Structure and Regulation of the *I* Factor Promoter of *Drosophila melanogaster*.

Carol McLean

A thesis presented for the degree of PhD

University of Edinburgh

1991

I dedicate this thesis to my family.

Declaration:

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

Carol McLean

December, 1991.

ACKNOWLEDGEMENTS:

There are many people to whom I am indebted for their help and encouragement during the three years of this thesis, and before. I would firstly like to thank my supervisor, David Finnegan, for adopting me as a secondhand PhD student, and for his advice, encouragement and enthusiasm throughout this project. I am also grateful to Mary Bownes for acting as my advisor.

I would especially like to thank the two people whose support and persuasion encouraged me to embark on this PhD. They are Xavi Estevill and Gordon Alan.

I would also extend my thanks to Alain Bucheton and the many people I had a pleasure to work with at the CNRS, Gif Sur Yvettes, including Maki, Neel, and Janet, for their advice, humour, and hospitality during the three months I spent there.

I am also indebted to many members of the ICMB, past and present for their friendship and support. These people are Nik, Alison, Trevor, Glen, Susan, Melanie, Vicky, Gerry, and Michael. I would particularly like to thank Nik whose advice, humour, and tolerance has been relentless. Thanks go also to Irina Arkhipova for permission to used some of her unpublished results. I would also like to thank Helen and Lorraine for keeping me supplied with clean scintillation vials (and other glassware), particularly Helen for her good humour and gossip. I would also like to thank Kerr Hatrick, with whom it was a pleasure to work during summer '91, and for his determined effort and success at making plasmid p186T.2 (pity about pCAT.2!).

I am also grateful to Graham Brown and Frank Johnston for an excellent photography service (even at short notice), and David Melton and Angela Pow for use of their mammalian tissue culture facilities and advice during these experiments. A second thanks goes to David for use of his laser printer and drawperfect package that generated many of the figures in this thesis, and to Nik for explaining how to use it.

I would also like to thank the people outside the lab - Christine, Raymond, Ruth and Douglas, for making life easier and who were often a welcome distraction.

Finally I would like to thank my parents, Jean and Jim, my sister, Elizabeth, brother John, and my gran, Lily, for their support and encouragement.

ABBREVIATIONS:

A	- adenosine
amp	- ampicillin
ATP	- adenosine 5'-triphosphate
βgal	- β-galactosidase
bp	- basepair
BSA	- bovine serum albumin
C	- cytidine
°C	- degrees celsius
CAT	- chloramphenicol acetyl transferase
cDNA	- complementary DNA
Ci	- curie
CMV	- cytomegalo virus
cpm	- counts per minute
d	- deoxy
dd	- dideoxy
dig	- digoxygenin
DMSO	- dimethyl sulphoxide
DNase	- deoxyribonuclease
dNTP	- deoxyribonucleoside triphosphate
DTT	- dithiothreitol
g	- gram
G	- guanosine
HEPES	- N-2-hydroxyethyl piperazine-N-2-ethane sulphonic acid
IAA	- isoamyl alcohol
IPTG	- isopropyl-β-D-thiogalactopyranoside
kb(s)	- kilobase(s)
M	- molar
mg	- miligram
min	- minute

ml	- millilitre
mol	- mole
MW	- molecular weight
N	- any nucleoside
ng	- nanogram
nm	- nanometre
OD	- optical density
ONPG	- O-nitrophenyl-β-D-galactopyranoside
p	- plasmid
p	- pico
PCR	- polymerase chain reaction
PEG	- polyethylene glycol
RNase	- ribonuclease
rpm	- revolutions per minute
SDS	- sodium dodecyl sulphate
SDM	- site directed mutagenesis
SV40	- simian virus 40
T	- thymidine
TEMED	- N,N,N'-tetramethylethylenediamine
TLC	- thin layer chromatography
Tris	- tris (hydroxymethyl) aminomethane
U	- uridine
u	- unit
μg	- microgram
μl	- microlitre
UV	- ultraviolet
v	- volt
v	- volume
v/v	- volume per volume
w/v	- weight per volume

X-gal - 5-bromo-4-chloro-3-indol-β-D-galactopyranoside

ABSTRACT:

The *I* factor of *Drosophila melanogaster* is the transposable element that controls I-R hybrid dysgenesis. I-R hybrid dysgenesis is seen when males of strains that have *I* factors (inducer strains) are crossed with females that lack them (reactive strains). Dysgenesis is manifested only in the female progeny of such a cross (SF females) as reduced hatchability of eggs. The females progeny of the reciprocal cross (RSF females) are fertile. SF sterility correlates with high rates of *I* factor transposition in the SF germline.

The *I* factor is a LINE-like transposable element, mobility of which is thought to occur by reverse transcription of a full length transcript. This mechanism predicts that the *I* factor has a promoter internal to the transcription unit. Occurrence of the putative RNA intermediate for *I* factor transposition correlates with transposition frequency, suggesting that *I* factor mobility is controlled at the level of transcription or RNA stability.

The experiments described in this thesis identify a promoter within the 5' UTR of the *I* factor that initiates transcription at position +1. Characterisation of this promoter fused to the CAT reporter gene in transient assays in *Drosophila* tissue culture cells has identified sequences that modulate expression positively and negatively.

Transformation of a reactive strain with analogous *I*-CAT expression vectors demonstrated that sequences within the first 186 nucleotides of *I* contribute to elevated levels of expression in ovaries, suppression of promoter activity in males, female somatic tissues, and during development of both males and females.

I factor transposition and transcripts are rare in inducer lines. Conversion of transgenic lines to the inducer cytotype reduced CAT expression in the ovaries indicating that the 5' 186bp of *I* is a target for cytotype regulation. SF and RSF flies produces by matings between transformed and host lines showed that the non-reciprocal maternal effect of I-R dysgenesis also acts on nucleotides 1-186 of *I*.

The implications of these results for transcription mechanisms and transposition regulation are discussed.

TABLE OF CONTENTS:

CHAPTER 1

INTRODUCTION

1.1 I-R Hybrid Dysgenesis in *Drosophila melanogaster*:

Hybrid dysgenesis is the name given to a set of characteristics manifested in the progeny of particular *Drosophila melanogaster* interstrain crosses. This phenomenon was first noticed when females of long established laboratory stocks were crossed with males from recently isolated populations. The progeny of such crosses were partially sterile and the progeny of the reciprocal cross were fertile. Two independent systems of hybrid dysgenesis were distinguished and called P-M, denoting paternal and maternal, and I-R denoting inducer and reactive (Kidwell, 1979). All strains of *D. melanogaster* can be categorised as inducer or reactive, and paternal or maternal. The P-M system of hybrid dysgenesis shall be discussed in detail in section 1.4.1.

I-R hybrid dysgenesis is seen only in the female progeny of crosses between males of an inducer strains and females of a reactive strains. The female progeny are partially sterile in that they lay a normal number of eggs but a proportion fail to hatch. This type of sterility is termed SF (*stérilité femelle*) and these females are called SF females. Female progeny of the reciprocal cross are fertile and are called RSF females. Crosses within inducer or reactive stocks also give fertile progeny. Table 1.1 summarises the outcome of inducer and reactive stock matings.

TABLE 1.1

	♀ REACTIVE	♀ INDUCER
♂ REACTIVE	FERTILE	RSF ♀
♂ INDUCER	SF ♀	FERTILE

The genetic determinant controlling I-R hybrid dysgenesis is a transposable element called the *I* factor. When discussing the various characteristics of this system I shall refer to functional transposons as *I* factors, and those that are defective or have not been shown to transpose as *I* elements.

1.1.1 The Dysgenic Traits Associated with I-R Hybrid Dysgenesis:

SF Sterility:

SF females have normal ovaries but a percentage of their eggs fail to hatch due to developmental arrest during early cleavage divisions (Picard *et al.*, 1977; Lavige and Lecher, 1982; Lavige, 1986). The percentage of eggs that hatch is called the 'hatchability' and is used as a measure of SF sterility. Hatchability can range from 0% to 100% depending on the inducer and reactive parents used, the age of the SF females, and the temperature at which they are reared. Different inducer and reactive strains vary in 'strength' - strong inducer strains being those that father SF females with very low hatchability, and weak inducer strains fathering SF females with higher hatchability. Similarly strong reactive females produce SF daughters that are very infertile and weak strains, more fertile SF daughters. The strength of inducer ability does not vary a great deal from strain to strain. However there is considerable variation in so called 'reactivity' (Bucheton *et al.*, 1976) (see section 1.1.10). The weakest reactive strains are called neutral stocks and produce fertile daughters when crossed as females to inducer males, or as males to reactive females (Kidwell *et al.*, 1983).

Age affects SF sterility in that hatchability increases in older SF females and can reach normal levels in 15-20 day old adults (Bucheton, 1978). An increase in temperature reduces fertility at all developmental stages except if applied during oogenesis when it causes an increase in fertility. The affects of age and temperature shall be discussed in more detail in section 1.1.10.

Other Dysgenic Traits:

Besides sterility, the germline of SF females show abnormalities manifested in their surviving progeny, including high mutation frequency, and X-chromosome non disjunction (Kidwell *et al.*, 1977). Picard *et al.* (1978) showed that these dysgenic traits correlate with the severity of SF sterility by measuring the rate of mutation and X-chromosome non disjunction in the offspring of SF females derived from strong and weak inducer and reactive parents. Significantly higher frequencies of XO males and

visible mutation were obtained from SF females with low hatchability, than those with higher hatchability. In addition the frequency decreased as SF females aged. Proust and Prudhommeau (1982) found a similar correlation of mutation frequency with SF sterility, suggesting both abnormalities are caused by the same phenomenon. In both studies mutational hotspots were found which lie within the *white*, *cut*, and *yellow* loci. About 80% of these mutations arose in single individuals and the remainder in small clusters of two to four individuals. Picard *et al.*(1978) interpreted this to mean that the causal event happens shortly before meiosis. However these mutations were not analyzed at the molecular level and could represent different mutations producing the same phenotype, or mutational hotspots.

1.1.2 The Genetic Properties of the Inducer State:

The paternally inherited determinants contributing to hybrid dysgenesis are called *I* factors. *I* factors are chromosomally born on so called i^+ chromosomes of which inducer strains have at least one, and can be any one of the four *D. melanogaster* chromosomes (Picard, 1976; Picard and Pélisson, 1979). *I* factors are stably inherited in a Mendelian fashion in inducer strains and in the brothers of SF and RSF females (Picard, 1976). In contrast *I* factors are unstable in the germline of SF females and reactive chromosomes (*r*) (both homologous or non-homologous) can acquire *I* factors, becoming i^+ at a frequency of 100% per generation (Picard, 1976). i^+ chromosomes can also gain *I* factors but the frequency of these events is more difficult to measure. This process is called chromosome contamination. The germline of RSF females undergo chromosomal contamination but at a frequency five fold lower than in the SF female germline (Picard, 1976). Chromosome contamination was interpreted to mean that the *I* factor is a transposable element, the mobility of which is restricted to the germline of SF and RSF females.

In addition to i^+ chromosomes, inducer strains can carry i^o chromosomes, that in general have no inducer potential as males from strains with a genome consisting entirely of i^o chromosomes do not normally give rise to SF females when crossed with reactive females (Pélisson and Bregliano, 1981). However i^o chromosomes are not

reactive because when i^o females are crossed with inducer males their daughters are fertile. Pélisson and Bregliano (1981) noticed that i^o males could spontaneously induce SF sterility. i^o chromosomes therefore contain *I* factors, possibly silenced by a position effect, that can regulate SF sterility but do not normally transpose. Position effect could therefore be one factor influencing in inducer strength. Differences intrinsic to the *I* factor also contribute to inducer strength (Pélisson, 1978). However *I* factor copy number does not as there is no strong additive affect of i^+ chromosomes (Pélisson, 1979).

1.1.3 Isolation of the *I* Factor:

I factor insertions were argued to be the causal agent of the mutator effect of I-R hybrid dysgenesis as *I* factor activity could not be separated by recombination from two independent I-R inducer mutations in the *white* locus called w^{IR1} and w^{IR3} (Pélisson, 1981; Bucheton *et al.*, 1984). The region of interest in these mutants was cloned and an apparently identical insert of 5.4kb was found at the same location in *white* (Bucheton *et al.*, 1984). Sequences homologous to the insert are found in both reactive and inducer strains, but only inducer strains have intact copies of large internal restriction fragments. Bucheton *et al.* concluded that only inducer strains have complete *I* factors but both inducer and reactive strains have defective *I* elements. *In situ* hybridisation to polytene chromosomes showed that these common *I* elements are clustered in the chromocentre and only inducer strains have euchromatic *I* sequences located at about 15 sites per haploid genome.

1.1.4 The *I* Factor is Related to LINEs:

The complete sequence of the *I* factor in w^{IR1} and the terminal sequences of 13 others has been determined (Fawcett *et al.*, 1986; Crozatier *et al.*, 1988; Pritchard *et al.*, 1988; Busseau *et al.*, 1989a; Busseau *et al.*, 1989b; Jensen and Heidmann, 1991). The main features of the *I* factor are diagrammed in figure 1.1 and classify it with long interspersed nuclear elements (LINEs, discussed in detail in section 1.2). The *I* factor has no direct or inverted repeats but has the trinucleotide TAA repeated four

to seven times at its right hand end. It is flanked by duplications of target genomic sequences that vary in length from 10 to 14bp. The target sites have very little sequence similarity except that they are AT rich and some (5 out of 13) start with the sequence TAA.

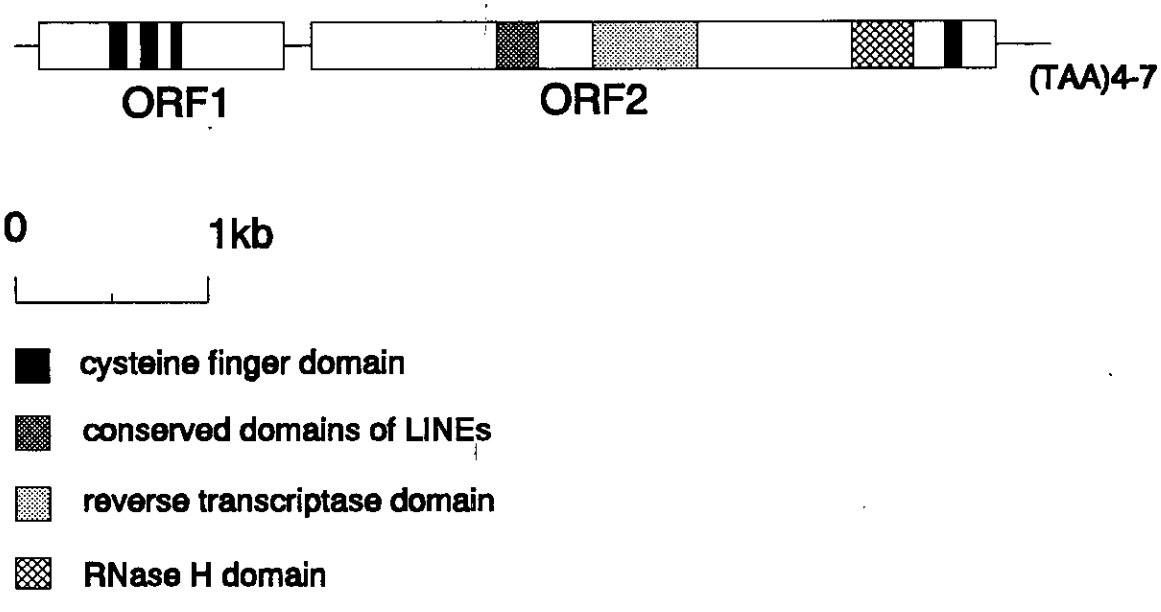
The *I* factor encodes two long open reading frames (ORFs) separated by 53bp. ORF1 is 1287bp long and encodes a polypeptide with one full and two partial copies of the cysteine motif $CX_2CX_4HX_4C$ (where X is any amino acid). This motif is found in the *gag* polypeptide of retroviruses, the *gag*-like ORF of retrovirus like transposons, the coat protein of cauliflower mosaic virus, and ORF1 of other LINE-like elements (Doolittle *et al.*, 1989). ORF1 of the *I* factor binds both DNA and RNA but is apparently not specific for *I* factor sequences (T. Paterson, personal communication). The binding domains span ORF1 as recombinant proteins with various deletions that cover the entire ORF retain non-specific nucleic acid binding properties.

