammalian LINEs. *Cell* 47: 1007-1015

Feinberg, A. P. and Vogelstein, B. (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132: 6-13

Finnegan, D. J. (1988) I-factors in *Drosophila melanogaster* and similar elements in other eukaryotes. in *Transposition.* eds. A. J. Kingsman, K. F. Chater and S. M. Kingsman. Cambridge University Press, Cambridge: 271-285

Finnegan, D. J. (1989a) Eukaryotic transposable elements and genome evolution. *TIG* 5: 103-107

Finnegan, D. J. (1989b) The I factor and I-R hybrid dysgenesis in *Drosophila melanogaster.* in *Mobile DNA.* eds. D. E. Berg and M. M. Howe. American Society for Microbiology, Washington, DC.: 503-517

Finnegan, D. J. (1990) Transposable elements and DNA transposition in eukaryotes. *Current Opinion in Cell Biology* 2: 471-477

Fischer, J. A. and Maniatis, T. (1986) Regulatory elements involved in Drosophila *Adh* gene expression are conserved in divergent species and separate elements mediate expression in different tissue. *EMBO J.* 5: 1275-1289

Fried, M. (1989) Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay. *Electrophoresis* 10: 366-376

Fujiwara, H., Ogura, T., Takada, N., Miyajima, N., Ishikawa, H. and Maekawa, H. (1984) Intron and their flanking sequences of *Bombyx mori* rDNA. *Nucleic Acids Res.* 12: 6861-6869

Furano, A. V., Robb, S. M. and Robb, F. T. (1988) The structur of the regulatory region of the rat L1 (L1Rn, long interspersed repeated) DNA family of transposable elements. *Nucleic Acids Res.* 16: 9215-9231

Gabriel, A. and Boeke, J. D. (1991) Reverse transcriptase encoded by a retrotranposon from the trypanpsomatid *Crithidia fasciculata. Proc. Natl. Acad. Sci. USA* 88: 9794-9798

Gabriel, A., Yen, T. J., Schwart, D. C., Smith, C. L., Boeke, J. D., Sollner-Webb, B. and Cleveland, D. W. (1990) A rapidly rearranging retrotranposon within the miniexon gene locus of *Crithidia fasciculata. Mol. Cell. Biol.* 10: 615-624

Galli, G., Hofsyetter, H. and Birnateil, M. L. (1981) Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. *Nature (London)* 294: 626-631

Gerster, T., matthias, P., Thali, M., Jiricny, J. and Schaffner, W. (1987) Cell-type-specificity elements of the immunoglobulin heavy chain gene enhancer. *EMBO J.* 6: 1323-1330

Gillies, S. D., Morrison, S. L., Oi, V. T. and Tonegawa, S. (1983) A tissue-specific transcription enhancer is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* 33: 717-728

Goodbourn, S., Burstein, H. and Maniatis, T. (1986) The human beta-interferon gene enhancer is under negative control. *Cell* 45: 601-610

Goodbourn, S., Zinn, K. and Maniatis, T. (1985) Human beta-interferon gene expression is regulated by an inducible enhancer element. *Cell* 41: 509-520

Grimaldi, G. and Singer, M. F. (1983) Members of the KpnI family of long interspersed repeated sequences join and interrupt α-satellite in the monkey genome. *Nucleic Acids Res.* 11: 321-338

Handler, A. M. and Gomez, S. P. (1995) The *hobo* transposable element has transposase-dependent and -independent excision activity in drosophilid species. *Mol. Gen. Genet.* 247: 399-408

Hasan, G., Turner, M. J. and Cordingley, J. S. (1982) Ribosomal RNA genes of trypanosoma brucei. Cloning of a rRNA gene containing a mobile element. *Nucleic Acids Res.* 10: 6747-6767

Hattori, M., Kuhara, S., Takenaka, O. and Sakaki, Y. (1986) L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein. *Nature (London)* 321: 625-628

Hohjoh, H., Minakami, R. and Sakaki, Y. (1990) Selective cloning and sequence analysis of the human L1 (LINE-1) sequences which transposed in the recent past. *Nucleic Acids Res.* 18: 4099-4104

Holmes, S. E., singer, M. F. and Swergold, G. D. (1992) Study on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element. *J. Biol. Chem.* 267: 19765-19768

Hutchison, C. A., Hardies, S. C., Loeb, D. D., Shehee, W. R. and Edgell, M. H. (1989) LINEs and related retroposons: long interspersed repeated sequences in the eucaryotic genome. in *Mobile DNA*. eds. D. E. Berg and M. M. Howe. American Society for Microbiology, Wachington, DC: 593-617

Ingolia, T. D., Craig, E. A. and McCarthy, B. J. (1980) Sequences of three copies of the gene for the major *Drosophila* heat shock induced protein and their flanking regions. *Cell* 21: 669-679

Ivanov, V. A., Melnikov, A. A., Siunov, A. V., Fodor, I. I. and Ilyin, Y. V. (1991) Authentic reverse transcriptase is coded by *jockey*, a molbile *Drosophila* element related to mammalian LINEs. *EMBO J.* 10: 2489-2495

Jakubczak, J. L., Burke, W. D. and Eickbush, T. H. (1991) Retrotransposable elements *R1* and *R2* interrupt the rRNA genes of most insects. *Proc. Natl. Acad. Sci. USA* 88: 3295-3299

Jakubczak, J. L., Xiong, Y. and Eickbush, T. H. (1990) Type I (R1) and type II (R2) ribosomal DNA insertions of *Drosophila melanogaster* are retrotransposable elements closely related to those of *Bombyx mori*. *J. Mol. Biol.* 212: 37-52

Jarrel, K. A. and Meselson, M. (1991) *Drosophila* retrotransposon promoter includes essential sequences at the initiation site and requires a downstream sequence for full activity. *Proc. Natl. Acad. Sci. USA* 88: 102-104

Jensen, S. and Heidmann, T. (1991) An indicator gene for detection of germline retrotrnsposition in transgenic *Drosophila* demonstrates RNA-mediated transposition of the LINE *I* element. *EMBO J.* 10: 1927-1937

Karess, R. E. and Rubin, G. M. (1984) Analysis of P transposable element functions in *Drosophila melanogaster*. *Cell* 38: 135-146

Katzir, N., Rechavi, G., Cohen, J. B., Unger, T., Simoni, F., Segal, S., Cohen, D. and Givol, D. (1985) "Retroposon" insertion into the cellular oncogene c-myc in canine transmissable venereal tumor. *Proc. Natl. Acad. Sci. USA* 82: 1054-1058

Kaufman, P. D., Doll, R. F. and Rio, D. C. (1989) *Drosophila* P element transposase recognizes internal P element DNA sequences. *Cell* 59: 359-371

Kazazian Jr., H. H., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G. and Antonarakis, S. (1988) Haemophilia A resulting from *de novo* indertion of L1 sequences represents a novel mechanism for mutation in man. *Nature (London)* 332: 164-166

Kidwell, M. G., Kidwell, J. F. and Sved, J. A. (1977) Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics* 86: 813-833

Kim, A. I. and Belayaeva, E. S. (1991) Transposition of mobile elements *gypsy*, (*MDG4*), and hobo in germ-line and somatic cells of genetically unstable mutator strains of Drosophila melanogaster. *Mol. Gen. Genet.* 229: 437-444

Kimmel, B. E., Ole-Moiyoi, O. K. and Young, J. R. (1987) Ingi, a 5.2 kb dispersed sequence element from *Trypanosoma brucei* that carries half of smaller mobile element at either end and has homology with mammalian LINEs. *Mol. Cell. Biol.* 7: 1465-1475

Kinsey, J. A. (1989) Restricted distribution of the Tad transposon in strains of *Neurospora. Curr. Genet.* 15: 271-275

Kinsey, J. A. (1990) Tad, a LINE-like transposable element of *Neurospora*, can transpose between nuclei in heterokaryons. *Genetics* 126: 317-323

Kinsey, J. A. (1993) Transnuclear retrotranposition of the Tad element of *Neurospora. Proc. Natl. Acad. Sci. USA* 90: 9384-9387

Kinsey, J. A. and Helber, J. (1989) Isolation of a transposable element from *Neurospora crassa. Proc. Natl. Acad. Sci. USA* 86: 1929-1933

Klenmenz, R., Weber, U. and Gehring, W. J. (1987) The *white* gene as a marker in a new *P*-element vector for gene transfer in *Drosophila*. Nucleic Acids Res. 15: 3947-3957

Lachaume, P., Bouhidel, K., Mesure, M. and Pinon, H. (1992) Spatial and temporal expression of the *I* factor during oogenesis in *Drosophila melanogaster*. *Development* 115: 729-735

Lachaume, P. and Pinon, H. (1993) Germ-line expression of the *I* factor, a functional LINE from the fruit fly *Drosophila melanogaster*, is positively regulated by reactivity, a peculiar cellular state. *Mol. Gen. Genet.* 240: 277-285

Laski, F. A., Rio, D. C. and Rubin, G. M. (1986) Tissue specificity of *Drosophila* P element transposition is regulated at the level of mRNA splicing. *Cell* 44: 7-19

Lavige, J. M. (1986) I-R system of hybrid dysgenesis in *Drosophila melanogaster*: further data on the arrest of development of the embryos from SF females. *Biol. Cell.* 56: 207-216

Leblanc, B and Moss, T. (1994) DNAse I footprinting in Methods in molecular biology volume 30: DNA-protein interactions 'principles and protocols' ed. Kneale, G. G. Humana Press Inc., Totawa, New Jersy: 1-10

Leibold, D. M., Swergold, G. D., Singer, M. F., Thayer, R. E., Dombroski, B. A. and Fanning, T. G. (1990) Translation of LINE-1 DNA elements *in vitro* and in human cells. *Proc. Natl. Acad. Sci. USA* 87: 6990-6994

Lemaitre, B. and Coen, D. (1991) P regulatory products repress in vivo the P promoter activity in P-lacZ fusion genes. *Proc. Natl. Acad. Sci. USA* 88: 4419-4423

Lerman, M. I., Thayer, R. E. and Singer, M. F. (1983) KpnI family of long interspersed repeated DNA sequences in primates: Polymorphism of family members and evidence for transcription. *Proc. Natl. Acad. Sci. USA* 80: 3966-3970

Levis, R. W., Ganesan, R., Houtchens, K., Tolar, L. A. and Sheen, F. (1993) Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell* 75: 1083-1093

Loeb, D. D., Padgett, R. W., Hardies, S. C., Shehee, W. R., Comer, M. B., Edgell, M. H. and Hutchison III, C. A. (1986) The sequence of a large L1Md element reveals a tandemly repeated 5' end and several features found in retrotransposons. *Mol. Cell. Biol.* 6: 168-182

Logan, S. K. and Wensink, P. C. (1990) Ovarian follicle cell enhancers from the Drosophila yolk protein genes: different segments of one enhancer have differnet cell-type specificities that interact to give normal expression. *Gen. Dev.* 4: 613-623

Luning, K. G. (1981) Genetics of inbred *Drosophila melanogaster. Hereditas* 95: 181-188

Mack, D. H., Vaqrtikar, J., Pipas, J. M. and Laimins, L. A. (1993) Specific repression of TATA-mediated but not initiator-mediated transcription by wild-type p53. *Nature (London)* 363: 281-283

Maestre, J., Tchenio, T., Dhellin, O. and Heidmann, T. (1995) mRNA retroposition in human cells: processed pseudogene formation. *EMBO J.* 14: 6333-6338

Maldonado, E., Ha, I., Cortes, P., Weis, L. and Reinberg, D. (1990) Factors involved in specific transcription by mammalian RNA polymerase II: role of transcriptin factors IIA, IID and IIB during formation of transcription-component complex. *Mol. Cell. Biol.* 10: 6334-6347

Maniatis, T., Goodbourn, S. and Fischer, J. A. (1987) Regulation of inducible and tissue-specific gene expression. *Science* 230: 1237-1244

Mathias, S. L., Scott, A. F., Kazazian, J., H. H., Boeke, J. D. and Gabriel, A. (1991) Reverse transcriptase encoded by a human transposable element. *Science* 254: 1808-1810

McGinnes, W., Shermoen, A. W. and Beckendorf, S. K. (1983) A transposable element inserted just 5' to a *Drosophila* glue protein gene alters gene expression and chromatin structure. *Cell* 34: 75-84

McLean, C., Bucheton, A. and Finnegan, D. J. (1993) The 5' untranslated region of the *I* factor, a long interspersed nuclear element-like retrotransposon of *Drosophila melanogaster*, contains an internal promoter and sequences that regulate expression. *Mol. Cell. Biol.* 13: 1042-1050

McMillan, J. P. and Singer, M. F. (1993) Translation of the human LINE-1 element, L1Hs. *Proc. Natl. Acad. Sci. USA* 90: 11533-11537

Melton, D. W., Koneck, D. S., Brennand, J. and Caskey, C. T. (1984) Structure, expression, and mutation of the hypoxanthine phosphoribosyltransferase gene. *Proc. Natl. Acad. Sci. USA* 81: 2147-2151

Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomia, J., Kinzler, K. W., Vogelstein, B. and Nakamura, Y. (1992) Disruption of the APC gene by a retrotransposable insertion of L1 sequence in a colon cancer. *Cancer Res.* 52: 643-645

Miklos, G. L. G., Yamamoyo, M. T., Davies, J. and Pirrotta, V. (1988) Microcloning reveals a high frequency of repetitive sequences characteristic of chromosome 4 and the β-heterochromatin of *Drosophila melanogaster. Proc. Natl. Acad. Sci. USA* 85: 2051-2055

Minakami, R., Kurose, K., Etoh, K., Furuhata, Y., Hattori, M. and Sakaki, Y. (1992) Identification of an internal *cis*-element essential for the human L1 transcription and a nuclear factor(s) binding to the element. *Nucleic Acids Res.* 20: 3139-3145

Minchiotti, G. and Di Nocera, P. P. (1991a) Convergent transcription initiated from oppositely orientated promoters within the 5' end region of *drosophila melanogaster* F elements. *Mol. Cell. Biol.* 11: 5171-5180

Minchiotti, G. and Di Nocera, P. P. (1991b) convergent transcription initiates from oppositely oriented promoters within the 5' end regions of *Drosophila melanogaster* F elements. *Mol. Cell. Biol.* 11: 5171-5180

Misra, S. and Rio, D. C. (1990) Cytotype control of the *Drosophila* P element transposition: the 66 kd protein is a repressor of transposase activity. *Cell* 62: 269-284

Mizrokhi, L. J., Georgieva, S. G. and Ilyin, Y. V. (1988) Jockey, a mobile Drosophila elements similar to mammalian LINEs, is transcribed from an internal promoter by RNA polymerase II. *Cell* 54: 685-691

Mizrokhi, L. J., Obolenkova, L. A., Priimagi, A. F., Illyin, Y. V., Gerasimova, T. I. and Georgiev, G. P. (1985) The nature of unstable insertion mutations and reversions in the locus *cut* of *Drosophila melanogaster*: molecular mechanism of transposition memory. *EMBO J.* 4: 3781-3787

Morse, B., Rotherg, P. G., South, V. J., Spandorfer, J. M. and Astrin, S. M. (1988) Insertional mutagenesis of the *myc* locus by a LINE-1 sequence in a human breast carcinoma. *Nature (London)* 333: 87-90

Mullins, M. C., Rio, D. C. and Rubin, G. M. (1989) *Cis*-acting DNA sequemce requirements for P-element transposition. *Gen. Dev.* 3: 729-738

Murphy, N. B., Pays, A., Tebabi, P., Coquelet, H., Guyaux, M., Steinert, M. and Pays, E. (1987) *Trypanosoma brucei* repeated elementswith unusual structural and transcriptional properties. *J. Mol. Biol.* 195: 855-871

Nakajima, N., Horikoshi, M. and Roeder, R. G. (1988) Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box-promoter interaction of TFIID. *Mol. Cell. Biol.* 8: 4028-4033

Nur, I., Pascale, E. and Furano, A. V. (1988) The left and of rat L1 (L1Rn, long interspersed repeated) DNA which is a CpG island can function as a promoter. *Nucleic Acids Res.* 16: 9233-9251

O'Hare, K., Alley, M. R. K., Cullingford, T. E., Driver, A. and Sanderson, M. J. (1991) DNA sequence of the *Doc* retroposon in the white-one mutant of *Drosophila melanogaster* and of secondary insertions in the phenotypically altered derivatives white-honey and white-eosin. *Mol. Gen. Genet.* 225: 17-24

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

Padgett, R. W., Hutchison, C. A., III and Edgell, M. H. (1988) The F-type 5' motif of mouse L1 elements: a major class of L1 termini similar to the A-type in organization but unrelated in sequence. *Nucleic Acids Res.* 16: 739-749

Pardue, M. L. and Dawid, I. B. (1981) Chromosomal location of two DNA segments that flank ribosomal insertion like sequences are mobile elements. *Chromosoma* 83: 29-43

Parker, C. S. and Topol, J. (1984) A *Drosophila* DNA polymerase II transcription factor binds to the regulatory site of an *hsp70* gene. *Cell* 37: 273-283

Pélisson, A. (1981) The I-R system of hybrid dysgenesis in *drosophila melanogaster*: are I factor insertions responsible for the mutator effect of the I-R interaction? *Mol. Gen. Genet.* 183: 123-129

Pélisson, A. and Bregliano, J. C. (1981) The I-R system of hybrid dysgenesis in *Drosophila melanogaster*: construction and characterization of a non-inducer stock. *Biol. Cell.* 40: 159-164