ORF2 has blocks of amino acid similarity with viral reverse transcriptase proteins. Retroviral polymerases have three enzymatic functions: an RNA dependent DNA polymerase, RNase H, and integrase. ORF2 of *I* has domains with similarity to the first two functions but not to integrase. RNA dependent DNA activity has been shown for ORF2 (T. Paterson, personal communication). ORF2 also has a domain particular to LINE-like elements (figure 1.1)(Finnegan *et al.*, 1988; Xiong and Eickbush, 1988a). Absence of this domain from other reverse transcriptase proteins suggests that it functions in part of LINE element lifecycle not shared with retroviral-like transposons or retroviruses. Near the carboxyl terminus of ORF2 there is a cysteine motif with the spacing $CX_2CX_7HX_4C$. Similar motifs have been found at this location in other LINE-like elements but no function has yet been assigned, although it too could be involved in nucleic acid binding.

The ends of the 13 full length *I* factors that have been sequenced are highly conserved showing variation only in the number of TAA repeats and the identity of the nucleotide at position +3 that can be a G or a T.

Figure 1.1 : **The structural features of the *I* factor.** The key defines each region referred to in the text. Untranslated regions are shown as a single line.

Figure 1.1


ORF1
ORF2
(TAA)4-7
0
1kb
cysteine finger domain
conserved domains of LINEs
reverse transcriptase domain
RNase H domain

1.1.5 The 5.4kb *I* Factor is the Determinant of I-R Dysgenesis:

Conclusive proof that the 5.4kb element in w^{IR3} is the inducer determinant of I-R hybrid dysgenesis was provided by Pritchard *et al.* (1988) by introducing the cloned w^{IR3} *I* factor into reactive flies by P-element transformation. The resultant transformants became strong inducer strains shown by the ability of males to induce SF sterility, and the absence of SF sterility when females were crossed with inducer males.

P-element transformation was used because transformed lines were not recovered when the plasmid born *I* factor with no *P* transformation sequences were injected into reactive or dysgenic embryos. Autonomous *I* factor transposition events possibly do not occur because transposition is much later than *P* by which time the DNA is too dilute or degraded. The *I* factor is however able to transpose autonomously once integrated, as copies were found at new euchromatic locations that were no longer associated with vector or *P* element sequences. One of the new copies of *I* was cloned and its ends sequenced. The only difference was that the number of TAA repeats had increased from five to six, presumably due to polymerase slippage during replication. Such slippage is not uncommon as Jensen and Heidmann (1991) reported a similar event following *I* transposition.

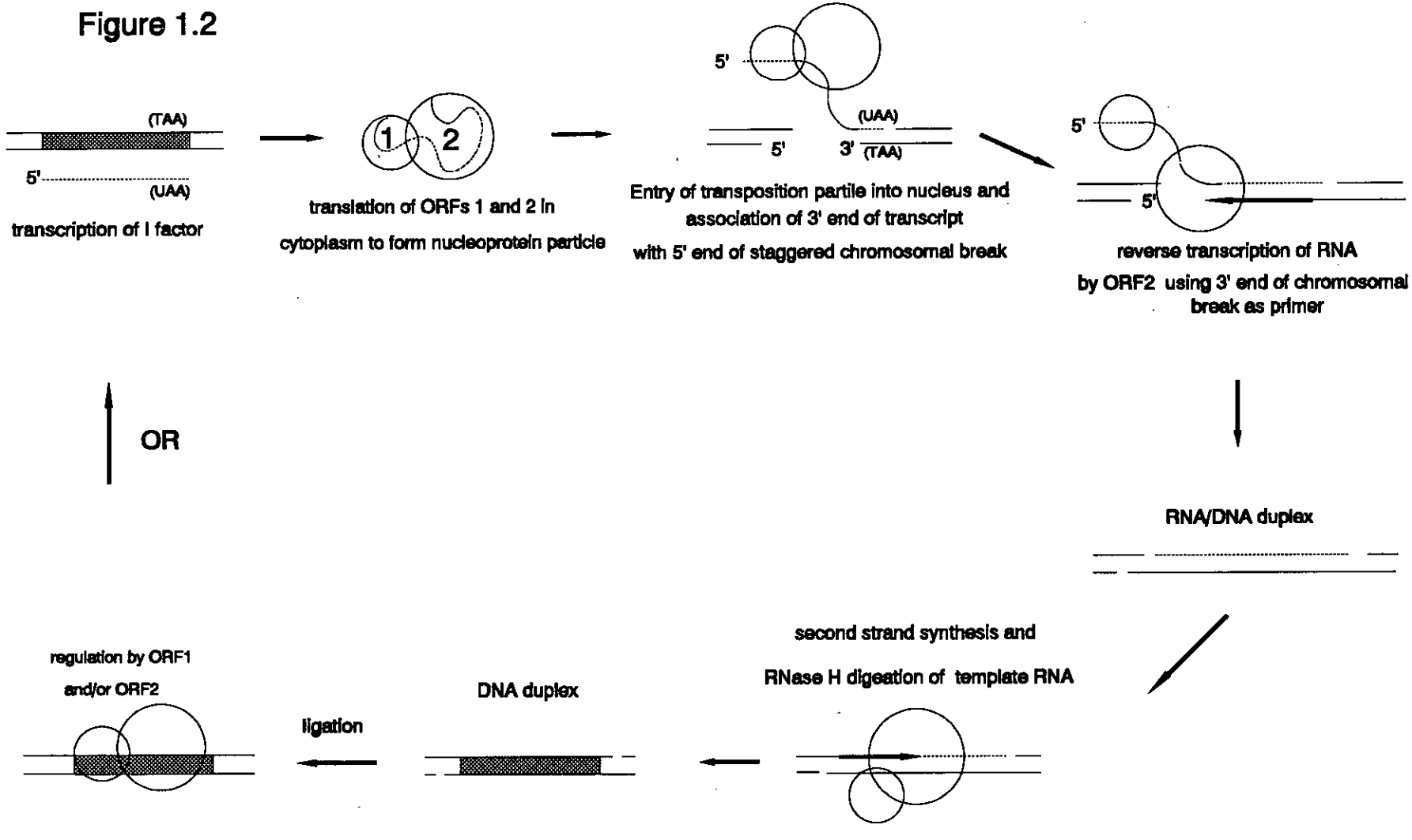
A complete *I* factor from the *D. melanogaster* sibling species, *D. teissieri*, has been cloned and sequenced (Abad *et al.*, 1989). It shows 85% homology (at the nucleotide level) to the *D.melanogaster I* factor and can also transpose autonomously and confer inducer cytotype when introduced into reactive *D. melanogaster* strains by P-element transformation.

1.1.6 A Model for *I* Factor Transposition:

Pritchard *et al.*, (1988) proposed a model for *I* factor transposition, based on that suggested for the *Zea mays* LINE-like transposon, *Cin4* (Schwarz-Sommer *et al.*, 1987). Figure 1.2 illustrates the model. Firstly, a full length transcript of an *I* factor is synthesised, initiating at the first nucleotide of *I* from an internal promoter. Translation of this RNA in the cytoplasm and association of the *I* factor polypeptides

Figure 1.2 : **A model for *I* factor transposition.** The full length *I* factor is hatched. RNA is shown as a dotted line and the ORFs as circles.

Figure 1.2


(TAA)
5'
(UAA)
transcription of I factor
1
2
translation of ORFs 1 and 2 in
cytoplasm to form nucleoprotein particle
5'
(UAA)
5'
3'
(TAA)
Entry of transposition partile into nucleus and
association of 3' end of transcript
with 5' end of staggered chromosomal break
5'
5'
reverse transcription of RNA
by ORF2 using 3' end of chromosomal
break as primer
RNA/DNA duplex
second strand synthesis and
RNase H digestion of template RNA
DNA duplex
ligation
regulation by ORF1
and/or ORF2
OR

with the RNA forms a 'transposition particle'. The particle enters the nucleus and associates with naturally occurring staggered nicks in the genome possibly by base pairing between the UAA at the 3' end of the RNA intermediate and TAA if present at the target site. Using the 3' end of the staggered DNA as a primer, the RNA dependent DNA polymerase activity of ORF2 makes a cDNA of the *I* factor RNA. Second strand synthesis is catalysed by the DNA dependent DNA polymerase function of ORF2. *Drosophila* DNA ligase seals the ends and further transcription and transposition is regulated by ORF1/ORF2 binding.

The evidence for this model comes from several observations:

1. i^+ chromosomes never become i^o (Picard, 1979), and I-R induced mutations never revert (Pélisson, 1981), indicative of a replicative process with no excision.

2. The structure and coding capacity of LINEs suggest replication is through reverse transcription of an RNA intermediate.

3. Euchromatic *I* elements are often truncated at their 5' end possibly due to premature termination of reverse transcription.

Retrotransposition has been confirmed for *I* by insertion of an intron which is removed in transposed copies of the element (Pélisson, 1991; Jensen and Heidmann, 1991). Pélisson inserted the intron within ORF2 thereby disrupting it. Transposition events were recovered, albeit at a low frequency (at less than 10^{-2} of the normal frequency), indicating that the product of ORF2 can work in *trans* but rather inefficiently. In the experiments of Jensen and Heidmann, the intron was in the 3' untranslated region (UTR), and transposition events were observed in the germline of SF females when males carrying the transgene were crossed with reactive females. The *I* transgene was therefore autonomous but transposed very inefficiently (10^{-3}-10^{-4} fold lower than normal), and did not induce SF sterility.

Transposition through an RNA intermediate has been demonstrated for only one other LINE-like element, L1Md of mouse (Evans and Palmiter, 1991). Retrotransposition probably mediates mobility of all LINE-like elements. However the *I* factor offers the best system for elucidation of mechanistic details as it is the

only element for which the conditions of high transposition rates are known.

1.1.7 A Candidate for the Transposition Intermediate:

A 5.4kb *I* factor transcript has been identified, the presence of which correlates with *I* factor transposition (Chaboissier *et al.*,1990). This transcript is confined to ovary tissues, and is most abundant in young SF females and decreases as they age. The transcript is found in lower amounts in the ovaries of RSF females, and may be present in inducer ovaries as a band of the predicted size has been found in S1 mapping experiments of the 5' end, but not on Northern blots. The correlation of full length RNA with transposition frequency suggests transposition is controlled at the level of transcription or RNA stability. The 5' end of the RNA is homogenous, being at or very near position +1, and the 3' end is not polyadenylated and lies at the TAA repeats. The 3' end of most LINE-like elements have a polyA sequence, presumably from a polyadenylated RNA intermediate. Termination of *I* transcription at the TAA repeats suggest a mechanism similar to that of the sea urchin H2A histone RNA that is synthesised by RNA polymerase II and is poly(A)$^-$. Termination of transcription occurs at, or near to, and requires a sequence that includes $TAA(TAAA)_3$ (Briggs *et al.*, 1989). Both euchromatic and pericentromeric *I* elements have been describe that have the sequence $TAA(TAAA)_n$ rather than $(TAA)_n$ (Busseau *et al.*, 1989a; Crozatier *et al.*, 1988). This types of 3' end might be present in a different subgroup of *I* factor as yet unidentified, or $TAA(TAAA)_3$ could represent the original termination signal from which the *I* factor has diverged.

No splicing has been detected in the region between the two ORFs. Translation of ORF2 must therefore rely on reinitiation of ribosomes or translational frameshifting. Both of these mechanisms are inefficient, thus translation of ORF2 is a possible means of transposition regulation.

1.1.8 *I* Elements:

1. *I* Elements Specific to Inducer Strains:

I factors can loose 5' sequences upon transposition rendering the transposon

defective and as such are termed *I* elements. The 5' end is lost prior to integration as deleted elements are flanked by target site duplication sequences. Molecular analysis of several euchromatic *I* sequences revealed 5' truncation to be a common occurrence (about 50% of euchromatic sequences analyzed are truncated) (Busseau *et al.*, 1989a; Busseau *et al.*, 1989b; Crozatier *et al.*, 1988; Jensen and Heidmann, 1991). The size of 5' truncated *I* elements ranges from 3893 to 38bp (excluding TAA repeats). The point at which the truncation lies is apparently random for *I*, unlike the *Cin4* element for which truncation occurs immediately 3' to a sequence that resembles the target site duplication sequence (Schwarz-Sommer *et al.*, 1987). Busseau *et al.* (1989a) suggested arrest of reverse transcription of *Cin4* would occur at such sequences if the integration site is a 5' overhang (figure 1.3a). *I* truncation would be common and occur at sequences with no similarity to target DNA if the RNA template associates with a 3' overhang (using the 3' UAA to base pair with genomic ATT sequence)(figure 1.3b).

2. *I* Elements Common to Inducer and Reactive Strains:

Pericentromeric *I* elements have been cloned from inducer and reactive strains and characterised at the molecular level (Crozatier *et al.*, 1988). Restriction analysis indicates that many of them have long regions with homology to ORF1 but are highly rearranged and are often repeated in tandem arrays with flanking sequences that are themselves transposable elements (Vaury *et al.*, 1989). Sequence analysis of three *I* elements from a strong reactive strain showed that the 5' 1286bp to be 93-95% homologous to the *I* factor, with several substitutions, insertions and deletions that disrupt ORF1.

Centromeric *I* elements are transcribed in both inducer and reactive strains giving rise to two abundant RNAs of 1.1 and 2.2kb, that hybridise strongly with ORF1 sequences, and a smear of other transcripts (Chaboissier *et al.*, 1990). The 1.1 and 2.2 kb RNAs are absent from ovary tissues.

1.1.9 Specific Rearrangements Generated by Euchromatic *I* Elements:

Several of the mutations arising in the progeny of SF females have been

Figure 1.3 **Generation of 5' truncated elements of *Cin4* (a) and the *I* factor (b)** (taken from Busseau *et al.* (1989a). The model in figure 1.3a envisages that the *Cin4* transposition intermediate integrates preferentially at nicks in the genome with 5' overhangs thus facilitating base pairing between the left hand 5' overhang of the integration site and the RNA intermediate at a sequences with some similarity. The model in figure 1.3b envisages *I* integration preferentially at DNA nicks with 3' overhangs. Premature termination of reverse transcription would be facilitated by template switching from the RNA intermediate to the left hand 3' overhang.

Figure 1.3a

Cin 4

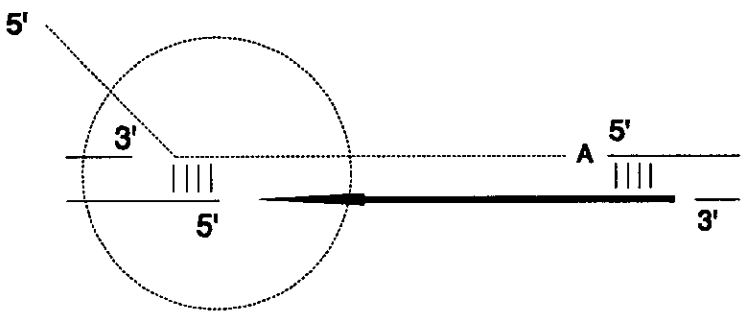
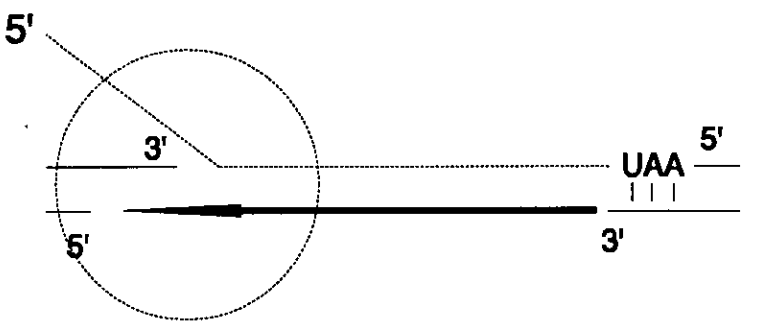


Figure 1.3b

I element



analyzed at the molecular level (Sang *et al.*, 1984; Fawcett *et al.*, 1986; Busseau *et al.*, 1989a; Lynch, 1989). Association of the *I* factor with an I-R induced mutation can be straight forward insertion of a full length or 5' truncated element as with the w^{IR1} and y^{IR3} insertions respectively (Sang *et al.*, 1984; Busseau *et al.*, 1989a), or can involve more complex rearrangements. Busseau *et al.* (1989b) showed y^{IR5} and y^{IR6} both involve inversions in which *I* factor sequences are associated with at least one breakpoint. Lynch (1989) analyzed several *white* mutations arising in the progeny of SF females that inherited the w^{IR1} mutation paternally. Some of these so called w^{IR1} derivatives were deletions downstream of the w^{IR1} *I* factor that may have arisen by insertion of a second *I* factor downstream in direct orientation, followed by recombination between the two and deletion of the intervening sequences. Three I-R induced mutations described to date are deletions with no apparent involvement of *I* factor sequences. These mutations are w^{IR7} and w^{IR8} in the *white* gene, and y^{IR7} in the *yellow* gene (Lynch, 1989; Busseau *et al.*, 1989b). The origin of these mutations is not understood.