Pélisson, A. and Bregliano, J. C. (1987) Evidence for rapid limitation of the I element copy number in a genome submitted to several generations of I-R hybrid dysgenesis in *Drosophila melanogaster*. *Mol. Gen. Genet.* 207: 306-313

Pélisson, A., Finnegan, D. J. and Bucheton, A. (1991) Evidence for retrotransposition of the *I* factor, a LINE element of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* 88: 4907-4910

Perkins, K. K., Dailey, G. M. and Tjian, R. (1988) In vitro analysis of the Antennapedia P2 promoter: identification of a new *Drosophila* transcription factor. *Gene & Dev.* 2: 1615-1626

Picard, G. (1976) Non-mendelian female sterility in *Drosophila melanogaster*: heredity and transmission of *I* factor. *Genetics* 83: 107-123

Picard, G. (1978) Non-mendelian female sterility in *Drosophila melanogaster*: sterilituy in the daughter progeny of SF and RSF females. *Biol. Cell.* 31: 235-244

Picard, G. (1979) Non-mendelian female sterility in *Drosophila melanogaster*: principal characteristics of chromosomes from inducer and reactive origin after chromosomal contamination. *Genetics* 91: 455-471

Picard, G., Bregliano, J. C., Bucheton, A., Lavige, J. M., Pélisson, A. and Kidwell, M. G. (1978) Non-mendelian female sterility and hybrid dysgenesis in *Drosophila melanogaster*. *Genet. Res. Camb.* 32: 275-287

Picard, G., Lavige, J. M., Bucheton, A. and Bregliano, J. C. (1977) Non-mendelian female sterility in *Drosophila melanogaster*: physiological pattern of embryo lethality. *Biol. Cell.* 29: 89-98

Picard, G. and Pélisson, A. (1979) Non-mendelian female steriliry in *Drosophila melanogaster*: characterization of the non-inducer chromosomes of inducer strains. *Genetics* 91: 473-489

Preston, C. R. and Engels, W. R. (1984) Movement of *P* elements within a P strain. *Dros. Inf. Serv.* 60: 169-170

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

Pritchard, M. A., Dura, J.-M., Pélisson, A., bucheton, A. and Finnegen, D. J. (1988) A cloned *I* factor is fully functional in *Drosophila melanogaster*. *Mol. Gen. Genet.* 214: 533-540

Proust, J. and Prudhommeau, C. (1982) Hybrid dysgenesis in *Drosophila melanogaster* I. Further evidence for and characterization of the mutator effect of the inducer-reactive interaction. *Mutation Res.* 95: 225-235

Ptashne, M. (1986) Gene regulation by proteins acting nearby and at a distance. *Nature (London)* 322: 697-701

Pugh, B. F. and Tjian, R. (1991) Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Gen. Dev.* 5: 1935-1945

Rae, P. M. M., Kohorn, B. D. and Wade, R. P. (1980) The 10 kb *drosophila virilis* 28S rDNA intervening sequence is flanked by a direct repeat of 14 base pairs of coding sequence. *Nucleic Acids Res.* 8: 3491-3505

Rio, D. C. (1991) Regulation of *Drosophila* P element transposition. *TIG* 7: 282-287

Rio, D. C., Laski, F. A. and Rubin, G. M. (1986) Identification and immunochemical analysis of biologically active *Drosophila* P element transposase. *Cell* 44: 21-32

Rio, D. C. and Rubin, G. M. (1988) Identification and purification of a *Drosphila* protein that binds to the 31-base pair inverted repeats of the P transposable element. *Proc. Natl. Acad. Sci. USA* 85: 8929-8933

Robertson, H. M. and Engels, W. R. (1989) Modified *P* elements that mimic the P cytotype in *Drosophila melanogaster*. *Genetics* 123: 815-824

Roche, S. E., Schiff, M. and Rio, D. C. (1995) P-element repressor autoregulation involves germ-line transcriptional repression and reduction of third intron splicing. *Gen. Dev.* 9: 1278-1288

Roiha, H., Rubin, G. M. and O'Hare, K. (1988) *P* element insertions and rearrangements at the *singed* locus of *drosophila melanogaster*. *Genetics* 119: 75-83

Ronaldson, E. and Bownes, M. (1995) Two independent *cis*-acting elements regulate the sex-specific and tissue-specific expression of *yp3* in *Drosophila melanogaster*. *Genet. Res. Camb.* 66: 9-17

Ronsseray, S. Lemaitre, B. and Coen, D. (1993) Maternal inheritance of P cytotype in Drosophila melanogaster: a "pre-P cytotype" is strictly extra-chromosomally transmitted. *Mol. Gen. Genet.* 241: 115-123

Rubin, G. M., Kidwell, M. G. and Bingham, P. M. (1982) The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. *Cell* 29: 987-994

Rubin, G. M. and Spradling, A. C. (1982) Genetic transformation of *Drosophila* with transposable element vectors. *Science* 218: 348-353

Sang, H. M., Pélisson, A., Bucheton, A. and Finnegan, D. J. (1984) Molecular lesions associated with *white* gene mutations induced by I-R hybrid dysgenesis in *Drosophila melanogaster*. *EMBO J.* 3: 3079-3085

Schneider, I. (1972) Cell lines derived fromlate embryonic stages of *Drosophila melanogaster*. *J. Embryol. Exp. Morph.* 27: 353-367

Schwarz-Sommer, Z., Leclercq, L. and Saedler, H. (1987) Cin4, an insert altering the structure of the A1 gene in *Zea mays*, exhibits properties of nonviral retrotransposons. *EMBO J.* 6: 3873-3880

Searle, P. E., Stuart, G. W. and Palmiter, R. D. (1985) Building a metal-responsive promoter with synthetic regulatory elements. *Mol. Cell. Biol.* 5: 1480-1489

Sehgal, A., Patil, N. and Chao, M. (1988) A constitutive promoter directs expression of the nerve growth factor receptor gene. *Mol. Cell. Biol.* 8: 3160-3167

Sen, R. and Baltimore, D. (1986) Inducibility of κ immunoglobulin enhancer binding protein NF-κB by a posttranslational mechanism. *Cell* 47: 921-928

Serfling, E., Jasin, M. and Schaffner, W. (1985) Enhancers and eukaryotic gene transcription. *TIG* 1: 224-230

Sheen, F. and Levis, R. W. (1994) Transposition of the LINE-like retrotransposon TART to *Drosophila* chromosome termini. *Proc. Natl. Acad. Sci. USA* 91: 12510-12514

Shehee, W. R., Chao, S.-F., Loeb, D. D., Comer, M. B., Hutchison, C. A., III and Edgell, M. H. (1987) Determination of a functional ancestral sequence and definition of the 5' end of A-type mouse L1 elements. *J. Mol. Biol.* 196: 757-767

Siebel, C. W., Admon, A. and Rio, D. C. (1995) Soma-specific expression and cloning of PSI,a negative regulator of P element pre-mRNA splicing. *Gen. Dev.* 9: 269-283

Siebel, C. W. and Rio, D. C. (1990) Regulated splicing of the *Drosophila* P transposable element third intron in vitro: somatic repression. *Science* 248: 1200-1208

Simonelig, M., Bazin, A., Pélisson, A. and Bucheton, A. (1987) Transposable and non-transposable elements homologous to the *I* factor involved in I-R hybrid dysgenesis in *Drosophila melanogaster* coexist in various Drosophila species. *Proc. Natl. Acad. Sci. USA* 84: 1141-1145

Singer, M. F. (1982) SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. *Cell* 28: 433-434

Singer, M. F., Krek, V., McMillan, J. P., Swergold, G. D. and Thayer, R. E. (1993) *LINE-1*: a human transposable element. *Gene* 135: 183-188

Skowronski, J., Fanning, T. G. and Singer, M. F. (1988) Unit-length Line-1 transcripts in human teratocarcinoma cells. *Mol. Cell. Biol.* 8: 1385-1397

Skowronski, J. and Singer, M. F. (1985) Expression of a cytoplasmic LINE-1 transcript is regulated in a human teratocarcinoma cell line. *Proc. Natl. Acad. Sci. USA* 82: 6050-6054

Smale, S. T. and Baltimore, D. (1989) The initiator as a transcription control element. *Cell* 57: 103-113

Smale, S. T., Schmidt, M. C., Berk, A. J. and Baltimore, D. (1990) Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirements for mammalian transcription factor IID. *Proc. Natl. Acad. Sci. USA* 87: 4509-4513

Spradling, A. C. and Rubin, G. M. (1982) Transposition of cloned *P* elements into *Drosophila* germline chromosomes. *Science* 218: 341-347

Streck, R. D., MacGaffey, J. E. and Beckendorf, S. K. (1986) The structure of *hobo* transposable elements and their insertion. *EMBO J.* 5: 3615-3623

Swergold, G. D. (1990) Identification, characterization, and cell soecificity of a human LINE-1 promoter. *Mol. Cell. Biol.* 10: 6718-6729

Temin, H. M. and Mizutani, S. (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* 226: 1211-1213

Thayer, R. E., Singer, M. F. and Fanning, T. G. (1993) Undermethylation of specific *LINE-1* sequences in human cells producing a *LINE-1*-encoded protein. *Gene* 133: 273-277

Thummel, C. S., Boulet, A. M. and Lipshitz, H. D. (1988) Vectors for *Drosophila* P-element-mediated transformation and tissue culture transfection. *Gene* 74: 445-456

Treisman, R. and Maniatis, T. (1985) Simian virus 40 enhancer increases number of RNA polymerase II molecules on linked DNA. *Nature (London)* 315: 72-75

Udomkit, A., Forbes, S., Dalgleish, G. and Finnegan, D. J. (1995) BS a novel LINE-like element in *Drosophila melanogaster*. *Nucleic Acids Res.* 23: 1354-1358

Usheva, A. and Shenk, T. (1994) TATA-binding protein-independentinitiation: YY1, TFIIb And RNA polymerase II direct basal transcription on supercoiled template DNA. *Cell* 76: 1115-1121

Valgeirsdottir, K., Traverse, K. L. and Pardue, M. L. (1990), Het DNA: a family of mosaic repeated sequences specific for heterochromatin in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 87: 7998-8002

VanDyke, M. W., Roeder, R. G. and Sawadogo, M. (1988) Physical analysis of transcription preinitiation complex assembly on class II promoters. *Science* 241: 1335-1338

Vaury, C., Abad, P., Péisson, A., Lenoir, A. and Bucheton, A. (1990) Molecular characteristics of the heterochromatic *I* elements from a reactive strain of *Drosophila melanogaster*. *J. Mol. Evol.* 31: 424-431

Vaury, C., Bucheton, A. and Pélisson, A. (1989) The β heterochromatic sequences flanking the *I* elements are themselves defective transposable elements. *Chromosoma* 98: 215-224

Vaury, C., Chaboissier, M. C., Drake, M. E., Lajoinie, O., Dastugue, B. and Pélisson, A. (1994) The Doc transposable element in *Drosophila melanogaster* and *Drosophila simulans*: genomic distribution and transcription. *Genetica* 93: 117-124

Verma, I. M. (1975) Studies on reverse transcription of RNA tumor viruses. I. Localization of thermolabile DNA polymerase and RNase H activities on one polypepetide. *J. virol.* 15: 121-126

Voliva, C. F., Martin, S. L., Hutchison, C. A., III and Edgell, M. H. (1984) The dispersal process associated with the L1 family of interspersed repetitive sequences. *J. Mol. Biol.* 178: 795-813

Weiner, A. M., Deininger, P. L. and Efstratiadis, A. (1986) Non-viral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Ann. Rev. Biochem.* 55: 631-661

Weis, L. and Reinberg, D. (1992) Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. *FASEB* 6: 3300-3309

Witney, F. R. and Furano, A. V. (1984) Highly repeated DNA families in the rat. *J. Biol. Chem.* 259: 10481-10492

Woods-Samuels, P., Wong, C., Mathias, S. L., Scott, A. F., Kazazian, J., H. H. and Antonarakis, S. E. (1989) Characterisation of a nondeleterious L1-insertion in an intron of the human factor-VIII gene and further evidence of open reading frames in functional L1-elements. *Genomics* 4: 290-296

Xiong, Y., Burke, W. D., Jakubczak, J. L. and Eickbush, T. H. (1988) Ribosomal DNA insertion elements R1Bm and R2Bm can transpose in a sequence specific manner to locations outside the 28S genes. *Nucleic Acids Res.* 16: 10561-10573

Xiong, Y. and Eickbush, T. H. (1988a) Functional expression of a sequence-specific endonuclease encoded by the retrotransposon R2Bm. *Cell* 55: 235-246

Xiong, Y. and Eickbush, T. H. (1988b) The site specific ribosomal DNA insertion element R1Bm belongs to a class of non-long terminal repeat retrotransposons. *Mol. Cell. Biol.* 8: 114-123

Yannopolous, G., Stamatis, N., Monastirioti, M., Hatzopoulos, P. and Louis, C. (1987) *hobo* is responsible for induction of hybrid dysgenesis by strains of *Drosophila melanogaster* besring the male recombination factor 23.5 MRF. *Cell* 49: 487-495

Yannopoulos, G., Stamatis, N. and Eeken, J. C. J. (1986) Differences in the cytotype and hybrid dysgenesis inducing abilities of different P strains of *Drosophila melanogaster*. *Experimentia* 42: 1283-1285

Zenke, M., Grundstrom, T., Matthes, h., Wintzerith, M., Schaltz, C., Wildeman, A. and Chambon, P. (1986) Multiple sequence motifs are involved in SV40 enhancer function. *EMBO J.* 5: 387-397

Zenzie-Gregory, B., O'Shea-Greenfield, A. and Smale, S. T. (1992) Similar mechanisms for transcription initiation mediated through a TATA box or an initiator element. *J. Biol. Chem.* 267: 2823-2830

Zinn, K. and Maniatis, T. (1986) Detection of factors that interact with human beta-interferon regulatory region *in vivo* by DNase I footprinting. *Cell* 45: 611-618

APPENDIX A

Statistical analysis of enzyme assays

The CAT assays for all of the transformed lines, except for those that were transformed with p186Δ4.2 W8 (lines 227 and 228) and line 230 transformed with p1-Δ186W8, were done on two separate extracts and two assays were performed on each extract. There are at least three sources of variation that could account for the results observed in the assays. These are the variation of assays, the variation of extracts and the variation of lines in which the same transgene was inserted at different positions on the chromosomes.

In order to test the variation of assays, *t* -test was applied to the data. The mean values and standard errors of the two assays on each extracts were calculated and the P values were used to confirm the significance of the data. Table A.1 shows the *t*-values obtained from separate assays on the same extract from each transformed line. Reference to table of *t*-distribution (table A.5) shows that to achieve significance at the 5 % level (P=0.05) with 1 degree of freedom requires a *t*-value of 12.71 or more. Most of the *t*-values calculated for two assays on the same extract given in table A.1 are less than 12.71 indicating that the values from two assays on the same extract are not significantly different except for those that give the *t*-value of more than 12.71.

Table A.1 Calculated *t*-values of separate assays on the same extract.

CONSTRUCT	LINE	t values				
		MALE	FEMALE	OVARY	CARCASS	O/C
p186Δ4.2W8	227	7.792	16.18	18.61	3.646	21.11
	228	8.500	13.27	8.233	4.489	8.267

CONSTRUCT	LINE	t values				
		MALE	FEMALE	OVARY	CARCASS	O/C
p1-Δ186W8	230	11.54	11.93	12.60	26.71	22.33
	231(E1)	4.097	3.148	3.706	4.455	24.71
	(E2)	2.864	5.069	6.238	4.379	5.285
	232(E1)	53.03	10.21	36.40	39.09	104.0
	(E2)	18.33	6.333	20.00	5.429	7.364
	233(E1)	31.00	19.67	11.00	6.000	3.965
	(E2)	4.067	10.91	55.29	25.372	14.60
	234(E1)	33.00	8.091	5.125	12.01	7.000
	(E2)	12.58	7.350	8.835	21.79	2.080
p138W8	283(E1)	9.493	4.846	820.2	16.52	12.00
	(E2)	16.79	2.598	27.32	94.43	20.33
	284(E1)	147.8	2.464	7.2064	5.913	29.01
	(E2)	8.219	42.74	157.6	1.503	3.000
	285(E1)	2.240	2.595	2.910	30.83	3.242
	(E2)	2.539	7.543	8.644	3.991	7.286
	286(E1)	15.23	9.869	6.989	3.020	5.333
	(E2)	10.56	8.922	10.84	8.255	33.02
	287(E1)	1.699	12.13	272.1	7.370	6.867
	(E2)	3.405	3.791	11.55	2.163	2.532
	288(E1)	3.021	3.236	3.077	6.368	5.571
	(E2)	14.26	4.628	1.143	1.152	80.61

CONSTRUCT	LINE	t values				
		MALE	FEMALE	OVARY	CARCASS	O/C
p186 Δ101-157W8	289(E1)	3.946	3.412	12.85	2.450	1.829
	(E2)	2.670	3.662	9.149	3.821	2.789
	290(E1)	3.109	4.571	1.692	8.231	1.966
	(E2)	23.53	2.645	2.515	3.169	9.667
	291(E1)	9.143	7.263	4.168	2.883	8.636
	(E2)	3.585	3.255	2.167	12.44	2.647
	292(E1)	4.951	2.659	2.869	3.165	11.00
	(E2)	2.165	2.426	2.651	2.273	17.01

Variation of extracts was determined by comparing assays on two separate extracts (two assays on each extract). Independent *t*-test was performed at 2 degrees of freedom and the *t*-values comparing assays on two extracts of lines transformed with p1-Δ186W8, p138W8 and p186Δ101-157W8 are shown in table A.2 respectively. At P=0.05 significance level, the *t*-values need to exceed 4.303 to say that the assays on two separate extracts are significantly different. Only a few cases of the values shown in table A.2 are more than 4.303 therefore almost all of the assays done on separate extracts gave consistent results.