1.1.10 Reactivity:

Reactivity is the maternally inherited component of I-R hybrid dysgenesis. It is a cytoplasmic condition within the oocyte that, on introduction of *I* factors paternally, causes SF sterility, permits transposition of the *I* factor at a high frequency, and increases mutational frequency and chromosome non disjunction in the germline of females. Reactivity is measured by the sterility of SF daughters when females are crossed with males from standard inducer stocks. Variation in reactivity strength is seen from one strain to another, manifested as different severities of SF sterility (Bucheton *et al.*, 1976). Within a particular strain, there can be different strengths of reactivity, but homogenous stocks can be established from such lines by selection (Bucheton, 1973: Picard *et al.*, 1972).

Maternal Inheritance of Reactivity:

Reactivity is transmitted from mother to daughter largely irrespective of the

father's genotype. This was demonstrates by measuring the reactivity of daughters from strong reactive mothers and weak reactive fathers. Reactivity of the daughters is strong whereas, daughters of the reciprocal cross are weakly reactive. The level of reactivity is therefore determined over a single generation by maternal factors. However, when the genotype of a strong reactive strain is replaced by that of a weak strain, reactivity evolves over several generations to that of the weak strain. Again the same is true of the reciprocal cross (Bucheton and Picard, 1978). Reactivity is therefore determined by maternal factors but is ultimately dictated by the genome. Each of the three major chromosomes contribute to the genetic control, their effect being additive. This indicates several genetic factors located on each chromosome are involved. The switch in reactivity strength takes at least 10 generations despite completion of genome replacement after only two.

Transmission of Reactivity from SF and RSF Females:

The fertility of an SF female is independent of her mate. However the fertility of the daughters of both SF and RSF females depends on the genotype of their fathers (Picard, 1978a). SF females from a strong reactive maternal line give rise to daughters with reduced fertility whatever the SF mate. However the degree of fertility of the daughters of SF females is lowest when their father is from a strongly reactive line and highest when from an inducer line. The same effect, albeit less severe, occurs in the daughters of RSF females. Sterility in each category therefore correlates with a higher ratio of $r{:}i^{+}$ chromosomes, the ratio being determined by the paternal genome.

Effect of Age and Thermic Treatment on Reactivity:

SF sterility decreases with age indicative of a decrease in reactivity. This effect is heritable and cumulative in that aged reactive females give rise to more fertile SF daughters, than young reactive females. The SF sterility of the daughters of aged reactive females is further reduced if they are from a maternal line that has been maintained through aged females over several generations. This cumulative effect

plateaus after three or four generations at which time hatchability of SF females can reach near normal levels (Bucheton, 1979a).

Reactivity is also influenced by heat treatment but the effect is more complex with a decrease in reactivity only when heat treatment (a shift in temperature from 20°C to 29°C) is applied during late oogenesis (5-9 day old adults), again this effect is heritable and cumulative. At all other times during development heat shock causes an increase in reactivity. Removal of the heat treatment after late oogenesis causes reactivity to increase to a level greater than if no heat treatment were applied (Bucheton, 1979b).

The effects of both ageing and heat shock are reversible with original levels of reactivity recovered five or so generations after removal of either treatment (Bucheton, 1979a).

Ageing has the same effects on the poorly fertile daughters of SF and RSF females, and heat treatment on SF daughters (Picard, 1978a).

No simple solution exists to explain the involvement of reactivity in I-R hybrid dysgenesis. It has been established that the genotype ultimately controls reactivity strength but perpetuation of maternal levels of reactivity through several generations after genome replacement is not understood. One can envisage positive feedback control that maintains reactivity at a level dictated by chromosomal factors in previous generations genome. However persistence of such control through the numerous cell divisions that occur over the eight or so generations that follow genome replacement is a difficult to explain.

1.1.11 Regulation of the Inducer-Reactive Interaction:

In discussing regulation of the I-R interaction one must consider several points:

1. SF sterility in the daughters of SF and RSF females is proportional to the number of reactive chromosomes present.
2. Transposition of *I* is necessary but not sufficient for SF sterility, as the *I* factor is mobile in fertile RSF females.
3. The *I* factor encodes all components that confer the inducer phenotype,

shown by conversion of a strong reactive strain into a strong inducer strain following P-element transformation of a cloned *I* factor (Pritchard *et al.*, 1988).

4. The I-R interaction is non-reciprocal suggestive of maternal inheritance of an *I* factor regulator. The concentration of the maternal factor would be slightly reduced in the RSF zygote. In contrast, in SF individuals, the regulator would be greatly diluted upon introduction from sperm (or perhaps absent from sperm), thus alleviating regulation, and activating *I* factor expression/transposition in the female germline.

5. Reactivity is not simply the absence of this regulatory substance as reactivity can vary in strength.

Point 1 suggest that there may be a threshold number of *I* factors above which SF sterility is no longer manifested. To attain this threshold copy number transposition must proceed, which brings in point 2. In RSF flies transposition occurs at a frequency five times lower than in isogenic SF females, presumably because of a maternally inherited *I* factor encoded regulator. The severity of SF sterility correlates with *I* factor transposition (Picard, 1978b), therefore the reduced rate of transposition in RSF females may be insufficient to manifest sterility.

Pélisson and Bregliano (1987) investigated whether the number of *I* factors influences the level of fertility in females. They did this by reconstructing the reactive genotype in the progeny of SF females. The resultant progeny, that differ only from the original reactive strain by *I* factors spread by chromosome contamination in the SF germline, were intercrossed to maintain the line. SF sterility was perpetuated by selecting females that had lowest fertility at each generation. The proportion of females with SF sterility decrease from 80% to 20% over the first four generations and stabilised at this level. The number of euchromatic *I* elements was measured by *in situ* hybridisation to larval polytene chromosomes and was found to reach an equilibrium of six by the third generation. The authors suggested this meant that SF sterility could be maintained in the absence of transposition. This is not necessarily the case as the experiment may select against individuals in which transposition has occurred. Selection of mothers for continuation of the next generation is based on

poor fertility - presumably manifested as a result of transposition in her germline. The germline cells that undergo high levels of transposition may die, or give rise to fertile individuals if a threshold number of *I* factors is attained that confers inducer cytotype and therefore fertility. Such females would be discarded. The females selected to continue the next generation would be those originating from germline cells that escaped transposition. Because the copy number has not increased it appears as if transposition is not a prerequisite for SF sterility. The copy number would however still be below the threshold level necessary to suppress high frequencies of transposition in her germline and thus result in SF sterility.

This theory was tested by measuring the *I* element copy number at the 15th generation for the 60 most fertile and the 60 least fertile females. The mean number of euchromatic *I* elements was 10.7 and 8.7 (with t=2.2 for 12 degrees of freedom) respectively. Because of the insignificant difference between high and low fertility, the authors concluded that the SF sterility is not switched off when *I* factors copy number exceeds a certain level. However *in situ* hybridisation does not distinguish between functional and defective euchromatic *I* elements. Therefore the slight difference in copy number that correlates with fertility may indeed reflect a threshold of *I* factors above which transposition is suppressed to level that does not trigger SF sterility.

1.1.12 Phylogenetic Distribution of *I* factor Sequences and Evolutionary Implications:

D. melanogaster is classified in the *melanogaster* species group that is within the *Sophophora* subgenus. Of the 19 species tested in the *melanogaster* species group, sequences that hybridise to the *I* factor have been found in 18. In addition *I* factor similarity is found in three more distantly related species within the *Sophophora* subgenus, and one in the *Dorsilopha* subgenus (Bucheton *et al.*, 1986), suggesting *I* elements are an old component of the genome. Greatest similarity to the *I* factor is found in its closest sibling species, *D. simulans*, *D. mauritiana*, and *D. sechellia* of which *D. simulans* is most closely related, being separated from *D. melanogaster* 2-3

million years ago (Carlou, 1987). The *I* element in these sibling species differ from the *D. melanogaster I* factor in having an extra *Hind*III site. All the apparent full length *I* elements of *D. mauritiana* and *D. seichellia* have this site and approximately 60% of the *D. simulans* elements have it, the remainder having an restriction pattern identical to the *D. melanogaster I* factor.

In situ hybridisation of *I* factor sequences to the polytene chromosomes of *D. simulans*, *D. mauritiana*, and *D. teissieri* show their distribution to be similar to that of *D. melanogaster* inducer strains, with *I* elements at dispersed euchromatic sites, and clustered at the chromocentre (Bucheton *et al.*, 1986; Simonelig *et al.*, 1988). On Southern blots all *D. simulans* and *D. teissieri* strains tested to date appear to be like inducer strains as they have homology to large internal restriction fragments of the *D. melanogaster* and *D. teissieri I* factors respectively. At least some of these *I* elements are functional as the pattern of euchromatic sites differs from strain to strain (Simonelig *et al.*, 1988).

The phenomenon of I-R hybrid dysgenesis is apparently a consequence of introduction of the *I* factor into *D. melanogaster* in recent times. This is based on the observations that long established laboratory strains are usually reactive, and recently isolated wildtype strains are inducer. The alternative explanation that reactive flies result from selection against *I* factors in the laboratory environment is unlikely as inducer strains maintained under laboratory conditions for 5-10 years do not become reactive (Kidwell, 1983). The origin of the *I* factor in *D. melanogaster* is possibly through horizontal transfer from *D. simulans* as it is the only species that have *I* factors with an identical restriction map. Transfer of the *I* factor could have been via a vector such as a virus or a mite as suggested for the *P* factor (Houck *et al.*, 1991), or by productive mating between the species that normally give rise to sterile offspring.

Vaury *et al.* (1990) have suggested that the *I* elements situated in the chromocentre are vestiges of the original *I* factor that spread through the *melanogaster* subgroup. During evolution some species such as *D. simulans* would have retained

the *I* factor and others, such as *D. melanogaster*, lost it. Evidence for this comes from sequence analysis of three *I* elements from *D. melanogaster* reactive strains. These *I* elements have a significant number of silent base substitutions common to at least two of them, that differ at the corresponding position in both the *D. melanogaster* and *D. teissieri I* factors. The *I* factors of *D. melanogaster* and *D. teissieri* in turn share a number of bases that differ at the corresponding location in the euchromatic *I* elements. The shared nucleotides in the two *I* factors may represent the sequence that existed before the divergence of *D.teissieri* and *D. simulans* (assuming this is where the I factor of *D. melanogaster* came from), and the positions not present in either *I* factor may represent the sequence of the progenitor *I* factor that was immobilised in the chromocentre before the divergence of the two species.

1.2 LINES AND RELATED ELEMENTS:

The *I* factor falls into a class of transposable elements collectively known as LINEs or non-viral retrotransposons (Weiner *et al.*, 1986). LINEs are found in a diverse array of eucaryotic organisms ranging from fungi to man, but transposition has been demonstrated for only some. LINEs are characterised by several structural features. Full length elements are between four and seven kb long. They have no terminal direct, or inverted repeats but are flanked by rather long target site duplications with no sequence specificity (in most cases). They encode one, or more frequently, two long ORFs on the same DNA strand (plus strand) with putative enzymatic domains as described for the *I* factor. The 3' end of the plus strand ends in a poly(A) or adenine rich sequence. The number of $CX_2CX_4HX_4C$ motifs in ORF1 varies from one element to another but ranges from one to three copies. The structural features of various LINE-like elements are summarised in table 1.2.

1.2.1 Mammalian LINEs:

LINEs (or L1s) have been found in all marsupials and placental mammals examined (Burton, *et al.*, 1986). The most thoroughly studied mammalian LINE is *L1Hs* (L1 of *Homo sapiens*) only a few copies of which are thought to be functional

TABLE 1.2: T.S.D. = target site duplication; a - possibly ORFs but there are a few interruptions by stop codons; b - *Ara= A. arabiensis* element, *Aga- A. gambiensis* element; c - elements may represent tandem repeats of the same element; d - only one example sequenced; e - all elements appear to be trauncated at their 5' end so don't know if 5' heterogeneity exists; NM - not measured; NS -not sequenced; +/- - cannot distinguish between 5' heterogeneity and truncation as heterogeneity could be deletion of a few 5' bases.

References: 1 - Skowronski and Singer, 1985; 2 - Skowronski *et al.*, 1987; 3 - Loeb *et al.*, 1986; 4 - D'Ambrosio *et al.*, 1986; 5 - Fawcett *et al.*, 1986; 6 - Bucheton *et al.*, 1984; 7 - Bucheton, 1990; 8 - Di Nocera and Casari, 1987; 9 - Pardue and Dawid, 1981; 10 - Di Nocera, 1988; 11 - Driver *et al.*, 1989; 12 - Priimägi *et al.*, 1988; 13 - Jacubczak *et al.*, 1991; 14 - Dawid and Rebbert, 1981; 15 - Roiha *et al.*, 1981; 16 - Xiong and Eickbush, 1988a; 17 - Burke *et al.*, 1987; 18 -Paskewitz and Collins, 1989; 19 - Kimmel *et al.*, 1987; 20 - Murphy *et al.*, 1987; 21 - Askoy *et al.*, 1987; 22 - Askoy *et al.*, 1990; 23 - Gabriel *et al.*, 1990; 24 - Kinsey and Helber, 1989; 25 - Schwarz-Sommer *et al.*, 1987; 26 - Leeton, 1991.

Element	Size(kbs)	Copy No.	No. of ORFs	5' truncation	5' hetero-geneity	3'end	T.S.D. (bp)	Reference
Mammalian LINEs	6-7	~10^5	2	+	+	A rich with some T's	5-15	1,2,3,4
I	5.4	10-15	2	+	-	TAA_{4-8}	7-14	5,6,7
F	4.7	25-30	2	+/-	+	A_{12-30}	8-13	8,9
G	4.3	20	2[a]	-[c]	-[c]	A_{19}	9[c]	10
doc	4.9	NM	NS	+	+	A_{15-29}	7-13	11
jockey	5 and 2.5	NM	2	+/-	+	A_{5-29}	8-10	12
R1Dm	5.3	variable	2	+	+	Not A rich	14	13,14,15
R1Bm	5.1	25	2	-	+	Not A rich	14	16
R2Dm	3.6	variable	1	+	+	A_{13-22}	-	13
R2Bm	4.2	20-25	1	-	+	GA_4	-	17
Aga, Ara[b]	4.7-8.2	6, ≥20	NS	+	?[c]	A_{7-15}	17	18
ingi	5.2	200	3	+/-	+	A_{7-14}	4	19,20
SLACS	6-7	9	2	-	-	$A_{14,17}$	49	21,22
CRE-1	3.9	10	1	?[d]	?[d]	A_{27}	29	23
Tad	7	NM	NS	NS	NS	NS	14,17	24
Cin4	≥7	50-100	2	+	?[c]	A_{6-11}	3-16	25
del-1	4.5	270,000	2[a]	-	+	A_n	4-13	26

as the majority are truncated at their 5' end or internally rearranged (Fanning and Singer, 1988). At least some *L1Hs* are transposable as three *de novo* insertion events have been identified. Two of the insertion were into the *Factor VIII* gene of unrelated patients and lead to haemophilia A. The mothers of both patients had normal *Factor VIII* genes indicating the insertion event took place in the mothers' germline (Kazazian *et al.*, 1988). The third insertion was found in a patient with adenocarcinoma of the breast. An element transposed into one *myc* allele of the affected tissue, the surrounding normal tissue being free of the insertion (Morse *et al.*, 1988). *L1Hs* can therefore transpose in both germline and somatic tissues. The insertions in each patient were of defective *L1Hs* elements - the element in the *myc* locus was highly rearranged and both insertion events into the *Factor VIII* gene were by 5' truncated elements. *De novo* insertion of a highly defective elements suggests *L1Hs* 'transposase' can work in *trans*. However truncation of the 5' ends of the two insertions in the *Factor III* gene may have occurred during transposition.

The conditions under which *L1Hs* transposes are not known. However, high levels of a full length *L1Hs* mRNA corresponding to the plus strand have been found in a human teratocarcinoma cell line, Ntera2D1 (Skowronski and Singer, 1985). This transcript is approximately 6.5kb long, polyadenylated, and initiates with slight heterogeneity within one residue from the position defined as the 5' end of *L1Hs* (Skowronski *et al.*, 1988). As such, this RNA represents a putative transposition intermediate. Nineteen cDNAs from Ntera2D1 L1Hs RNA have been analyzed in detail (Skowronski *et al.*, 1988). Complete sequencing of one (cD11), and partial sequencing of the others, showed that all were slightly different indicating that they were transcribed from different *L1Hs* elements. The 5' UTR of complete elements is approximately 900bp and contains an internal promoter which directs transcription initiation, in a tissue specific fashion, at or near the beginning of the element (Swergold, 1990).

ORF1 is translated in Ntera2D1 cells as a 38kD polypeptide (Leibold, *et al.*, 1990) but its function remains to be tested.

The role of ORF2 has yet to be tested directly, however reverse transcriptase

activity has been detected in a high molecular weight complex isolated from Ntera2D1 cells (Deragon, *et al.*, 1990). *L1Hs* plus strand mRNA and a 37kD protein component (very similar in size to ORF1) were associated with the complex. It is possible that ORF2 encodes the reverse transcriptase activity and the complex represents a transposition intermediate.

The LINEs of some other mammals have been studied but in less detail. *L1Md* of *Mus domesticus* is repeated 10^5 times per haploid genome. The majority have a common 3' end but are truncated at their 5' end (Fanning, 1983; Voliva, *et al.*, 1983). The 5' end of full length *L1Md* elements differs from most other LINE-like elements in that they contain tandem repeats of one of two types of non-homologous sequences (type A=208bp; type F=206bp) (Fanning, 1983; Loeb *et al.*, 1986; Padgett *et al.*, 1988; Wincker *et al.*, 1988). Sequencing data of three elements revealed the copy number of A type repeats to be 4 and two-thirds, 2 and two-thirds and one and two thirds (Shehee *et al.*, 1987). F-type arrays have also been found with only two-thirds of a copy at the 5' most repeat but others which terminate within a few nucleotides of a complete unit also exist (Padgett *et al.*, 1988). Variability in copy number may be a consequence of the presumed mechanism of transposition. If each unit contains a promoter that directs transcription initiation one third of a unit distance from its 5' end, following reverse transcription of the transposition intermediate, an element that had n+1 promoter units will have n units. Repeats could be gained by unequal crossing over. Promoter activity has been detected from the 5' 1.6kb of the *L1Md* element but the precise sequences controlling transcription are unknown (Evans and Palmitter, 1991).

The rat, *Rattus norvegicus* L1 element (*L1Rn*) is quite distinct from other mammalian LINEs in that most of the $4x10^4$ copies are full length (6.7kb) (D'Ambrosio *et al.*, 1986) and at least some copies are active as allelic variation due to their insertion at a number of genetic loci has been found (Economou-Pachnis *et al.*, 1985; Lakshmikumaran *et al.*, 1985). The left hand ends of four full length *L1Rn* elements have been sequenced (Furano *et al.*, 1988) and one complete and often one partial copy of a 600bp promoter region identified. The promoter is CpG rich (>50%

G+C), with several putative Sp1 binding sites. *L1Rn* and *L1Md* 5' UTRs both have tandem repeats and are GC rich but they bear no sequence homology. The 600bp repeat of *L1Rn* can direct transcription in an orientation dependent manner in rat cell lines, and is sensitive to methylation (Nur *et al.*, 1988). However the precise sequences that direct transcription are unknown and the promoter detected in tissue culture remains to be shown to be the internal promoter used in transposition.

1.2.2 LINE-like Elements in *Drosophila*:

As well as the *I* factor, *D. melanogaster* contains repeated sequences called *F*, *G*, *Doc*, and *jockey* each possessing features of LINE-like retrotransposons. *F* is transposable as its location is highly polymorphic in different *Drosophila* stocks (Pardue and Dawid, 1981; Di Nocera *et al.*, 1983), and a recent *F* insertion event has been described (w^{i+A}) that causes reversion of the w^{ivory} mutation (Di Nocera *et al.*, 1983). The chromosomal distribution of *F* is like that of *I* in inducer strains, occupying both euchromatic and clustered centromeric sites (Pardue and Dawid, 1981). Apparent full length *F* element poly(A)$^+$ RNA is found in *Drosophila* Schneider line 2 tissue culture cells. These transcript are transcribed from an internal promoter situated at the 5' end of *F* that initiates transcription around position +11 (Minchiotti and Di Nocera, 1991).

The *G* element was first identified as an insert within an *F* element (Di Nocera and Dawid, 1983). It is structurally similar but bears no DNA sequence homology to *F* or *I* (Di Nocera, 1988).

Doc was isolated after it inserted into the *white* and *suppressor of forked* loci causing w^1 and $su(f)^{s2}$ mutations respectively (Driver *et al.*, 1989). The 5' end of eight *doc* elements have been sequenced. One is truncated, and in the other seven two classes of 5' end can be distinguished: five elements retain the sequence CATT and CAGT very close to the 5' end; the other two have quite different 5' sequences, but have the sequence CATT 15 or 19bp from the 5' end.

Jockey exists in two major classes - a 5kb full length element, and a 2.5kb internally deleted element. Both forms are transcribed in larvae, pupae, and adults of

different strains, and in *Drosophila* Schneider line 2 tissue culture cells (Mizrokhi *et al.*, 1988). Primer extension of transcripts from chromosomally encoded elements showed initiation is homogenous at nucleotide +7 (as defined by the sequence of a full length element, Priimägi *et al.*, 1988) from an RNA polymerase II promoter. Transient expression assays in Schneider line 2 cells demonstrated that the promoter is within the 5' 350bp of *jockey* and nucleotides +1 to +13 are essential.

1.2.3 Ribosomal Insertion Elements:

Elements resembling LINEs interrupt the 28S ribosomal genes (encoding the large subunit rRNA genes of eucaryotes) of most insects, and the nematode *Ascaris lumbricoides* (Jakubczak *et al.*, 1991; Back *et al.*, 1984; Neuhaus *et al.*, 1987). These sequences fall into two classes called type I (or *R1*) and type II (or *R2*) depending on their insertion site and presence of target site duplications (see table 1.1). The target site region of both elements lies in a highly conserved region - type I interrupts the 28S genes approximately two thirds from the 5' of a gene, and type II insert 74bp upstream of type I. A few copies of type I and II elements are found outside the 28S rDNA genes. The examples characterised to date have very similar target sequence specificity but are defective (Brown *et al.*, 1984; Xiong *et al.*, 1988).

Complete *R1* and *R2* elements of *D. melanogaster* and *B. mori* (*R1Dm*, *R2Dm*, *R1Bm* and *R2Bm* respectively) have been sequenced (Burke *et al.*, 1987; Xiong and Eickbush, 1988b; Jakubczak, *et al.*, 1990). *R2Bm* has been shown to encode an endonuclease with site specificity for the *R2Bm* target sequence (Xiong and Eickbush, 1988c).

Insertion sequences interrupt the 28S rRNA genes of the two sibling mosquito species *Anopheles gambiae* and *Anopheles arabiensis*, 634bp 3' to the *R1* insertion site (Paskewitz and Collins, 1989). These elements probably represent a third class of 28S rDNA insertion sequences because of their unique location, their polyadenylation (found in *R2* but not *R1*), and their target site duplications (found in *R1* but not *R2*).

R1 and *R2* elements are always in the same orientation as rDNA transcription

whereas the *Anopheles* insertions are in the opposite orientation. It has been suggested that rDNA insertion elements lack a promoter and instead transcripts are readthrough from the rDNA promoter (Burke *et al.*, 1987). The insert in *Anopheles* cannot rely on this mechanism for transcription and mobilisation implying that this element, and possibly *R1* and *R2* as well, contain an internal promoter that directs transcription of the transposition intermediate.

1.2.4 LINEs in *Trypanosomids*:

Three *Trypanosomid* LINEs have been characterised to the DNA sequence level - *ingi* (or *TRS*) of *Trypanosoma brucei*, *SLACS* of *T. brucei gambiense*, and *CRE1* of *Crethidia fasciculata* (Kimmel *et al.*, 1987; Murphy *et al.*, 1987; Aksoy *et al.*, 1990; Gabriel *et al.*, 1990). *Ingi* was first found in a sequence highly homologous to a short poly adenylated mobile element, *RIME* (Hasan *et al.*, 1984). In several independent clones, *ingi* is found inserted at a precise nucleotide in *RIME* with no target site duplications suggesting that the ends of *ingi* are *RIME*. Both strands of *ingi* are transcribed but the most abundant RNAs are 5-9kb long, poly(A)$^+$, transcribed from the coding strand, and their synthesis is moderately sensitive to α-amanatin (Murphy *et al.*, 1987).

SLACS is found inserted into nine of the 300 spliced-leader RNA (SL or miniexon) genes in *T.b. gambiense*. The 5' end of all nine copies are identical but their size ranges from 6-7.2kb, thought to be due to variation in the number of repeats of a 185bp sequence found in the 5'UTR (Aksoy *et al.*, 1990; Aksoy *et al.*, 1987). This structure is reminiscent of that in the 5' UTR of *L1Md* and *L1Rn*.

CRE1 inserts into the miniexon genes at exactly the same site as *SLACS*, but in *Crithidia fasciculata*. *CRE1* is thought to be actively transposing as its rate of rearrangement is higher than can be explained by meiotic unequal crossing over of the miniexon repeats alone (Gabriel *et al.*, 1990). Gabriel and Boeke (1991) have demonstrated that the single large ORF of *CRE1* encodes a reverse transcriptase, activity of which is enhanced upon micrococcal nuclease digestion. This suggests that the protein is tightly associated with nucleic acids, possibly to ensure correct template

usage.

1.2.5 *Tad* - a Retrotransposon in *Neurospora crassa*:

Like *I*, the *Neurospora crassa* element *tad*, is mobilised when strains carrying it are crossed with strains that do not. *Tad* is only found in the Ivory Coast strain of Adiopodume origin (Kinsey, 1989). It was isolated as a 7kb insert in the *am* locus following a cross between the Adiopodume strain and a laboratory strain (Kinsey and Helber, 1989). It has been suggested that *tad* is a LINE like element because it creates long target site duplications upon insertion, and has no terminal inverted or direct repeats, but to date the complete *tad* sequence is lacking (Kinsey and Helber, 1989). *Tad* transposes via a cytoplasmic intermediate, shown in heterokaryons formed between tad^+ and tad^- strains (Kinsey, 1990). The naive nuclei acquired *tad* and upon serial transfer the copy number appeared to increase indicating transposition was in progress. The nuclei of heterokaryons do not fuse therefore *tad* mobility must be mediated through a cytoplasmic factor. The identity of the transposition intermediate remains unknown and to date only heterogenous RNAs complementary to either strand of *tad* have been detected (Kinsey, personal communication).

1.2.6 LINEs in Plants:

Only two LINEs have been reported in plants - both in monocots: *cin4* of *Zea mays* and *del2* of *Lilium speciosum* (Schwarz-Sommer *et al.*, 1987; Leeton, 1991). There are 50-100 copies of *cin4* per haploid genome, and a staggering 270,000 copies of *del2* making up approximately 4.2% of *L. speciosum* genome. All *cin4* elements isolated to date are truncated at their 5' which is thought to be a consequence of preferential termination of cDNA synthesis at a sequence resembling the genomic insertion site (see section 1.1.8).

LINEs, other than those described here, have been found in other organisms, and almost certainly new elements will continue to be reported in the future. Despite individual variations, it is obvious that LINEs are related. But the question of whether

they have diverged from a common ancestral element, or have converged to adopt a similar structure is open to debate.

1.3 RETROVIRUS-LIKE TRANSPOSONS:

Retrovirus-like transposons, as the name suggests, resemble retroviruses structurally, and in their method of replication. Both have long terminal repeats (LTRs) flanking a region encoding all proteins required for mobility. They are between 5 and 9kb long and are flanked by short duplications of host DNA, the length of which is constant for a given virus or transposon family. Like LINEs their mobility is mediated through reverse transcription of a RNA intermediate. However the two classes of element are structurally distinct, and consequently transposition differs mechanistically.

In *D. melanogaster* 10 independent retrovirus-like transposon (or *copia*-like) families have been characterised, each having 10-80 copies per haploid genome distributed, for the most part, randomly (Bingham and Zachar, 1989). Four families of retrotransposons, Ty1, Ty2, Ty3 and Ty4 are found in *Saccharomyces cerevisiae.* Characterisation of elements in the above two species, in combination with current understanding of retroviral structure and lifecycle, has led to much of what is known about retrotransposition of these elements.

1.3.1 The Structure of Retrovirus-like transposons:

Figure 1.4 illustrates the structure of retrovirus-like transposons. They terminate in LTRs, 200-500bp in length that themselves terminate in short imperfect inverted repeats of 6-12bp that often begin with TG or AG, and ending with CA or CT respectively. There are three distinct regions in the LTRs - U5, R, and U3. Within U3 lies an RNA polymerase II promoter which, unlike retroviral promoters, sometimes lacks a TATA box. In some *D. melanogaster* elements the sequence TCAGT is located near the initiation site (Arkhipova *et al.*, 1986; Sneddon and Flavell, 1989; Yuki *et al.*, 1986; Matsuo, *et al.*, 1986). This motif is within a region essential for promoter activity in *mdg-1* and *mdg-4* (also known as *gypsy*) of *D.*

Figure 1.4 : **The Structure of Retrovirus-like transposons.**

Figure 1.4

melanogaster (Jarrel and Meselson, 1991; Arkhipova and Ilyin, 1991). The promoters of *mdg-1* and *mdg-4* also require sequences downstream of the transcription initiation site for full activity (Jarrel and Meselson, 1991; Arkhipova and Ilyin, 1991). The LTR (or δ sequence) Ty elements has a TATA box that, in Ty2, has been shown to functions in transcription strength but not start site selection (Liao *et al.*, 1987). Sequences downstream of δ also affect transcription, some of which act through proteins involved in yeast mating type.

20-60bp downstream of the transcription initiation site lies one, or sometimes two, polyadenylation signals, AATAAA. Proximity to the promoter is thought to prevent transcription termination of nascent RNAs at the left hand LTR (Coffin, 1985; Sanfacon and Hohn, 1990). However in *copia* premature transcription termination is prevented because sequences internal to the body of the retrotransposon are required in addition to the poly(A) signal for transcription termination (Kurkulos, *et al.*, 1991).

The central portion of the element encodes one, but more frequently two or three, overlapping ORFs. The first ORF often bears similarity to the nucleic acid binding region of the retroviral *gag* polyprotein and ORF2 encodes domains similar to the retroviral *pol* gene. As well as the reverse transcriptase domain (*rt*), ORF2 has similarity to protease (*pro*), endonuclease (*int*), and RNase H (*rnh*) domains in the order 5' *pro rt int rnh* 3'(Doolittle *et al.*, 1989). However in *copia* and Ty the order is 5' *pro int rt rnh* 3' suggesting they belong to a separate subgroup of retroviral-like transposons.

In retroviruses and Ty large polyprotein fusions are produced from the overlapping ORFs encoded on the same mRNA molecule (Mellor *et al.*, 1985a). Translation of the downstream ORF (ORFB) of Ty1 is achieved by translational frameshifting at a 7base sequence, CUU-AGG-C, located at the overlapping codons in a region conserved between Ty families. This sequence allows frameshifting in the +1 direction (Belcourt and Farabaugh, 1990). In retroviruses an AU rich 7base sequence at the site of the frameshift, and a sequence 3' to this with potential to form a hairpin are sufficient for frameshifting (Jacks *et al.* 1988). It is thought ribosomes stall at the hairpin allowing slippage at the 7base sequence in the -1 frame. The

mechanism adopted in *Drosophila* retrotransposons is unknown but a shift in the -1 frame is required suggesting a strategy similar to retroviruses. Translational frameshifting is very inefficient resulting in production of as little as 5% of ORF2 relative to ORF1 (Jacks and Varmus, 1985). Retroviruses use this inefficiency to ensure higher levels of capsid protein. *copia* has only one ORF. Excess structural protein is made by differential post-transcriptional expression of two mRNAs (2kb and 5kb long) (Brierley and Flavell, 1990). The 2kb mRNA encodes all sequences required for formation of virus-like particle (Yoshioka *et al.*, 1990) and its protein is expressed at least 10 times more efficiently.