Table A.2 Calculated *t*-values of the assays on two separate extracts (2 assays on each extract) of transformed lines

CONSTRUCT	LINE	t values				
		MALE	FEMALE	OVARY	CARCASS	O/C
p1-Δ186W8	231	1.022	0.4432	0.0190	0.4469	0.3362
	232	0.6325	3.296	0.1943	1.243	2.091
	233	0.0661	0.6860	0.9570	1.437	1.011
	234	6.187	0.0614	1.118	1.590	1.534
p138W8	283	0.2249	1.265	8.606	7.130	10.26
	284	2.443	0.536	2.506	0.6142	2.495
	285	0.3299	3.777	0.6975	1.574	1.660
	286	1.204	0.0796	0.2617	0.1661	0.1649
	287	0.2256	0.1894	0.3556	0.1279	0.3243
	288	0.6600	1.213	0.6831	0.3206	0.0712
p186 Δ101-157W8	289	0.0386	0.8293	1.361	0.2595	0.4868
	290	0.7590	0.0190	0.0552	0.1455	0.0338
	291	1.131	0.2959	0.5225	1.424	2.470
	292	0.3595	0.2010	0.0912	0.8170	4.243

Variation on different lines transformed with the same construct could be due to position effect. Since the transgene is inserted at different sites on the chromosomes, the neighbour chromosomal DNA sequences could affect the expression of the transgene at different level. This type of variation was determined by looking at how much the value of individual sample of each line disperse from the mean value. The result of this analysis (not shown here) shows that the values obtained from assays on different lines are significantly different. To reduce this effect, more transform lines should be obtained and used in the assays. Despite the variation of the CAT values in

each tissue, the O/C ratios are consistent among the lines. As discussed earlier in chapter 6, there is one line of the transformants carrying p138W8 transgene (line 287) and one line of those carrying p186Δ101-157W8 transgene (line 290) that give exceedingly high O/C ratios as compare to the ratios of the other lines carrying the same transgene. The *t*-values of O/C ratios of lines transformed with p138W8 and p186Δ101-157W8 are shown in table A.3. With 23 degrees of freedom line 287 is the only line that gives the *t*-value of O/C ratio higher than 2.069 that is the *t*-value at P=0.05 level of significance, in other word this line gives the value of $P \leq 0.05$. Similarly, line 290 is the only line that gives higher *t*-values than the value of 2.131 at p=0.05 with 15 degrees of freedom. Therefore these two line could be considered as the lines that express different O/C ratios from the other.

Table A.3 Calculated *t*-values of the O/C ratio of different lines

CONSTRUCT	LINE	O/C
p138W8	283	2.011
	284	1.500
	285	1.689
	286	2.020
	287	10.01
	288	2.052
p186 Δ101-157W8	289	1.759
	290	6.014
	291	1.600
	292	2.050

β-galactosidase assays were done on a single extract of each sample. The variation of assays and the variation of lines should therefore account for the results observed. With 1 degree of freedom the two assays on each individual extract show no significant difference as they give t-values of less than 12.71 that correspond to P=0.05 significance level (not shown here). The variation of lines was determined by comparing the activity between two or three lines that carry the same transgene. Table A.4 shows the *t*-values calculated for transformants carrying each transgene. With 3 degrees of freedom the assays without heat-shock between those two lines transformed with pI821 and between those transformed with pI822 do not show any significant difference (with the *t*-values of 2.6167 and 0.3226 respectively). The lines transformed with pI824 give the *t*-values of 4.9806 which exceed the t-values at P=0.05 (3.182) and therefore could be said that there is significant variation between lines. Lines transformed with pI824 also show significant different with 5 degrees of freedom (with the *t*-value of 4.041 comparing to 2.571 at P=0.05). The assays on heat-shocked flies also give similar level of significance.

Table A.4 Calculated t-values of different lines carrying the same transgene

CONSTRUCT	t- value	
	non heat-shock	heat-shock
pI821	2.617	2.827
pI822	0.323	2.133
pI823	4.041	3.072
pI824	4.981	14.856

Table A.5 *t*-distribution

Degrees of freedom	Value of P					
	0.10	0.05	0.02	0.01	0.002	0.001
1	6.314	12.71	31.82	63.66	318.3	636.6
2	2.920	4.303	6.965	9.925	22.33	31.60
3	2.353	3.182	4.541	5.841	10.21	12.92
4	2.132	2.776	3.747	4.604	7.173	8.610
5	2.015	2.571	3.365	4.032	5.893	6.869
6	1.943	2.447	3.143	3.707	5.208	5.959
7	1.895	2.365	2.998	3.499	4.785	5.408
8	1.860	2.306	2.896	3.355	4.501	5.041
9	1.833	2.262	2.821	3.250	4.297	4.781
10	1.812	2.228	2.764	3.169	4.144	4.587
11	1.796	2.201	2.718	3.106	4.025	4.437
12	1.782	2.179	2.681	3.055	3.930	4.318
13	1.771	2.160	2.650	3.012	3.852	4.221
14	1.761	2.145	2.624	2.977	3.787	4.140
15	1.753	2.131	2.602	2.947	3.733	4.073
16	1.746	2.120	2.583	2.921	3.686	4.015
17	1.740	2.110	2.567	2.898	3.646	3.965
18	1.734	2.101	2.552	2.878	3.610	3.922
19	1.729	2.093	2.539	2.861	3.579	3.883
20	1.725	2.086	2.528	2.845	3.552	3.850
21	1.721	2.080	2.518	2.831	3.527	3.819
22	1.717	2.074	2.508	2.819	3.505	3.792
23	1.714	2.069	2.500	2.807	3.485	3.767
24	1.711	2.064	2.492	2.797	3.467	3.745
25	1.708	2.060	2.485	2.787	3.450	3.725
26	1.706	2.056	2.479	2.779	3.435	3.707
27	1.703	2.052	2.473	2.771	3.421	3.690
28	1.701	2.048	2.467	2.763	3.408	3.674
29	1.699	2.045	2.462	2.756	3.396	3.659
30	1.697	2.042	2.457	2.750	3.385	3.646

The table gives the percentage points most frequently required for significance tests and confidence limits based on 'Student's' *t*-distribution. Thus, the probability of observing a value of t, with 10 degrees of freedom, greater in *absolute value* than 3.169 (i.e. < -3.169 or $> +3.169$) is exactly 0.01 or 1 per cent.

Taken from: **Statistical methods in biology** (third edition) 1995 by Norman T. J. Bailey, University Press, Cambridge.

APPENDIX B

The work on sequence analysis of a novel LINE-like element called BS has been published in:

Udomkit, A., Forbes, S., Dalgleish, G. and Finnegan, D. J. (1995) BS a novel LINE-like element in Drosophila melanogaster. *Nucleic Acids Res.* 23: 1354-1358

1354–1358 Nucleic Acids Research, 1995, Vol. 23, No. 8

BS a novel LINE-like element in *Drosophila melanogaster*

A. Udomkit, S. Forbes, G. Dalgleish and D. J. Finnegan*

Institute of Cell and Molecular Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JR, UK

Received December 30, 1994; Revised and Accepted March 7, 1995 EMBL accession no. X77571

ABSTRACT

Transposable elements with long terminal inverted repeats are rare and only one family of elements of this sort has been identified in the genome of *Drosophila melanogaster*. An insertion associated with the Hs^{BS} mutation of the *achaete–scute* complex has been reported to be a second element of this type. We have determined the complete sequence of this insertion and have shown that it is in fact two copies of a new LINE-like transposable element, that we have called *BS*, inserted in opposite orientation 337 bp apart. Like other elements of this type, *BS* has two open reading frames that appear to encode a *gag*-like polypeptide and a reverse transcriptase. There are few complete *BS* elements in the five strains of *D.melanogaster* that we have tested and they appear to transpose infrequently. The events that may have lead to the double *BS* insertion are discussed in terms of the supposed mechanism of transposition of LINE-like elements.

INTRODUCTION

Transposable elements are an important component of eukaryotic genomes, making up 15% or more of the total DNA. They occur as families of dispersed repeat sequences scattered throughout the genome, the number of copies varying from less than ten to several hundred thousand depending on the element and species concerned. They can be classified according to their structure and presumed mechanism of transposition, and fall into two main classes: elements that transpose by reverse transcription of an RNA intermediate, Class I elements, and elements that transpose directly from DNA to DNA, Class II elements.