1.3.2 Transposition of Retroviral-like Elements:

Mobility of retroviral-like transposons is mediated through an RNA intermediate. This was demonstrated for Ty1 and murine intracisternal A-type particle (IAP) by precise removal of an intron upon transposition (Boeke *et al.*, 1985; Heidmann and Heidmann, 1991).

Because transcription of retroviral-like transposons initiates downstream of the elements 5' end, the mechanism of transposition is not as simple as that proposed for LINEs. Instead, use is made of the elements terminal redundancy to synthesise a full length double stranded DNA replica.

Priming of the first strand cDNA on the retrotransposon transcript is by a tRNA that binds a region between the left hand LTR and ORF1. This so called 'tRNA primer binding site' (PBS) is complementary to the 3' terminal 18 bases of a particular tRNA that differs from one retrotransposon family to another. The PBS of *copia* has complementarity with the central rather than 3' region of $tRNA^{MET}$ (Kikuchi *et al.*, 1986) and the putative PBS of *mdg-3* is complementary to 13 nucleotides of $tRNA^{LEU}$ that begin 5 nucleotides from the 3' end of the tRNA (Saigo, 1986). For *copia* it has been suggested that the tRNA is cleaved internally to produce the 3'OH used in priming and that *Drosophila* RNase P catalyses the cleavage event (Kikuchi *et al.*, 1986; Kikuchi *et al.*, 1990).

Synthesis of the plus strand is primed from a polypurine ribonucleotide

generated by RNase H at site between the ORFs and the right hand LTR (Resnick *et al.*, 1984).

The mechanism of transposition is based mainly on information from retroviral replication. However, some intermediate molecules of transposition corresponding to those predicted by retroviral replication have been identified for retroviral-like transposons. These are 'strong stop minus DNA' i.e. the DNA produced when reverse transcription initiating at the PBS reaches the 5' end of the RNA template, and 'strong stop plus DNA' i.e. the DNA primed from the polypurine ribonucleotide primer that stops at the PBS. Both species have been identified in *Drosophila* tissue culture cells for *mdg-1*, *mdg-3*, and *mdg-4* (Arkhipova *et al.*, 1986). Variation may exist between the mechanisms of retroviruses and retroviral-like transposons as strong stop plus DNA of *mdg-1*, *mdg-3*, and *mgd-4* lacks a detectable RNA primer (Arkhipova *et al.*, 1986). In addition it has been suggested that the plus and/or minus strand in Ty1 is primed from the 2'OH of an unknown RNA as Ty1 transposition is reduced 9 fold in strains that carry a mutated debranching enzyme, *DBR-1*, which cleaves 2'-5' phosphodiester bonds (Chapman and Boeke, 1991).

Further evidence to support a common mechanism of replication is that extrachromasomal circular copies of *copia*, *412*, *297*, *mdg-1*, *mdg-3*, and *mdg-4* some predominantly with one LTR, others with two (Shepard and Finnegan, 1984; Mossie *et al.*, 1985; Ilyin *et al.*, 1984; Flavell and Horowicz, 1981), and linear copies of *copia* have been found in *Drosophila* tissue culture cells (Flavell and Brierley, 1986). Only single LTR circular copies of Ty have been reported to date (Ballario *et al.*, 1983). In retroviral replication such molecules are used as substrates for integration. In addition viral-like particles containing *copia* RNA and reverse transcriptase activity have been found in *Drosophila* tissue culture (Shiba and Saigo, 1983). The causal agent of such particle formation is *copia* DNA (Miyake *et al.*, 1987). Similarly Ty RNA, DNA, and reverse transcriptase activity associated with virus-like particles in *S. cerevisiae* cytoplasm have been described (Mellor *et al.*, 1985b; Eichinger and Boeke, 1988).

There is no disputing the relatedness of retroviruses and retroviral-like

transposable elements. However whether the transposons are vestiges of retroviruses that have lost the extracellular phase, or retroviruses have evolved from cell bound transposable elements by transduction of host membrane proteins, or evolving envelope proteins independently is conjecture.

1.4 OTHER SYSTEMS OF HYBRID DYSGENESIS:

1.4.1 P-M Hybrid dysgenesis:

The P-M system of hybrid dysgenesis in *D. melanogaster*, like I-R, is seen when males of P strains are crossed with females of M strains. However, both male and female progeny display sterility manifested as gonadal dysgenesis (GD: a failure of the gonads to develop). Fertile G_1 flies have abnormalities manifested in their germline as high mutation frequency, chromosome rearrangements, and male recombination (Kidwell *et al.*, 1977; Bregliano and Kidwell, 1983). The reciprocal cross produces fertile flies with a normal level of mutation in the G_2 generation. Crosses between P or M strains are normal. The phenomenologies of P-M and I-R dysgenesis are therefore rather alike but at the molecular level the relatedness breaks down.

1.4.1.1 The P Factor:

The *P* factor is 2.9kb long but non-autonomous elements that have internal deletions can also be found (O'Hare and Rubin, 1983). 30-50 *P* elements are found in typical P strains located at euchromatic sites and are absent from M strains, but M' strains exist that have apparently only non-autonomous *P* elements (Bingham, *et al.*, 1982). The *P* factor has four exons designated ORF0,1,2, and 3, terminates in 31bp inverted repeats, and creates 8bp duplications of host DNA upon insertion (O'Hare and Rubin, 1983). *P* elements can excise precisely or imprecisely, as well as transpose. These features classify *P* elements with transposons thought to move via a DNA rather than an RNA intermediate.

The *P* factor was demonstrated to be the causal agent of P-M hybrid dysgenesis using a sensitive assay for *P* element transposition. This assay makes use

of a P-M induced mutation at the *singed* (*sn*) locus called sn^w (moderately singed bristles)(Engels, 1979). The sn^w allele is caused by two defective *P* elements situated at the locus. These elements are stable in an M cytotype but under dysgenic conditions one or other element excises to form the sn^+ (pseudo wildtype) or sn^e (extreme singed phenotype) (Roiha *et al.*, 1988). Spradling and Rubin (1982) introduced plasmid born *P* factor DNA into embryos of an M cytotype strain that carried the sn^w allele. sn^+ and sn^e individuals were recovered in the progeny of injected embryos indicating the injected DNA contained sequences that control P-M hybrid dysgenesis. Spradling and Rubin also showed that plasmid born *P* factor DNA transposed onto the germline chromosomes of recipient M strain embryos and thus the *P* factor provides transposase for automobilisation. Rubin and Spradling (1982) went on to show that *P* can catalyse mobilisation of defective *P* elements in *trans*. These experiments pioneered the P-element-mediated transformation system of *D. melanogaster*.

1.4.1.2 Regulation of Transposition:

A - Germline Specificity

Germline restriction of *P* element transposition is not controlled at the level of the promoter. This was demonstrated by Laski *et al.* (1986) by transforming flies with a construct in which *P* expression is controlled from the *hsp70* promoter. The *hsp70* promoter functions in both somatic and germline tissues. Somatic excision could be monitored because the *P* element transgene included the *white* gene and the host background was *white*$^-$. Somatic excision would produce white patches on a red eye background. No mosaic eyed flies were found indicating that somatic transcription of *P* is insufficient for transposition. In addition, *P* factor transcripts of 2.5kb and 3kb long are made in somatic tissues of P strains - absence of somatic transposition in P strains implies these RNAs do not produce transposase. Mapping the splice site boundaries of somatic transcripts showed no splice between ORF2 and ORF3 (IVS3) which would result translation termination before ORF3. Karess and Rubin (1984) had already established that ORF3 is essential for transposition by showing that a *P*

element with a frameshift in ORF3 is unable to destabilise the sn^w locus. Laski *et al.*(1986), therefore created a *P* element with IVS3 precisely removed, P[*ry*(Δ2-3)], and tested its ability to stimulate somatic excision and transposition of *P* elements. Both activities were found, therefore transposase production requires splicing of IVS3 which is restricted to the germline. The transposase encoded by the [Δ2-3] element is 87kd and can catalyse excision of *P* element DNA precisely in tissue culture (Rio *et al.*, 1986).

The tissue specific splice is regulated by splicing inhibitor(s), present in somatic tissue, but absent in the germline (Seibel and Rio, 1990; Chain *et al.*, 1991). Seibel and Rio showed that when *P* element premRNA is incubated with human splicing extract IVS3 is removed, but the reaction is inhibited if preincubated with *Drosophila* splicing extract of somatic origin. Several proteins have been identified in the *Drosophila* somatic extract which specifically bind ORF2 premRNA immediately upstream of the IVS3 5' splice site. Mutation of these sequences allows splicing in both germline and somatic tissues *in vivo* (Chain *et al.*, 1991).

B - Repression of Transposition:

P element transposition and excision are repressed in P cytotype (Engels, 1983). Following introduction of a single *P* factor into an M cytotype the cytotype can switch to P indicating the *P* factor is involved in cytotype determination (Daniels *et al.*,1987; Preston and Engels, 1989). By monitoring the *P* factor copy number by *in situ* hybridisation, and cytotype state using s^w destabilisation and gonadal dysgenesis, Preston and Engels (1989) showed the cytotype switch can take up to 30 generations after the *P* element copy number has reached 20. This delay in cytotype switch is not understood.

As well as the transposase, the *P* factor encodes a 66kd protein translated from mRNA retaining IVS3. Misra and Rio (1990) tested its ability to confer P cytotype by transforming M strain flies with elements encoding only the 66kd protein. The 66kd protein can repress transposase activity in both somatic and germline tissues, but the strength of repression depends on the position of integration. No maternal effect

of repression is seen, that is repression of transposition is seen when the 66kd transgene is inherited maternally or paternally. Robertson and Engels (1989) reported similar results with *P* elements harbouring mutations which allow expression of the 66kd protein only. Again these effects were position dependent. The position effect may help explain the delay in onset of P cytotype in that many *P* elements may exist in a strain but P cytotype is not established because their chromosomal location is unsuitable for correct expression.

P cytotype regulates *P* factor expression at the level of the promoter (Lemaitre and Coen, 1991). This was shown by comparing *lacZ* mRNA levels in P and M cytotypes in transgenic flies with the *lacZ* gene expressed from the *P* element promoter. The level of *lacZ* mRNA was much lower in P than M strains. Transcription repression in the germline could be mediated by the 87kd *P* transposase as it specifically binds the *P* element ends and inhibits transcription *in vitro* by competing with TFIID for their overlapping binding sites in the *P* promoter (Kaufman *et al.*, 1989; Kaufman and Rio, 1991). Lemaitre and Coen (1991) showed β-gal expression from the *P* promoter is repressed in all tissues and developmental stages in P cytotype. A likely candidate for the repressor in somatic tissues (and possibly in germline as well) is the 66kd protein. Its DNA binding properties have not yet been shown but its coding sequence overlaps the transposase making it possible that they share sequence specific DNA binding domains.

1.4.1.3 Excision and Transposition:

Unlike retrotransposons, *P* element mobility is mediated via a DNA intermediate and excision as well as transposition occur. *P* transposition could be either replicative, with synthesis and insertion of a DNA copy at a new chromosomal location, or non-replicative in which the element excises, leaving a double-strand gap that must be repaired, and inserts at another chromosomal location. Engels *et al.* (1990) provided evidence in support of non-replicative transposition by showing the DNA at the donor site can become like that of the homologous chromosome following excision. This indicates that the gap created by excision is repaired using the

homologous chromosome as a template by a mechanism similar to gene conversion. This would result in an apparent replicative event if the homologous chromosome is identical to the donor, and a precise excision event if the homologous chromosome lacks a *P* element insert. Gap repair using the homologous chromosome was estimated to occur 15% of the time, the remaining repair events are thought to use the sister chromatid, resulting in events that appear as replicative transposition. This model also explains the frequent occurrence of internally deleted *P* elements - these could arise from arrest of repair synthesis followed by joining of copied *P* element ends.

Details of transposition and excision at the molecular level have yet to be elucidated. However it is known that around 150bp at each end of the *P* element are necessary (O'Hare and Rubin, 1983). Functional analysis of the ends has shown the 31bp inverted repeats, 11bp inverted repeats (approximately 130bp from the ends), and 10bp inverted repeats to which the transposase specifically binds are important for transposition (Mullins *et al.*, 1989; Misra and Rio, 1990). In addition a host protein which binds the 31bp repeats has been identified but its role in transposition, if any, remains unknown (Rio and Rubin, 1988).

1.4.2 H-E hybrid dysgenesis:

The least well understood system of hybrid dysgenesis in *D. melanogaster* is H-E, mediated by the transposable element *hobo*. *Hobo* resembles *P* in structure and in and the dysgenic traits it produces. H-E dysgenesis occurs when strains carrying *hobo* (H) are crossed with strains lacking it (E for empty), inducing hypermutability in the germline and occasional gonadal dysgenesis in G_1 flies (Blackman *et al.*, 1987; Yannopoulos *et al.*, 1987). However unlike *I* or *P*, *hobo* is not influenced by maternal cytotype as an almost equal rate of transposition occurs in HxE versus ExH crosses. In addition, *hobo* induced rearrangements are occasionally found in the somatic tissues of H strains indicating mobility is not restricted to the germline or is exclusively E strain induced (Yannopoulos *et al.*, 1987; Blackman and Gelbart, 1989).

Most H strains have 2-10 putative full length *hobo* elements and numerous internally deleted elements distributed around the chromosome arms (Blackman *et al.*, 1987). Blackman *et al.*, (1989) showed that a full length element can catalyse integration of a marked non-autonomous *hobo* into germline chromosomes of E strain embryos. The element is therefore fully functional and can mobilise a defective element in *trans*. In addition this marked element can transpose onto recipient embryo chromosomes in the absence of the full length element if the embryos are from a HxE cross. The full length *hobo* element has been sequenced (Calvi *et al.*, 1991). It is 3kb long with short inverted repeats of 12bp at each end, and creates target site duplications of 8bp upon insertion. It encodes one short ORF (ORF0) and one long ORF (ORF1), ORF1 being the transposase as mutation of its carboxy terminus disrupts the elements ability to catalyse transposition.

1.4.3 Other Systems of Hybrid Dysgenesis:

A system of hybrid dysgenesis has recently been described in *Drosophila virilis* which resembles P-M dysgenesis in that G_1, males and females display gonadal dysgenesis and G_2 flies have a high frequency of mutation, chromosome non disjunction, and male recombination (Lozovskaya *et al.*, 1990). However unlike P-M or I-R dysgenesis the traits are seen when males from a long established laboratory strains are crossed with females from a wildtype strain. The temperature effects also differ with highest sterility at 23-25°C. The retroviral-like transposon *Ulysses* is implicated as the genetic determinant because it inserted into the *white* locus during a dysgenic cross.

CHAPTER 2

MATERIALS and METHODS

2.1 MEDIA:

2.1.1 Bacterial Media:

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

Luria Agar: As Luria broth with 15g per litre Difco agar.

BBL-top Agar: Baltimore Biological Laboratories trypticase, 10g; NaCl, 5g; Difco Agar, 6.5g; per litre.

Minimal Agar: Difco Bacto Agar, 6g; 5 x Spizizen salts, 80mls; 20% glucose, 4mls; 5mg/ml Vitamin B1, 0.1mls; per 400mls.

2xTY Broth: Difco Bacto tryptone, 16g; Yeast extract, 10g; NaCl, 10g; per litre adjusted to pH 7.4.

2xTY Agar: As 2xTY broth with 15g per litre Difco agar.

Ampicillin to a final concentration of 100μg/ml was added to media immediately prior to use when required.

2.1.2 *Drosophila* Media:

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

Egg collection media: Difco Bacto Agar, 1.5g per 100mls pure apple juice.

2.1.3*Drosophila* Tissue Culture Media:

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

Schneider's Drosophila medium (revised) (Imperial Laboratories) supplemented as above.

2.1.4 Mammalian Tissue Culture Media:

Glasgow Modified Eagles Medium (Flow laboratories) supplemented with 1x non-essential amino acids, 1mM sodium pyruvate, 2mM L-Glutamine, 50μg/ml

streptomycin sulphate, 50u/ml penicillin, and 6% foetal calf serum (Sera-lab).

2.2 MATERIALS

2.2.1 Solutions:

50 x Denharts: Ficol (MW 400,000), 1g; Polyvinyl pyrrolidone (MW 40,000), 1g; BSA, 1g; made up to 100mls with water.

10 x HBS: NaCl, 8.18g; HEPES, 5.94g; Na_2HPO_4, 0.2g; made up to 100 mls with water.

LTB: 20mM Tris/HCl pH 7.5; 10mM $MgCl_2$; 20mM NaCl.

10 x PBS: NaCl, 80g; KCl, 20g; Na_2HPO_4, 15g; KH_2PO_4, 20g; made up to 1 litre with water.

20 x TAE: Tris Base, 96.8g; Glacial acetic acid, 22.8mls; 0.5M EDTA (pH 8.0), 40mls; made up to one litre with water.

20 x TBE: Tris Base, 216g; Boric Acid, 110g; EDTA, 18.6g, made up to one litre with water.

20 x SSC: NaCl, 175g; *tri*-Na Citrate, 88g, made up to one litre with water.

Ringers solution: NaCl, 6.5g; KCl, 0.14g; $NaHCO_3$, 0.2g; $NaH_2PO_4.2H_2O$, 0.01g; $CaCl_2.2H_2O$, 0.12g, made up to one litre with water.

2.2.2 Isotopes:

α - ^{32}P - dCTP (3000 Ci/mM)

α - ^{35}S - dATP (400 Ci/mM)

γ - ^{32}P - dATP (5000 Ci/mM)

D-threo-[dichloroacetyl-1-^{14}C] chloramphenicol (53 or 57mCi/ml)

2.2.3 Suppliers:

Enzymes - Bethesda Research Laboratories, unless otherwise stated.

Isotopes - Amersham International

General laboratory chemicals - Sigma chemical company; Fisons; BDH; Boehringer Mannheim; Merck; Pierce.

2.2.4 Bacterial Stocks:

NAME	GENOTYPE	REFERENCE
RR1	*supE44, hsdS20($r_B^-m_B^-$), ara-14, proA2, lacY1, galK2, rpsL20(Sm^r), xyl-5, mtl-1*	Bolivar *et al.*, 1977
TG1	*supE, hsdΔ5, thi, Δ(lac-proAB), F'[traD36, proAB$^+$, lacI^q, lacZΔM15*	Gibson, 1984

2.2.5 Phages:

NAME	DESCRIPTION and USE	REFERENCE
M13mp18	Sequencing vector based on the *E. coli* phage M13. Used for cloning and sequencing.	Messing, 1983; Norrander *et al.*, 1983
M13mp19	As M13mp18	Messing, 1983; Norrander *et al.*, 1983
λcI857	λ phage with temperature sensitive *cI* gene product. DNA restricted with *Hind*III and used as size markers.	Sussman and Jacob, 1962

2.2.6 Plasmids:

NAME	DESCRIPTION and USE	REFERENCE
pUC18	General cloning vector used to make pCAT.1.	Norrander *et al.*, 1983; Messing, 1983
pBR322	General cloning vector used as size markers when restricted with *Hpa*II.	Bolivar *et al.*, 1977
pC4CAT	CAT reporter plasmid from which CAT gene was cloned to form pCAT.1	Thummel *et al.*, 1988
pSV2CAT	Plasmid containing SV40 early promoter fused to CAT. Used as control in tissue culture transfection.	Gorman *et al.*, 1982
p-194.70Z	Plasmid containing *hsp70* promoter fused to *E. coli lacZ*. Used as internal control in tissue culture transfections.	Xiao and Lis, 1988

pCMVβgal	Plasmid containing CMV immediate early promoter sequences fused to *E. coli lacZ*. Used as an internal control in tissue culture transfection.	Gift from Peter Rigby
pW8	P-element trans-formation vector containing *white* as the selectible marker. Used as vector for *I* factor promoter *CAT* fusions transformed into flies.	Klemenz *et al.*, 1987
pΠ25.7wc	Helper plasmid used in P-element trans-formation.	Karess and Rubin, 1984
phs70Δ2-3wc	Helper plasmid for P-element transformation containing a P-element, lacking the germline specifically spliced intron, transcribed from the *hsp70* promoter	Gift from Jean-Maurice Dura

pI407	Complete *I* factor from w^{IR3} cloned into pAT153. Internal *I* fragment used in Southern hybridisation.	Bucheton *et al.*, 1984
pI954	Complete *I* factor in the P-element transformation vector, pUC-*hsneo* (Steller and Pirrotta, 1985). Used as template in PCR reactions.	Pritchard *et al.*, 1988

2.2.7 Drosophila Stocks:

NAME	GENOTYPE	I-R/P-M STATE	REFERENCE
Cha	+	R/M	Pélisson, 1981
Cha-RC^{+}	+	I/M	Pelisson and Bregliano, 1987
w^{K}	*white*$^{-}$	R/M	Lüning, 1981
p[a-87]	+, carries transgene with LSP-1a promoter fused to CAT	Not tested	Gift from D. Glover
3	*SM6/Cy; Dcxf/SbIn(Mo)*	I/M	Gift from A. Bucheton
M5 B Birm	*In(1)sc*^{SIL} *sc*^{8R}*+S,* *sc*^{SI}*sc*^{8}*, w*^{a}*,B*	I/M	Gift from A. Bucheton
175	*Cy/Pl;Dcxf/H*	R/M	Gift from A. Bucheton

2.2.8 Cell Culture Lines:

NAME	COMMENTS	REFERENCE
Schneider Line 2	Derived from Oregan-R late embryos, 60-80% tetraploid and reactive.	Schneider, 1972
RJK88	Chinese hamster lung fibroblasts, *HPRT*$^{-}$.	Fuscoe *et al.*, 1983

2.2.9 Oligonucleotides:

NAME	SEQUENCE 5'-3'	COMMENTS
664C	GTC TAG ACA TTA CCA CTT CAA CCT CCG	'+' strand PCR oligonucleotide from base +1 to +20 on *I* factor sequence with +3=T.
134D	GTC TAG ACC GAA GAG ATA AGT CGT GCC	'+'strand PCR oligonucleotide from base +18 to +37 on *I* factor sequence.
133D	GTC TAG ACT CAG TCT AAA GCC TCG CTT	'+' strand PCR oligonucleotide from base +39 to +58 on *I* factor sequence.
135D	GTC TAG AAA TAT CAA CCA CAA AGA GAA	'+' strand PCR oligonucleotide from base +101 to +120 on *I* factor sequence.
737F	GTC TAG ACA GTA CCA CTT CAA C	'+' strand PCR oligonucleotide from base +1 to +15 on *I* factor sequence with base +3=G.
468E	GTC TAG AAC CAC TTC AAC CTC CGA AGA	'+' strand PCR oligonucleotide to change bases +1 and +4 of the *I* factor from C and T to T and A respectively.
442F	GTC TAG ACC GAC CAC TTC AAC CTC CGA CGC	'+' strand PCR oligonucleotide to change bases +1 to +4 of the *I* factor from CAGT to ACCG.
188F	GTC TAG AGT CGA CTT CGG GCC TCC CTC	'+' strand PCR oligonucleotide to include 99 bases of *white* upstream of the *I* factor in pI954.

403D	CGG ATC CGA TTG TTG GTT AAG GGC TTG	'-' strand PCR oligonucleotide from bases +186 to +167 of the *I* factor.
883D	CGG ATC CTG TTT ATC AAG ATT TTG CTG	'-' strand PCR oligonucleotide from bases +100 to +81 of the *I* factor.
884D	CGG ATC CCT TTA GAC TGA GAG GCA CGA	'-' strand PCR oligonucleotide from bases +50 to +31 of the *I* factor.
490E	CGG ATC CAG AGG CAC GAC TTA TCT CTT	'-' strand PCR oligonucleotide from bases +40 to +21 of the *I* factor.
510E	CGG ATC CTA TCT CTT CGG AGG TTG AAG	'-' strand PCR oligonucleotide from bases +28 to +9 of the *I* factor.
472K	CTA GAC AGT ACC ACT TCG	'+' strand oligonucleotide from bases +1 to +12 of the *I* factor, with *Xba*I sticky end sequence at 5' end.
471K	GAT CCG AAG TGG TAC TGT	'-' strand oligonucleotide from bases +12 to +1 of the *I* factor with *Bam*HI sticky end sequence at 5' end.
508E	CGG ATC CCA TTA CCA CTT CAA CCT CCG	'-' strand PCR oligonucleotide for reverse promoter with T at base +3.
509E	GTC TAG ACT TTA GAC TGA GAG GCA CGA	'+' strand PCR oligonucleotide for reverse promoter containing bases +1 to +50 of the *I* factor.
122H	CGG ATC CCA GTA CCA CTT CAA CCT CCG	'-' strand PCR oligonucleotide for reverse promoter with G at base +3.

123H	GTC TAG AGA TTG TTG GTT AAG GGC TTG	'+' strand PCR oligonucleotide for reverse promoter containing bases +1 to +186 of the *I* factor.
551D	TAG ACT GAG ACC TCG AGG TTA TCT CTT CG	SDM oligonucleotide to substitute 8 out of 9 conserved bases from +30 to +37 of the *I* factor.
467E	GAG GCA CGA CNT ATC TCT TCG N=A,G or C	SDM oligonucleotide to substitute A at +29 of the *I* factor to T,C or G.
466E	AGA GGC ACG ANT TAT CTC TTC N=G,A or T	SDM oligonucleotide to substitute G at +30 of the *I* factor to A,T or C.
465E	GAG AGG CAC GNC TTA TCT CTT N=T,G or C	SDM oligonucleotide to substitute T at +31 of the *I* factor to A,G or C.
462E	TGA GAG GCA CNA CTT ATC TCT N=A,T or C	SDM oligonucleotide to substitute C at +32 of the *I* factor to A,T or G.
463E	CTG AGA GGC ANG ACT TAT CTC N=A,T or G	SDM oligonucleotide to substitute G at +33 of the *I* factor to A,T or C.
464E	ACT GAG AGG CNC GAC TTA TCT N=T,C or G	SDM oligonucleotide to substitute T at +34 of the *I* factor to A,C or G.
663C	GTA CTG CAGTAC AGC T	Self complementary oligonucleotide with *Sac*I sticky ends and *Pst*I site. Used to insert a *Pst*I site into pCAT.0.
M13 Sequencing Primer	GTA AAA CGA CGG CCA GT	-20 M13 17mer sequencing primer.
M13 Sequencing Primer	GTT TTC CCA GTC ACG AC	-40 M13 17mer sequencing primer.

332G	CAA CGG TGG TAT ATC CAG TGA TTT T	Oligonucleotide complementary to CAT sequences, used as primer in plasmid sequencing and primer extension.

I factor sequences are shown in bold type; restriction endonuclease recognition sites are showm underlined - *Xba*I=5'TCTAGA3'; *Bam*HI=5'GGATCC3'; *Xho*I=5'CTCAGA3'; *Pst*I=5'CTGCAG3'; *Sac*I=5'GAGCTC3'.

2.3 METHODS

2.3.1 MANIPULATIONS OF BACTERIA AND PHAGE

2.3.1.1 Growth of *E. coli* Bacterial Cultures:

Liquid culture of *E. coli* was either in Luria broth (RR1) or 2xTY broth (TG1) by inoculating a given volume of broth with a single colony using a sterile inoculating loop. Cultures of greater than 10mls were grown in conical flasks with a total capacity of 5-10x that of the culture volume. Cultures of less than 10mls were grown in 2oz glass bottles. Cells were shaken at 37°C for an appropriate length of time.

2.3.1.2 Storage of *E. coli* Bacterial Cultures:

For long term storage 1ml of fresh overnight culture of bacteria, grown in Luria broth supplemented with antibiotic if necessary, was mixed with 1ml of 100% glycerol (autoclaved), and stored in a sterile vial at -70°C. Upon recovery, vials were thawed quickly at 37°C and the culture streaked out on Luria broth agar plates, with antibiotic if required. TG1 was streaked onto minimal agar to maintain the F' plasmid. After overnight incubation at 37°C, a single colony was picked to propagate a fresh bacterial culture.

For short-term storage (4-6 weeks), bacteria were stored as 5ml overnight liquid cultures or streaked onto agar plates.

2.3.1.3 Small Scale Preparation of Plasmid DNA: (Birnboim and Doly, 1979)

3mls of Luria broth, supplemented with ampicillin to a final concentration of 100μg/ml, was inoculated with a single colony of plasmid carrying strain and incubated at 37°C overnight with aeration. 1.5mls of the culture was pelleted by centrifugation in an Eppendorf microfuge and resuspended in 100μl 25mM Tris.HCl pH8, 10mM EDTA, 50mM glucose. 200μl of lysis buffer (0.2M NaOH, 1% SDS) was added and after gentle mixing, placed on ice for 20 minutes. Following lysis 150μl of 3M Na Acetate pH5 was added to precipitate chromosomal DNA, SDS and

proteins. After further incubation on ice for 20 minutes the precipitate was centrifuged in a microfuge for 10 minutes. To the supernatant, 1ml of absolute ethanol was added, mixed then immediately centrifuged in a microfuge to pellet the plasmid DNA. The DNA was washed with 70% ethanol, dried under vacuum, and resuspended in 50μl of 10μg/ml RNase made up in TE buffer. Plasmid 'miniprep' DNA was stored at 4°C.

2.3.1.4 Large Scale Plasmid Preparation:

A 250ml culture of bacteria carrying the desired plasmid was incubated overnight at 37°C with vigorous shaking in Luria broth supplemented with 100μg/ml of ampicillin. The cells were pelleted by centrifugation at 5000rpm (in Sorval GSA rotor) and resuspended in 4mls of 25mM Tris.HCl pH8, 10mM EDTA, 50mM glucose. Cells were lysed on ice by adding 8mls of 0.2M NaOH, 1% SDS and left on ice for 20 minutes. Addition of 6mls of 3M Na acetate pH5 precipitated chromosomal DNA, SDS and proteins which were spun down at 15000rpm (Sorval SS-34 rotor) for 15 minutes at 4°C. To the supernatant 11mls of isopropanol were added and left at room temperature for 30 minutes to precipitate plasmid DNA. The DNA was pelleted by centrifugation at 10000rpm (Sorval SS-34 rotor) for 20 minutes at 4°C. The pellet was washed with 70% ethanol, dried under vacuum, resuspended in 10mls of TE buffer plus 10μg/ml RNase and incubated at 37°C for 30 minutes. 9.4mls of plasmid DNA was transferred to a fresh tube to which 100μl of 10mg/ml ethidium bromide and 9.02g of CsCl were added, giving a density of 1.55g/ml. The DNA was banded by centrifugation at 38000rpm (in a Beckman Ti50 rotor) for 40-48 hours at 20°C. DNA was visualized by side illumination with UV light. The lower band containing supercoiled plasmid DNA was removed by puncturing the tube with a 19 gauge needle and syringe. A second 19 gauge needle was inserted at the top of the tube to allow pressure release. The ethidium bromide was removed by extraction several times with *sec*-butanol, and the CsCl was removed by dialysis against 2 litres of TE buffer for 8 to 15 hours at room temperature. The TE buffer was changed 3 to 4 times during dialysis. DNA to be used in tissue culture transfection experiments

was extracted with phenol:chloroform:IAA (see section 2.3.2.1) and stored at 4°C. The remainder was kept at 4°C for short to medium-term storage (up to 12 months), or under ethanol at 70°C for long-term storage.

2.3.1.5 Small Scale M13 Template DNA Preparation:

2mls of 2xTY broth was inoculated with 20μl of fresh TG1 overnight culture and a single recombinant M13 plaque, and incubated at 37°C with aeration for 4.5 hours. 1.5mls of culture was transferred to an Eppendorf tube and centrifuged for 5 minutes to pellet the cells. 1ml of the supernatant was transferred to a fresh tube and 200μl of 20% PEG 6000 in 2.5M NaCl added and left to stand at room temperature for 30 minutes to precipitate the phage. Phage were pelleted by centrifugation in a microfuge for 15 minutes and the supernatant removed. Centrifugation for 1 minute was repeated to allow removal of all residual PEG solution. The phage pellet was resuspended in 100μl of TE buffer and protein removed by extraction with 70μl of phenol. DNA was recovered by ethanol precipitation. Following a 70% ethanol wash the DNA was dried under vacuum, resuspended in 50μl of TE buffer, and stored at -20°C.