There are two types of Class I elements. Class I.1 elements resemble retroviruses in having long terminal direct repeats (LTRs) and open reading frames with the potential to code for polypeptides similar to viral *gag*, *pol* and, in some cases, *env* proteins. Class I.2 elements also have open reading frames with similarities to retroviral *gag* and *pol* genes, but have no terminal repeats and end with A-rich sequences at the 3′ end of their coding strands. Elements of this type are often referred to as LINE-like elements as the first examples to be detected were mammalian LINE or L1 elements (1).

The majority of Class II elements have short terminal inverted repeats, but some have been reported to have inverted repeats that are hundreds of base pairs long. Few elements with this structure have been identified, the best characterized being the FB elements of *Drosophila melanogaster* (2). These have inverted repeats 1–2 kb long that are themselves made up of short tandem repeats. Some FB elements contain nothing more than these tandem repeats, while others have a central region of unrelated and non-repeated sequence (3–5). In one case this central region appears to be due to insertion of a different Class II element (6), while the central region of another may code for a function that stimulates FB transposition (7,8).

Campuzano *et al.* (9) reported finding what appeared to be a member of a second family of elements with long inverted repeats that they named *BS* because it was found associated with the mutation Hw^{BS}. This is a derivative of Hw^{1}, a mutation of the *achaete–scute* complex (AS–C) that is associated with an insertion of a copy of the retrovirus-like element *gypsy* within the T5 AS-C transcription unit. Chromosomes carrying the Hw^{BS} mutation have a slightly reduced hairy wing phenotype and an additional 8 kb of DNA inserted within the Hw^{1} *gypsy* element. This extra DNA contains inverted repeats of 2.5 kb as judged by restriction site mapping and hybridization data. The 8 kb insertion does not contain *FB* sequences and was named *BS*.

In order to compare the structure of this putative new inverted repeat element with that of *FB* elements, we have determined the nucleotide sequence of the whole of the *BS* insertion. The results indicate that the Hw^{BS} mutation is not associated with insertion of a single inverted repeat element, but has two related sequences inserted 337 bp apart and in opposite orientation within the *gypsy* element of Hw^{1}. These are members of a new family of LINE-like elements as they show no strong sequence similarity to any of the five families of such elements, *Doc*, *F*, *G*, *I* and *jockey*, that have been identified in the genome of *D.melanogaster* (10,11).

MATERIALS AND METHODS

DNA sequencing

The sequence of the Hw^{BS} insertion was determined by subcloning the *Pst*I fragments BS0.6, BS816, BS817, BS818 and BS819, and the *Pst*I–*Bam*HI fragment BS2.0 (Fig. 1A) of the plasmid p14RRHw^{BS} (9) in the vector pUBS (12). Double-stranded templates were prepared using QIAGEN™ columns and these were sequenced by dideoxynucleotide chain termination (13). The *Hin*dIII fragments BS1.1, BS5.8 and BS7.6 were used to orient and link the sequences from the *Pst*I fragments. The entire

* To whom correspondence should be addressed

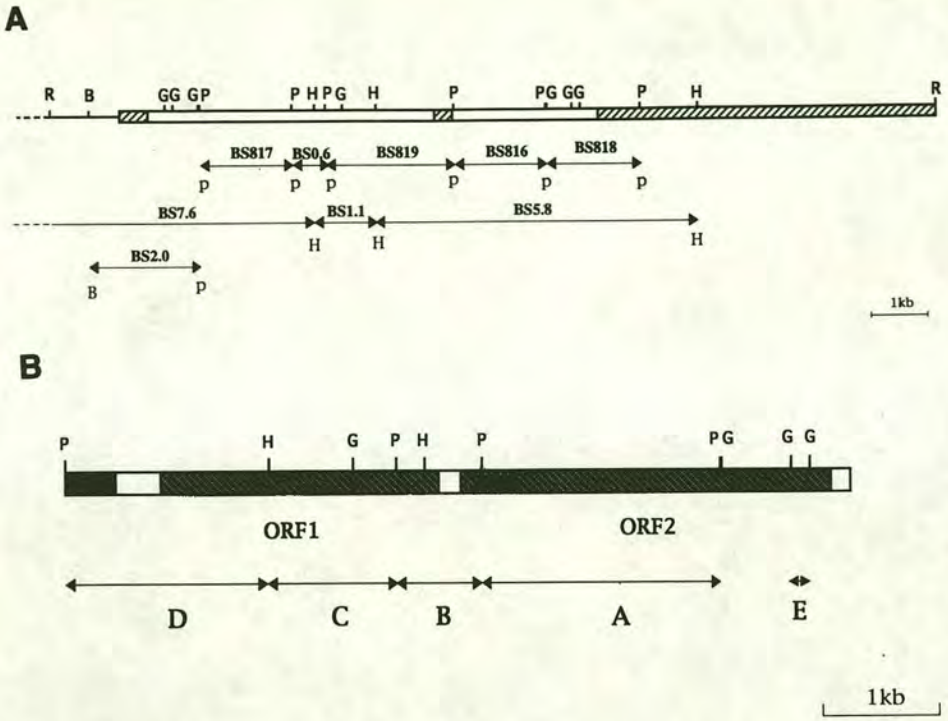


Figure 1. (**A**) Restriction map of an *Eco*RI fragment from the region of the AS-C region containing the Hw^{BS} insertions. The open boxes represent *BS* sequences. *BS*1 is on the left and *BS*2 on the right. The hatched boxes represent *gypsy* sequences, the thin lines represents AS-C sequences and the dashed line represents vector sequences. The lines underneath the map indicated fragments that were subcloned for DNA sequencing. (**B**) Restriction map of the *BS*1 insertion shown in the 5′→3′ orientation. The two open reading frames are indicated by hatched boxes while non-coding regions are indicated by open boxes. The filled box represents *gypsy* DNA. The lines underneath the map indicate restriction fragments used as probes for the Southern hybridizations shown in Figure 4. Restriction sites are indicated as follows: B, *Bam*HI; G, *Bgl*I; H. *Hin*dIII; P, *Pst*I; R, *Eco*RI.

sequence has been determined on both strands of the element. Sequence data were assembled using programs written by Staden (14) and were analysed using the suite of programs from the Wisconsin Genetics Computer Group (15). The program PILEUP is from Version 7.0 of the GCG package (Genetics Computer Group, Inc.).

Drosophila DNA preparation and Southern blotting

Genomic DNA was prepared from strains of *Drosophila* by homogenizing 10 flies in homogenization buffer (8 M urea, 0.35 M NaCl, 10 mM Tris–HCl pH 8.3, 10 mM EDTA). Homogenates were extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1) and then with chloroform:isoamyl alcohol (24:1). Nucleic acids were ethanol precipitated and dissolved in 50 ml of 20 mg/ml RNase in 10 mM Tris–HCl pH 8.3, 1 mM EDTA.

Five micrograms of DNA from each strain were digested with the appropriate restriction enzymes and electrophoresed on 0.7% agarose gels together with amounts of control plasmids equivalent to two, five, eight and 10 copies per genome. The DNA was then transferred to GeneScreen Plus™ nylon membrane (Dupont) by capillary action in 1.5 M NaCl, 0.15 M Na citrate. Hybridization was carried out in 0.5 M Na phosphate pH 7.2, 7% sodium dodecyl sulphate and 1 mM EDTA. After hybridization filters were washed in 40 mM Na phosphate, 1% sodium dodecyl sulphate, 1 mM EDTA. Fragments used as probe were purified by agarose gel electrophoresis and labelled with [^{32}P]adCTP by random priming (16). The amount of hybridization in each track was quantified with the aid of a PhosphorImager (Molecular Dynamics).

RESULTS

We have determined the sequence of the *BS* insertion starting from a 14.8 kb *Eco*RI fragment of Hw^{BS} genomic DNA kindly supplied by Dr S. Campuzano. This contains the entire *gypsy* and *BS* insertions associated with the Hw^{1} mutation (Fig. 1A). The sequence of the *BS* insertion is diagrammed in Figure 1B and is in the EMBL data base as accession X77571. It is not an insertion of a single inverted repeat element, as had been thought, but is made up of two copies of an element lying close together and in opposite orientation within *gypsy*. We have called the left-hand element *BS*1 and the right-hand element *BS*2.

The *BS*1 and *BS*2 insertions are 5126 and 2571 bp long, respectively, and each is flanked by a duplication of a 12 bp sequence present only once in the corresponding part of the *gypsy* sequence. These sequences are different for *BS*1 and *BS*2, and were presumably generated when these elements inserted at these sites. These insertions are separated by 337 bp of *gypsy* DNA. Geyer *et al.* (17) have determined what they believed to be the ends of the whole *BS* insertion when analysing insertions that modify the phenotype of mutations due to *gypsy* insertions. The sequences that they reported were those of the ends of *BS*2.