2.3.1.6 Large Scale M13 Template DNA Preparation:

To 20mls of 2xTY broth, one drop of fresh TG1 overnight culture was added and shaken at 37°C for 3 hours. 1ml of 2xTY medium was then inoculated with 100μl of the 3 hour culture and a recombinant M13 plaque. The culture was shaken at 37°C for 4 hours then transferred to an Eppendorf tube and centrifuged for 5 minutes in a microfuge. The supernatant was transferred to a new tube and used to inoculate a 100mls exponential culture of TG1 (OD_{550nm} of 0.3) in 2xTY broth. The culture was shaken at 37°C for 4 hours and the cells pelleted at 5000rpm (Sorval GSA rotor). The supernatant was transferred to new tubes and 0.2x volumes of 20% PEG 6000 in 2.5M NaCl added and phage allowed to precipitate at 4°C for 1 hour. The phage were collected by centrifugation at 7000rpm (Sorval GSA rotor) for 20 minutes. The supernatant was discarded and residual PEG removed following further centrifugation

for 5 minutes. The phage pellet was resuspended in 500μl of TE buffer and transferred to an Eppendorf tube. Centrifugation for 5 minutes in a microfuge ensured removal of any remaining cells. The supernatant was transferred to a fresh tube and 200μl of 20% PEG 6000 in 2.5M NaCl added, and left at room temperature for 15 minutes. Phage were pelleted by centrifugation for 5 minutes, the supernatant discarded, and all residual PEG removed after further centrifugation. The pellet was resuspended in 500μl TE buffer and protein removed by extraction with 200μl of phenol twice. The upper aqueous phase was then extracted with 500μl diethyl ether four times, followed by chloroform extraction twice. DNA was ethanol precipitated from the aqueous phase and after washing in 70% ethanol, and drying, the pellet was resuspended in 100μl of TE buffer and stored at -20°C.

2.3.1.7 Large Scale Preparation of M13 RF DNA:

To 250mls of 2xTY broth, 1ml of fresh TG1 overnight culture and 100μl of M13 lysate were added and shaken overnight at 37°C. The culture was then treated as for large scale plasmid DNA preparation.

2.3.1.8 Competent Cells:

25mls of broth (LB for RR1 and 2xTY for TG1) were inoculated with 0.5mls of an overnight culture and shaken at 37°C until an OD_{650nm} of 0.4-0.6 was reached. Cells were placed on ice for 10 minutes to arrest growth, then pelleted by centrifugation at 5000rpm for 5 minutes at 4°C. Cells were resuspended in 2.5mls of cold TSB (10% PEG 3000; 5% DMSO; 10mM $MgCl_2$; 10mM $MgSO_4$, made up in Luria broth for RR1 or 2xTY for TG1). After 30 minutes on ice cells were competent and remained competent for a further 24 hours if kept on ice. Alternatively cells could be snap frozen in liquid nitrogen when competent, and stored indefinitely at -70°C.

2.3.1.9 Transformation of Plasmid DNA into Competent Cells:

5 to 10ng of supercoiled plasmid DNA, or half of a ligation reaction, were

added to 100μl of competent RR1 cells in an Eppendorf tube and left on ice for 30 to 60 minutes. 0.9mls of cold TSB and 20μl of 20% glucose were added to the tube and shaken at 37°C for one hour. 10μl, 100μl, and 400μl alliquotes were spread onto Luria agar petri dishes with 100μg/ml ampicillin and incubated overnight at 37°C.

2.3.1.10 Transfection of M13 RF DNA into Competent Cells:

5 to 10ng of M13 RF DNA, or half of a ligation reaction, were added to 100μl of competent TG1 cells in a 5ml glass test tube and incubated on ice for 30 to 60 minutes. The tube was then transferred to a dry heating block preset to 45°C, and allowed to warm up for 3 minutes. 3mls of molten BBL top agar (at 45°C), 50μl of IPTG (25mg/ml) and 50μl of X-gal (20mg/ml in dimethyl formamide) were added and the mixture poured onto a minimal agar dish. The BBL top agar was allowed to set at room temperature then plates were incubated at 37°C overnight.

2.3.2 NUCLEIC ACID MANIPULATION AND DETECTION METHODS

2.3.2.1 Extraction of Proteins from Nucleic Acid by Phenol and Chloroform:

Distilled phenol was equilibrated to pH8 for extraction of proteins from DNA, by mixing the phenol with an equal volume of 1M Tris.HCl pH8. The phases were allowed to separate and the upper aqueous phase was discarded. This process was repeated until the pH of the phenol reached 8. The Tris layer was then replaced with TE buffer. Extraction of proteins from RNA required phenol to remain acidic (pH5-6). Phenol was therefore equilibrated as above except Tris was replaced by distilled water. The nucleic acid to be extracted was mixed thoroughly with 0.4 to 1x the volume of phenol then spun for 2-5 minutes. The upper aqueous phase was removed into a fresh tube, taking care not to collect any proteins located at the interface, and re-extracted with phenol if necessary. Instead of, or following phenol extraction, DNA or RNA was extracted with phenol:chloroform:IAA (25:24:1). The aqueous phase was then extracted with an equal volume of chloroform:IAA (24:1) to remove any

remaining phenol. After spinning the upper aqueous phase was transferred to a fresh tube.

2.3.2.2 Precipitation of Nucleic Acid with Ethanol:

DNA was precipitated from solution by addition of 0.1x volumes of 3M Na Acetate, pH4.8-5.2 and 2 to 3 volumes of absolute ethanol at room temperature. The solution was mixed thoroughly and allowed to stand at room temperature for 10 minutes, -20°C overnight, or -70°C for 15 minutes. DNA was pelleted by centrifugation for 15 to 30 minutes. The supernatant was discarded and the pellet washed with 70% ethanol and spun again for 2 minutes. The supernatant was discarded and the pellet dried under vacuum until no visible traces of ethanol remained (2 to 15 minutes). The DNA was dissolved in sterile distilled water or TE buffer. Precipitation of very small quantities of DNA was assisted by addition of tRNA carrier (approximately 10μg). Precipitation of RNA from solution with ethanol was identical except 0.033x volumes of 3M K Acetate, pH5 and 0.5x volumes of ethanol. Alternatively, RNA was precipitated by adding 2x volumes of 3M LiCl and left overnight at 4°C before centrifugation, washing, and drying.

2.3.2.3 Agarose Gel Electrophoresis:

DNA was separated in 0.7-1.2% (w/v) BRL electrophoresis grade agarose with 0.5μg/ml ethidium bromide in 1xTBE buffer (for Southern blotting) or 1xTAE buffer (for routine gels). Prior to loading, DNA samples were mixed with 0.1x volume of sample buffer (20% glycerol; 100mM EDTA; 0.1% bromophenol blue). Electrophoresis was carried horizontally out across a potential difference of 1-10v/cm. Bacteriophage λcI857 DNA cut with *Hind*III and pBR322 cut with *Hpa*II were used as size markers. DNA was visualised by UV illumination and photographed.

2.3.2.4 Polyacrylamide Gel Electrophoresis:

To separate DNA fragments less than 400 base pairs long, electrophoresis was done vertically in a 6-10% polyacrylamide (40:1 acrylamide:bis-acrylamide) in 1xTBE

at 10-70v/cm. When necessary pBR322 restricted with *Hpa*II was used as size markers and the gel stained with ethidium bromide to allow the DNA fragments to be visualised by UV illumination.

2.3.2.5 Cleavage of DNA with Restriction Enzymes:

All DNA restrictions were performed using BRL enzymes and buffers. 0.1-20μg of DNA were cut in 30-200μl of 1x appropriate 'react buffer' for 1-2 hours at 37°C. For double digests involving enzymes with different recommended buffers, the buffers were checked individually in double digests to determine which gave most efficient digestion. All double digests necessary for the results in this thesis worked efficiently in either 'react 2' (50mM Tris.HCl pH8; 10mM $MgCl_2$; 50mM NaCl) or 'react 3' (50mM Tris.HCl pH8; 10mM $MgCl_2$; 100mM NaCl).

2.3.2.6 Recovery of DNA from Agarose Gels:

DNA was electrophoresed through 1% BRL agarose in 1xTAE, 0.5μg/ml ethidium bromide. The desired fragment was visualised by UV illumination, cut out, and extracted from the agarose using 'Geneclean™' (Bio101). The agarose was weighed and 2-3x volumes of 6M NaI added. The agarose was dissolved by heating to 55°C for five minutes with occasional mixing, and then cooled on ice for five minutes. 5μl of 'glassmilk' (a silica matrix suspended in water) was added, and left for five minutes on ice with occasional mixing to allow the DNA to bind to the silica matrix. The 'glassmilk' was pelleted by centrifugation in a microfuge, the supernatant discarded and the pellet washed three times with 500μl of NEW wash (NaCl/ethanol/water mix). After a final spin all the NEW wash was discarded and the DNA eluted from the 'glassmilk' in 5-10μl of TE buffer at 55°C. The mixture was spun in a microfuge, and the supernatant, containing the DNA, transferred to a fresh Eppendorf and stored at -20°C. Fragments less than 300bp long were electrophoresed through 2% ultrapure agarose ('Mermaid™ Biogel', Bio101) in 1xTAE, 0.5μg/ml ethidium bromide. Agarose containing the DNA fragment was cut out, weighed, and mixed with 3x volumes of high salt binding solution (concentrated sodium perchlorate)

in an Eppendorf tube. 8μl of 'glassfog' (a silica based matrix in water) was added, the agarose melted and DNA bound to the 'glassfog' by incubation at 55°C for five minutes. Adsorption was allowed to continue at room temperature for five minute, with occasional mixing to keep the 'glassfog' in suspension. The 'glassfog' was centrifuged, the supernatant discarded, and the pellet washed three times with 300μl of ethanol wash. After the final wash the tube was spun again to ensure removal of all residual ethanol. The pellet of 'glassfog' was resuspended in 10μl of water, and the DNA eluted by incubation at room temperature for five minutes. The 'glassfog' was centrifuged and the supernatant containing the DNA transferred to a fresh Eppendorf tube. Recovery of DNA using Geneclean and Mermaid kits was usually around 80%.

2.3.2.7 Ligation of Fragments Greater than 30bp to Vector DNA:

50-100ng of vector (cut with the appropriate restriction enzyme) were incubated with a four molar excess of insert fragment in 1x ligation buffer (10mM Tris.HCl pH7.2; 1mM EDTA; 10mM $MgCl_2$; 10mM DTT; 1mM ATP) with 10 units of T4 DNA ligase (New England Biolabs). The reaction was made up to a final volume of 10μl and left overnight at 14°C.

2.3.2.8 Ligation of Double Stranded Oligonucleotides to Vector DNA: (Lathe *et al.*, 1984)

Complementary oligonucleotides (not phosphorylated at their 5' end) with appropriate overhanging ends were annealed at a concentration of 1μg/μl in TE buffer by heating briefly to 80°C then cooling from 65°C to less than 10°C over at least a one hour period. 1μg of double stranded oligonucleotide was mixed with 5μg of vector (restricted with enzyme yielding overhanging ends complementary to those of the oligonucleotide) in 1x oligonucleotide ligation buffer (30mM NaCl; 30mM Tris.HCl pH7.5; 7.5mM $MgCl_2$; 2mM DTT; 0.25mM ATP; 0.2mM EDTA; 1mM Spermidine; 0.1mg/ml BSA) in a total volume of 20μl. To this 5 units of T4 DNA ligase was added and incubated overnight at 10°C. Excess oligonucleotides were removed by

addition of 360μl of TE buffer, 40μl of ultrapure DMSO (stored frozen) and 10μl of 100mM spermine (adjusted to pH 6.8 with NaOH). The mixture was vortexed well and frozen briefly in a dry ice/ethanol bath and placed on ice to thaw over a 30-60 minute period. The plasmid DNA was pelleted in a microfuge at 4°C. The supernatant containing the excess oligonucleotides was discarded and the pellet washed once in 75% ethanol; 300mM Na OAc, pH5; 100mM $MgCl_2$, and twice in ethanol:TE buffer (3:1). Following vacuum drying, the DNA was resuspended in 20μl of TE buffer and stored at -20°C until transformation. If the oligonucleotide was designed to abolish a restriction site, digestion with the appropriate enzyme was done prior to transformation.

2.3.2.9 Random-Primed Labelling of DNA: (Feinberg and Vogelstein, 1983 and 1984)

50-100ng of linear DNA in a total volume of 35μl was heat denatured by boiling. After snap cooling on ice the DNA was mixed with 30μCi α^{32}P-dCTP (3000mCi/mmol), 10μg BSA, 10 μl 5xOLB (0.05M β-mercaptoethanol; 1M HEPES pH6.6; 0.03Upd$(N)_6$; 0.26M Tris.HCl pH8; 0.26M $MgCl_2$; 0.1 mMdATP, dTTP and dGTP) and 1μl *E. coli* DNA polymerase I Klenow fragment (5u/μl). The reaction took place at room temperature over 5-12 hour period. The radiolabelled DNA was separated from unincorporated nucleotides by firstly making the total volume of the reaction to 200μl with 4xSSC (final concentration of 3xSSC), then passing through a 1ml Sephadex G-50 column (made up in 3xSSC), (modified from Maniatis *et al.*, 1982).

2.3.2.10 Preparation of Digoxygenin Labelled DNA for *in situ* Hybridisation:

1μg of linear DNA (preferably less than 1kb long) was purified from an agarose gel and denatured by heating to 95°C for ten minutes, followed by quick cooling in an ice/NaCl bath. To this 4μl of OLB (see section 2.3.2.9), and 2μl of 10x dNTP labelling mix (1mM dATP; 1mM dCTP; 1mM dTTP; 0.35mM Dig-dUTP, pH6.5, from Boehringer Mannheim) were added and made up to 19μl with water. 1μl

of *E. coli* DNA polymerase I (Klenow fragment, 5u/μl) was added, mixed, and left overnight at 37°C. The reaction was stopped by addition of 1μl of 0.5M EDTA, pH8. The DNA was precipitated at -70°C for one hour with 1μl of 10mg/ml salmon sperm DNA, 20μl of 4M NH_4Acetate, pH6.6, and 80μl of ethanol. After centrifugation and washing in 70% ethanol, the DNA was dried and dissolved in 20μl of TE buffer.

2.3.2.11 Identification of Bacteria Harbouring Recombinant Plasmids by Colony Hybridisation: (Benton and Davis, 1977)

Colonies were transferred to 9cm nitrocellulose discs by placing filters for two minutes onto the agar. Cells were lysed and DNA denatured by soaking the discs in denaturation buffer (0.5M NaOH; 1.5M NaCl) for five minutes. The discs were then neutralised on 0.5M Tris/1.5M NaCl, pH7 for five minutes with one change of neutralisation buffer, and placed in 6xSSC for two minutes. After drying, the filters were baked *in vacuo* for two hours at 80°C. Prehybridisation in 50% formamide, 4xSSC, 5xDenharts was carried out at 37°C for 0.5-2 hours. Denatured sonicated salmon sperm DNA to a final concentration of 100μg/ml and denatured radioactively labelled probe (approximately 10^6cpm/ml) were added to the prehybridisation mix and incubated overnight at 37°C with gentle agitation. Filters were removed from the hybridisation bag and washed in 2xSSC twice at room temperature for five minutes, then 2xSSC, 0.1% SDS at 37°C for 20 minutes, and finally in 0.2xSSC, 0.1% SDS for 20 minutes at 37°C. Filters were sealed in a plastic bag when damp and autoradiographed for 2-24 hours at -70°C.

2.3.2.11 Polymerase Chain Reaction:

All reactions were carried out using a Techne Programmable Dryblock. The amplification strategy was based on that of Scherf, *et al.*, 1986. Primers were designed such that they share 20 nucleotides homology to the template and a 6 base restriction enzyme site preceded by a G at their 5' end. 100pg of template DNA (pI954, Pritchard, *et al.*, 1988) was mixed with 5μl of 10x amplification buffer (100mM Tris.HCl pH8.3; 500mM KCl; 15mM $MgCl_2$; 0.1% gelatin(w/v); 0.1% tween-

20(v/v); 0.1% NP40(v/v)), 2μl of 10mM dNTPs, 1-2μg of each primer, and 2.5 units of Taq polymerase (Boehringer Mannheim), were made up to a volume of 50μl with water and submerged under 50μl of mineral oil (Sigma). Three initial long denaturation cycles of 93°C for 5 minutes, annealing at 40°C for 2 minutes and polymerisation at 70°C for one minute were done. A further 27 cycles of 93°C for 1.5 minutes, 40°C for 1.5 minutes and 70°C for one minute completed the amplification process. PCR products were checked by polyacrylamide gel electrophoresis and excess primers removed by isopropanol precipitation (one volume of isopropanol and 0.1 volumes of Na Acetate, pH4.8 added mixed and centrifuged in a microfuge for 10 minutes - the DNA pellet was washed and dried as in ethanol precipitation). When the PCR product was less than 50 nucleotides long no isopropanol precipitation was done as the product yield was very low (<10%).