The structure of the *BS*1 element indicates that it is a LINE-like (Class I.2) transposable element. It has no terminal repeats, has an A-rich sequence at the 3′ end of the coding strand and has two long open reading frames (Fig. 1B). The first has three copies of

A

```
BS          CRRCQEYGHTAKYC
jockey      CQRCQIFGHSKNYC
Doc         CTNCQEYGHTKAYC
F           CTNCQEYGHTRGYC
G           CFRCQGFGHTQRYC
I           CKKCLRFGHPTPIC
R1Dm        CHRCVGFDHKVSEC
            *  *    *    *

BS          CARCGEN-HQTMQ--C
jockey      CGKCSGP-HMTGFALC
Doc         CVVCSEP-HTTAN--C
F           CVVCGDL-HDSKQ--C
G           CKVCGGL-HDSRA--C
I           CINCSETKHTNDGEKC
R1Dm        CRQCGQQ-HTAAK--C
            *  *    *    *

BS          CYHCSEN-HTASFKGC
jockey      CINCGGD-HVSTDKSC
Doc         CSNCGEK-HTANYRGC
F           CNNCGGN-HTANYRGC
G           CLHCQAD-HPASFKGC
I           CLNCRNNPELDHQHSP
R1Dm        CRNCRHRGQPSGHYML
            *  *    *    *
```

B

```
BS          YRPISLLSSLSKLWER
jockey      YRPTSLLPSLGKIMER
Doc         YRPISLLTCLSKLFEK
F           YRPISLLSCISKLFEK
G           YRPISLLAILSKILER
I           YRPISLNCCIAKILDK
R1Dm        YRPISLNCCIAKILDK
R2Dm        FRPISVPSVLVRQLNA
            *******  * *  **

BS          LATYADDIALLYS
jockey      IATYADDTAVLTK
Doc         TSTFADDTAILSR
F           VSTFADDTAILSR
G           VATYADDTAFLAS
I           FNAYADDFFLIIN
R1Dm        LSAYADDLLLLVE
R2Dm        AAAFADDLVLFAE
              *****  * *
```

Figure 2. (**A**) Alignment of the sequences of the first open reading frames of seven LINE-like elements from *D.melanogaster* showing the conserved CCHC motifs. Amino acid residues are represented by the standard one-letter code. Positions at which the same residue occurs in at least four of the sequences are marked with asterisks. (**B**) Alignments of amino acid sequences from parts of the second open reading frame of eight LINE-like elements of *D.melanogaster*. The first alignment shows the conserved motif YRPISLL that is characteristic of LINE-like elements (35,36) and YXDD that is found in all reverse transcriptases (37). Sequences corresponding to the other conserved motifs of reverse transcriptases are also present in ORF2 of *BS*. Positions at which the same residue is found in at least five of the sequences are marked with an asterisk. The sequences were obtained from the following *jockey* (38), *Doc* (39), *F* (40), *G* (41), *I* (42), *R1Dm* and *R2Dm* (20).

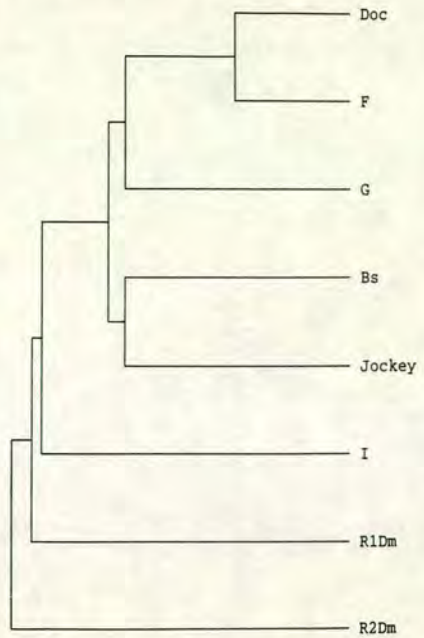


Figure 3. Dendrogram showing the relationships between the amino acid sequences of the putative reverse transcriptases encoded by the second open reading frames of the LINE-like transposable elements of *D.melanogaster* shown in Figure 2B. Distances along the horizontal axis are proportional to the differences between sequences.

a CCHC motif like those found in the *gag* genes of retroviruses and in the first open reading frames of LINE-like elements (Fig. 2A). The second has motifs characteristic of a reverse transcriptase (Fig. 2B). Five other transposable elements of this type, *Doc*, *F*, *G*, *I* and *jockey*, have been found in the genome of *D.melanogaster* (10,11), but the sequence of *BS*1 has not been reported previously. Elements of this type are frequently truncated by varying amounts at their 5′ ends (1,18). This is thought to happen at some point during transposition (19). The sequence of *BS*2 is identical to *BS*1, but is missing the first 2555 bp, suggesting that it has suffered such a deletion.

We have determined the relationships between the amino acid sequences of the putative reverse transcriptases of *BS*1 and the five *Drosophila* LINE-like elements mentioned above as well as those of two other potential transposable elements of this type, R1Dm and R2Dm (20), that are found inserted at specific sites within a proportion of the 28S rRNA coding sequences of *D.melanogaster*. These are believed to differ from the other LINE-like elements in being able to insert at these sites in a sequence-specific manner. This was done by comparing the complete amino acid sequence encoded by the second open reading frame of each element using the GCG program PILEUP. The result (Fig. 3) suggests that *BS* is most closely related to *jockey*.

We have used Southern transfer experiments to measure the number of *BS* elements in the genomes of five different stocks of *D.melanogaster* and to estimate their degree of conservation (Fig. 4). Three restriction fragments of *BS*1 were used to probe digests of genomic DNA using enzymes that should excise the corresponding fragments from complete genomic copies. Digests of a plasmid containing the fragment used as probe were used as size markers and to give an indication of the number of copies per genome. The intensity of hybridization of fragments corresponding in size to the probe decreases from the 3′ to 5′ end of the element (Fig. 4A–C) confirming that *BS* elements are a family of sequences containing variable 5′ deletions.

There are about five copies of the 3′ most fragment tested (fragment A in Fig. 1B) in the strains tested. The higher molecular weight bands presumably correspond to *BS* elements truncated somewhere within fragment A. This gives an upper limit of about 10 to the number of *BS* elements in the genome. The results from using fragment C as probe suggest that the strains tested contain few if any complete BS elements.

Campuzano *et al.* (9) estimated the copy number of *BS* elements to be 15 using the 1.7 kb *Pst*I fragment (fragment D in Fig. 1B) as probe. This fragment includes both internal *gypsy* sequences and sequences from the 5′ end of *BS* so this estimate reflects the number of copies of both elements. We were unable to detect fragments corresponding to this 1.7 kb *Hin*dIII–*Pst*I fragment in DNA of any of the five strains tested (Fig. 4D), indicating that the double insertion of *BS* elements within *gypsy* does not occur in any of these strains.

We have compared the distribution of *BS* elements in the genomes of five strains of *D.melanogaster* by probing *Pst*I-digested DNA with a *Bgl*II fragment from the 3′ end of *BS*1. Each *BS* element should give a band of hybridization that depends on its position in the genome. The patterns of hybridization are similar from strain to strain (Fig. 5) suggesting that *BS* elements transpose infrequently in them. Unfortunately we have not been able to obtain the Hw^{BS} strain to check that it does not have an unusually high frequency of *BS* transposition.

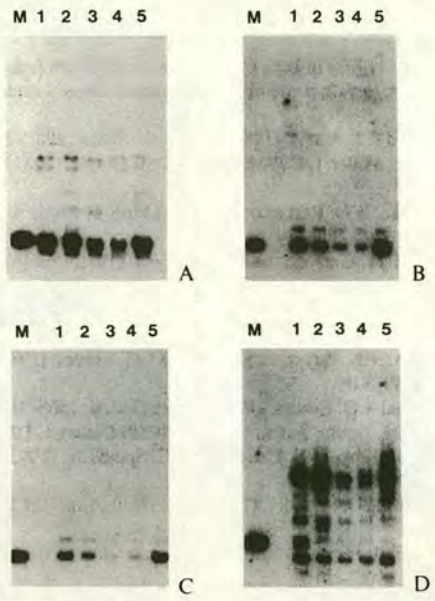


Figure 4. Southern transfer experiment showing the conservation of *BS* elements within the genomes of five strains of *D.melanogaster* isolated from the wild. Genomic DNAs were digested with *Pst*I in (**A**) and (**B**) and with *Pst*I and *Hin*dIII in (**C**) and (**D**). The fragments used to probe these DNAs are shown in Figure 1B. In each case a digest of a plasmid including the fragment used as probe was loaded on the gel in varying amounts to serve as a size marker and to allow the amount of hybridization of the genomic DNAs to be quantified. Track M shows the plasmid DNA loaded in an amount equivalent to 15 copies per haploid genome. The strains of *D.melanogaster* are as follows; 1, w^k; 2, Samarkand; 3, Charolles; 4, Oregon R; 5, Canton S.

Figure 5. Southern transfer experiment showing the degree of heterogeneity in the distribution of *BS* elements in the genomes of the strains of *D.melanogaster* analysed in Figure 4. Genomic DNAs were digested with *Pst*I and probed with the *Bgl*II fragment of *BS*, Fragment E in Figure 1B.

DISCUSSION

The insertion associated with the Hw^{BS} mutation is not an inverted repeat element like *FB*, as had been supposed previously, but is formed by two copies of a new element, *BS*, inserted in opposite orientation. These elements have all the characteristics of LINE-like elements and *BS*1 may be full-length.

Although the proportion of the *D.melanogaster* genome that comprises transposable elements is not unusually large, the number of different transposable elements of all kinds that can be found in this species is greater than that known for any other eukaryote (21). Eight LINE-like elements have now been identified in *D.melanogaster* as well as a short sequence that may represent a ninth family (22). This is the greatest number of elements of this type to have been found in any species. Five of these, *Doc F, I, jockey* and, presumably, *BS*, can insert at many sites in the genome while two of them, *R1Dm* and *R2Dm*, are site-specific elements (23,24). *G* elements are concentrated in the non-transcribed spacer DNA of ribosomal gene clusters and may not be active in most strains (25).

The *BS*2 element has inserted within one copy of a tandemly repeated 12 bp sequence that is the binding site for the product of the *suppresser of Hairy wing* gene (26). Binding of su(Hw) protein to these repeats in a *gypsy* element inserted upstream of the *yellow* gene alters *yellow* expression by interfering with the interaction of the *yellow* promoter and some of its upstream regulatory sequences (27,28). Insertion of the transposable elements *hobo* or *jockey* within these repeats suppresses the mutant phenotype of y^2 presumably because this reduces the binding of su(Hw) protein (17).

The *gypsy* element associated with Hw^1 is inserted within the transcribed region of the T5 (achaete) gene of the AS–C and the mutant allele is transcribed to give a hybrid T5–*gypsy* transcript that is about 10-fold more abundant than T5 transcripts from the wild-type allele (9). The level of this hybrid transcript is reduced in flies carrying the *su(Hw)* mutation suggesting that in this case binding of su(Hw) protein may prevent interaction of the T5 promoter and a *cis*-acting negative regulator. The binding of su(Hw) protein to its target may be reduced somewhat by insertion of the *BS*2 element within one of the 12 bp repeats as the Hw^1 phenotype is slightly attenuated in Hw^{BS} flies, although no effect on the level of the T5-*gypsy* transcript has been detected (9).

The Hw^{BS} insertion must have been produced by either two independent transposition events or a single transposition event that generated two copies of *BS*. Insertions of *P* elements have been found that are close together and in opposite orientation (29). These are probably the results of independent insertions as *P* elements transpose at high frequency and preferentially to sites adjacent to a donor element (30). Elements that transpose by reverse transcription of an RNA intermediate, as is the case for LINE-like elements (19,31,32), are unlikely to transpose preferentially to adjacent sites and the frequency of *BS* transposition appears to be low. This suggests that the Hw^{BS} double insertion is the result of a single transposition event.

LINE-like elements are thought to integrate by a mechanism in which the 3′-OH at a nick in a strand of chromosomal DNA is used to prime DNA synthesis by an element encoded reverse transcriptase that uses the RNA transposition intermediate as template (33,34). The DNA synthesized in this way is in turn used as the template for synthesis of the second strand to complete integration. If the RNA intermediate were to be released after synthesis of the first DNA strand it might to serve as primer for a second integration event close by resulting in a double insertion.

ACKNOWLEDGEMENTS

This work was supported by research grants from the Medical Research Council and the Human Frontier Science Program and a studentship to AU from the Government of Thailand. We are grateful to S. Campuzano for supplying the original *BS* clone.

REFERENCES

1 Hutchison III, C.A., S.C. Hardies, D.D. Loeb, W.R. Sehee, and M.H. Edgell. (1989) In D.E. Berg and M.M. Howe (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 593–617.
2 Bingham, P.M. and Z. Zachar. (1989) In D.E. Berg and M.M. Howe (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 485–502.
3 Truett, M.A., R.S. Jones, and S.S. Potter. (1981) *Cell*, **24**, 753–763.
4 Potter, S.S. (1982) *Mol. Gen. Genet.*, **188**, 107–110.
5 Potter, S.S. (1982) *Nature*, **297**, 201–204.
6 Brierley, H.L. and S.S. Potter. (1985) *Nucleic Acids Res.*, **13**, 485–500.
7 Templeton, N.S. and S.S. Potter. (1989) *EMBO J.*, **8**, 1887–1894.
8 Harden, N. and M. Ashburner. (1990) *Genetics*, **126**, 387–400.
9 Campuzano, S., L. Balcells, R. Villares, L. Carramolino, L. Garcoa-Alonso, and J. Modolell. (1986) *Cell*, **44**, 303–312.
10 Finnegan, D.J. (1989) In D.E. Berg and M.M. Howe (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 519–522.
11 Finnegan, D.J. (1989) In D.E. Berg and M.M. Howe (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 503–517.
12 Goman, M., B. Mons, and J. Scaife. (1991) *Mol. Biochem. Parasitol.*, **45**, 281–288.
13 Sanger, F., S. Nicklen, and A.R. Coulson. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
14 Staden, R. (1982) *Nucleic Acids Res.*, **10**, 4731–4751.
15 Devereux, J., P. Haeberli, and O. Smithies. (1984) *Nucleic Acids Res.*, **12**, 387–395.
16 Feinberg, A.P. and B. Vogelstein. (1983) *Anal. Biochem.*, **132**, 6–13.
17 Geyer, P.K., M.M. Green, and V.G. Corces. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8593–8597.
18 Bucheton, A. (1990) *Trends Genet.*, **6**, 16–21.
19 Jensen, S. and T. Heidmann. (1991) *EMBO J.*, **10**, 1927–1937.
20 Jakubczak, J.J., Y. Xiong, and T.H. Eickbush. (1990) *J. Mol. Biol.*, **212**, 37–52.
21 Finnegan, D.J. (1992) In D.L. Lindsley and G.G. Zimm (eds), *The Genome of Drosophila melanogaster*. Academic Press, London, pp. 1096–1107.
22 Pittler, S.J. and R.L. Davis. (1987) *Mol. Gen. Genet.*, **208**, 325–328.
23 Roiha, H., J.R. Miller, L.C. Woods, and D.M. Glover. (1981) *Nature*, **290**, 749–753.
24 Jakubczak, J.L., W.D. Burke, and T.H. Eickbush. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 3295–3299.
25 Di Nocera, P.P., F. Graziani, and G. Lavorgna. (1986) *Nucleic Acids Res.*, **14**, 675–691.
26 Spana, C., D.A. Harison, and V.G. Corces. (1988) *Genes Dev.*, **2**, 1414–1423.
27 Harrison, D.A., P.K. Geyer, C. Spana, and V.G. Corces. (1989) *Dev. Genet.*, **10**, 239–248.
28 Geyer, P.K. and V.G. Corces. (1992) *Genes Dev.*, **6**, 1865–1873.
29 Roiha, H., G.M. Rubin, and K. O'Hare. (1988) *Genetics*, **119**, 75–83.
30 Tower, J., G.H. Karpen, N. Craig, and A.C. Spradling. (1993) *Genetics*, **133**, 347–359.
31 Evans, J.P. and R.D. Palmiter. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 8792–8795.
32 Pélisson, A., D.J. Finnegan, and A. Bucheton. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 4907–4910.
33 Pritchard, M.A., J.-M. Dura, A. Pelisson, A. Bucheton, and D.J. Finnegan. (1988) *Mol. Gen. Genet.*, **214**, 533–540.
34 Schwartz-Sommer, Z., L. Leclercq, F. Gobel, and H. Saedler. (1987) *EMBO J.*, **6**, 3873–3880.
35 Xiong, Y. and T.E. Eickbush. (1990) *EMBO J.*, **9**, 3353–3362.
36 Finnegan, D.J. (1988) In *Transposition*. Cambridge University Press.
37 Toh, H., R. Kikuno, H. Hayashida, T. Miyata, W. Kugimiya, S. Inouye, S. Yuki, and K. Saigo. (1985) *EMBO J.*, **4**, 1267–1272.
38 Primagii, A.F., L.J. Mizrokhi, and Y.V. Ilyin. (1988) *Gene*, **70**, 253–262.
39 O'Hare, K., M.R.K. Alley, T.E. Cullingford, A. Driver, and M.J. Sanderson. (1991) *Mol. Gen. Genet.*, **225**, 17–24.
40 Di Nocera, P.P. and G. Casari. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5483–5487.
41 Di Nocera, P. (1988) *Nucleic Acids Res.*, **16**, 4041–4052.
42 Fawcett, D.H., C.K. Lister, E. Kellett, and D.J. Finnegan. (1986) *Cell*, **47**, 1007–1015.