2.3.2.13 DNA Sequencing using the Dideoxynucleotide Method:

Sequencing of DNA was carried out using the Sequenase™ Version 2.0 kit (United States Biochemicals).

a) Sequencing of single stranded M13 template DNA: 1μl of appropriate sequencing primer (3ng/μl) was annealed to 7μl of template DNA (approximately 1μg DNA) with 2μl of 5x reaction buffer (200mM Tris.HCl pH7.5; 100mM $MgCl_2$; 250mM NaCl) by heating to 65°C for two minutes then cooled slowly to below 30°C. Extension from the annealed primer was done by adding 1μl 0.1M DTT, 2μl dGTP label mix (a 1 in 4 dilution of 7.5μM dGTP, dCTP and dTTP), 0.5μl α-^{35}S dATP (400Ci/mmole) and 2μl dilute Sequenase™ (a 1 in 8 dilution of Sequenase at 13 u/μl in 10mM Tris.HCl pH7.5; 5mM DTT; 0.5mg/ml BSA). The extension mix was left at room temperature for 2-5 minutes. Further extension and termination was done by dispensing 3.5μl of extension mix into four tubes preheated to 37°C containing 2.5μl of one of the four termination mixes:

ddGTP mix - 80μM dNTPs; 8μM ddGTP; 50mM NaCl
ddATP mix - " " " ddATP; " "
ddCTP mix - " " " ddCTP; " "

ddTTP mix - " " " ddTTP; " "

The termination reaction was allowed to proceed at 37°C for 5-10 minutes and the reaction stopped by addition of 4μl of stop solution (95% formamide; 20mM EDTA; 0.05% bromophenol blue). Extension products were separated by electrophoresis through a 6% denaturing polyacrylamide gel made up by mixing 21.25g urea, 2mls 20x TBE, 7.5mls 40% acrylamide (40:1 acrylamide:*bis*-acrylamide) made up to 40mls with water. Polymerisation of the acrylamide was achieved by adding 300μl of 10% ammonium persulphate and 80μl of TEMED immediately prior to pouring the gel. Sequencing reactions were heat denatured by boiling for 3 minutes, loaded onto the gel and electrophoresed in 1xTBE buffer at 40v/cm for 2-3 hours. The gel was fixed in 10% methanol (v/v), 10% acetic acid (v/v) for 15 minutes, then dried under vacuum for 30-60 minutes at 80°C and autoradiographed at room temperature overnight.

2.3.2.14 Sequencing of Double Stranded Plasmid DNA by the Dideoxynucleotide Method:

5μg of plasmid DNA (prepared by CsCl density gradient centrifugation) was denatured in 0.2M NaOH, 0.2mM EDTA at 37°C for 30 minutes in a total volume of 10-15μl. The DNA was then placed on ice and neutralised by adding 0.1x volumes of 3M Na Acetate pH4.8. The DNA was ethanol precipitated at -70°C for 15 minutes, pelleted by centrifugation at 4°C, washed with 70% ethanol and dried under vacuum. The DNA was dissolved in 7μl water, 2μl 5x sequenase reaction buffer and 1μl primer (20ng/μl). The primer routinely used was a 25mer internal to the CAT gene (oligonucleotide 332G). The annealing, extension termination and electrophoresis conditions were as described for M13 single stranded DNA sequencing.

2.3.2.15 Genomic DNA Detection by Southern Blotting and Hybridisation:

(adapted from Southern, 1975)

1-5μg of genomic DNA was digested with the appropriate restriction enzyme in a total volume of 50μl for 2 hours at 37°C and separated according to size on a 1% agarose gel (w/v) in 1xTBE, 0.5μg/ml ethidium bromide. Routinely genomic DNA

was electrophoresed for 12-15 hours at 1v/cm then photographed. The gel was then soaked in denaturation buffer (0.5M NaOH, 1.5M NaCl) with gentle agitation for 30 minutes. The DNA was transferred to Genescreen Plus™ nylon membrane (Dupont) by capillary action using denaturation buffer as the transfer medium (Reed and Mann, 1985). Transfer was allowed to continue for 12-48 hours after which the membrane was neutralised for 30 minutes in 3M NaCl, 0.5M Tris.HCl pH7. Prehybridisation of the membrane was done in 30-50mls of 6xSSC, 1% SDS, 10% dextran sulphate and 100μg/ml of denatured salmon sperm DNA for 0.5-2 hours at 65°C. Denatured salmon sperm DNA to bring the final concentration to 250μg/ml and radiolabelled DNA probe (approximately 10^6cpm/ml of buffer) were added to the prehybridisation bag and hybridised at 65°C for 12-16 hours. The filter was washed twice in 2xSSC at room temperature for 5 minutes, then twice in 2xSSC, 0.1% SDS at 65°C for 30 minutes, then twice in 0.1xSSC at room temperature for 30 minutes. The filter was sealed in a plastic bag while damp and autoradiographed at -70°C.

2.3.2.16 Site Directed Mutagenesis: (Zoller and Smith, 1984)

1μg (approximately 160pmoles) of mutant oligonucleotide was phosphorylated at its 5' end in 1x T4 polynucleotide kinase buffer (100mM Tris.HCl pH8; 10mM $MgCl_2$; 5mM DTT; 1mM ATP) with 5u of polynucleotide kinase made up to 10μl with water. The reaction proceeded for 90 minutes at 37°C, and was stopped by heat inactivation of the enzyme at 65°C for 10 minutes. The volume was increased to 50μl with water and extracted with phenol:chloroform:IAA (25:24:1) and DNA precipitated from the aqueous phase with ethanol. To the DNA pellet, 3μg of M13 template DNA (prepared by the large scale method), 2μl of TM(100mM Tris.HCl pH8.5, 50mM $MgCl_2$), 3μl of non phosphorylated -20 M13 universal primer (3ng/μl) and 12μl of water were added. The mixture was heated to 80°C for 5 minutes and allowed to cool slowly to below 30°C to allow annealing of both oligonucleotides to the template DNA. Synthesis of heteroduplex mutant DNA was achieved by addition of 20μl chase (0.25mM dNTPs), 1μl 50mM dATP, 1.5μl Klenow(5u/μl), 1μl T4 ligase (5u/μl) and 6.5μl water. Extension and ligation occurred at room temperature over 90 minutes.

The reaction was then extracted with phenol:chloroform:IAA (25:24:1) and precipitated with ethanol. The DNA pellet was digested in the appropriate 'react buffer' with the appropriate enzymes (*Xba*I and *Bam*HI for all manipulations in this thesis) in a total volume of 40μl, and the heteroduplex DNA recovered after electrophoresis using Geneclean™ (see 2.3.2.6). The purified fragment was ligated to mp18 RF DNA and transfected into competent TG1 cells.

2.3.2.17 Radiolabelling of Oligonucleotides at their 5' End:

To 2-20pmoles of oligonucleotide, 3μl of 10x T4 polynucleotide kinase buffer (0.5M Tris.HCl pH8; 100mM $MgCl_2$; 100mM DTT), 2μl $\gamma[^{32}P]$ ATP(3000Ci/mmol) and 0.5μl (10u/μl) of T4 polynucleotide kinase were added and made up to 30μl with water. The reaction was carried out at 37°C for 45 minutes followed by heat inactivation of the enzyme at 65°C for 10 minutes.

2.3.2.18 *Drosophila* DNA Preparation: (from Pritchard *et al.*, 1988)

10-20 flies were homogenised in an Eppendorf tube on ice in 0.4mls of homogenisation buffer (8M urea; 0.35M NaCl; 10mM Tris.HCl pH8.3; 10mM EDTA) using an motor homogeniser with disposable pestles (Kontes). 0.4mls of phenol:chloroform:IAA (25:24:1) were added and spun on a rotating wheel for five minutes. Cell debris was removed by centrifugation in a microfuge for five minutes and the aqueous phase extracted with phenol:chloroform:IAA a second time. Following a third extraction with only chloroform:IAA (0.4mls). The nucleic acids were ethanol precipitated, washed with 70% ethanol and dissolved in 50-100μl of 20μg/ml RNase in TE buffer and stored at 4°C.

2.3.2.19 Mapping the 5' End of Transcripts by *in vitro* Transcription and Primer Extension: (Kamakaka *et al.*, 1991)

In vitro transcription was done using the eukaryotic *in vitro* transcription kit from Stratagene. 5μl of supercoiled template DNA (100ng/μl) was mixed with 7.5μl of incubation buffer, 7.5μl of nucleotide mix, and 5μl of *Drosophila* extract (made

from 0-12 hour embryos by high speed centrifugation of a nuclear pellet). Transcription was performed at 25°C for 30 minutes and stopped by addition of 100μl of 'stop mix' and 5μl of proteinase K (2.5mg/ml), and incubated at room temperature for 3.5 minutes. 300μl of 0.3M Na Acetate pH5.5 and 400μl of phenol:chloroform:IAA (25:24:1) were added, vortexed for one minute, and centrifuged for five minutes. The upper aqueous layer was extracted with chloroform:IAA and transferred to a fresh Eppendorf. 1μl of primer (0.5ng) labelled at its 5' end with γ-^{32}P-ATP (see section 2.3.2.17) was added and the mixture precipitated with ethanol at room temperature. Following a 15 minute centrifugation, the pellet was dissolved in 100μl of 0.3M Na Acetate pH5.5 reprecipitated with ethanol. Nucleic acids were pelleted by centrifugation, washed with 75% ethanol, and dried under vacuum. The pellet was dissolved in 10μl of 1x annealing mix and the nucleic acids melted by heating to 75°C for 1.5 minutes. The tube was immediately transferred to a water bath at 58°C to allow annealing of the primer to mRNA. After one hour the mix was cooled to room temperature and 40μl of 1x extension mix and 1μl of AMV reverse transcriptase (Boehringer Mannheim, 25u/μl) added. Polymerisation from the annealed primer was at 37°C for one hour. The nucleic acids were ethanol precipitated, spun, and washed with 75% ethanol. Following drying, the pellet was dissolved in 3μl of 0.1M NaOH to hydrolyse the RNA. 6μl of formamide dye mix (99.8% deionised formamide; 0.02% bromophenol blue, w/v; 0.02% xylene cyanol, w/v; 1mM EDTA) was added, the solution boiled for three minutes, loaded onto a 10% prewarmed denaturing acrylamide gel (see section 2.3.2.13) and electrophoresed at 60v/cm in 1x TBE. An appropriate sequencing ladder was loaded beside the primer extension reaction as size markers. The gel was fixed in 10% methanol(v/v), 10% acetic acid (v/v), dried under vacuum, and autoradiographed at -70°C.

2.3.3 TISSUE CULTURE TECHNIQUES

2.3.3.1 Growth and Maintenance of *Drosophila* Schneider Line 2 Cells:

Cells were maintained at around 2-4x10^6 cells/ml in 1x Schneider's *Drosophila* medium supplemented with 5% foetal calf serum, 100u/ml penicillin and 100μg/ml streptomycin. Cells were incubated at 23°C and passaged every 7 days i.e. 25-50% of the cells were discarded and the volume replaced with fresh medium.

2.3.3.2 Growth and Maintenance of Chinese Hamster Fibroblast Cells:

RJK88 cells were maintained by Angela Pow at 37°C in a CO_2 conditioned incubator.

2.3.3.3 Storage and Recovery of *Drosophila* Schneider Line 2 Cells:

10-20mls of cells (2x10^7-8x10^7 cells) were harvested and centrifuged at 1000rpm for two minutes at 4°C. The supernatant was discarded and the cells resuspended in 0.5-2mls of freezing mix (15% foetal calf serum (v/v); 5% glycerol(v/v); 80% Schneider's medium (v/v)) and transferred in 0.5ml alliquotes to sterile vials. Vials were placed in an insulated box to facilitate slow freezing to -70°C overnight. Cells were transferred to liquid nitrogen for long-term storage.

Recovery of cells was by quickly thawing at 37°C then seeding into a small or medium sized tissue culture flask (Nunclon™, 25cm^2 or 175cm^2 flasks) in 10-20mls of medium. 2-5mls of medium were added once a week until sufficient density was reached to allow normal passaging.

2.3.3.4 Calcium Phosphate Transfection of Schneider Line 2 Cells: (Gorman, 1985)

One day before transfection 7-8mls of Schneider Line 2 cells were seeded on 7cm dishes at a density of 0.5-1x10^6 cells/ml. Plasmid DNAs to be used in transfections were prepared as described in section 2.3.1.4. Transfection of individual constructs was always done in duplicate in any one experiment. The calcium phosphate-DNA precipitate was made by mixing 10μg of test plasmid with 5μg of

internal control plasmid (p-19470Z) in 328.5μl of water in an Eppendorf tube. To this 46.5μl of 2M$CaCl_2$ (sterilised through a 0.2μM syringe filter) was added and mixed. 2xHBS (made from a 10x stock of 8.18% NaCl (w/v); 5.94% HEPES (w/v); 0.2% Na_2HPO_4 (w/v)) adjusted to 7.12 with NaOH was prepared and filter sterilised. 375μl of the 2xHBS was placed in a 6ml polystyrene falcon tube and the DNA/$CaCl_2$ mix added slowly dropwise with gentle vortexing. The solution was left at room temperature for 20-30 minutes to allow formation of the calcium phosphate-DNA precipitate. The 750μl precipitate was added dropwise to the surface of the Schneider line 2 cells which were incubated for a further two days at 23°C before preparing extracts.

2.3.3.5 Preparation of Schneider Line 2 Cell Extracts for CAT and β-galactosidase Assays:

Two days after transfection cells were harvested from the culture dish by scraping into a 10ml conical bottom sterelin tube. The cells were placed at 37°C for 20 minutes to induce high levels of β-galactosidase expression from the *hsp70* promoter of the internal control plasmid. Tubes were gently inverted several times during the heat shock to allow aeration of those cells that had settled. Tubes were then placed horizontally in a 23°C incubator for 90 minutes to allow translation of the *lacZ* mRNA. Cells were spun at 1000rpm for five minutes at 4°C. The supernatant was discarded and the pellet washed with 5mls of cold 1xPBS twice. After a further spin, the PBS was discarded and the pellet resuspended in 100μl of cold 0.25M Tris.HCl pH7.8. Cell membranes were broken by sonication in a water bath - a 30 second burst of sonication followed by a 10 second rest, repeated four times was sufficient. The extract was transferred to an Eppendorf tube and spun for two minutes. The supernatant was used both for CAT and β-galactosidase assays, and could be stored at -20°C for several months.

2.3.3.6 Calcium Phosphate Transfection of Mammalian Tissue Culture Cells:

One day before transfection 7-8mls of RJK88 cells were seeded onto 7cm

dishes at a density of $3x10^4$ cells/ml. Medium was changed the next morning prior to transfection. The $CaPO_4$ precipitate was prepared as in section 2.3.3.4, except that the internal control plasmid was pCMVβgal. One day after transfection the medium was changed again and the cells left for a further 24 hours before preparing the extract.

2.3.3.7 Preparation of Mammalian Cell Extracts for CAT and β-galactosidase Assays:

Extracts were prepared as described for Schneider Line 2 cells (section 2.3.3.5) omitting the heat shock and translational time.

2.3.3.8 Assaying CAT Activity in Tissue Culture Extracts:

25μl of extract was mixed with 2μl of 25mM acetyl coenzyme A (Lithium salt, from Sigma), 0.5μl of D-threo-[dichloroacetyl-1-^{14}C] chloramphenicol (53 or 57mCi/ml) and 13μl of water and incubated at 37°C for 30 minutes. The reaction was stopped and radioactivity extracted by adding 200μl of ethyl acetate and vortexing for one minute. The phases were separated by centrifugation for two minutes in a microfuge and the upper organic layer transferred to a fresh tube. The ethyl acetate was evaporated either in a speedvac centrifuge or overnight on the bench. 20μl of ethyl acetate was added to the dried tube and vortexed to collect chloramphenicol from the Eppendorf walls. The ethyl acetate was concentrated to a single spot on TLC silica matrix (Merck) and acetylated chloramphenicol separated from chloramphenicol by silica gel TLC in a seeled tank using 95% chloroform (v/v), 5% methanol (v/v) as the solvent (Cohen *et al.*, 1980). The silica matrix plate was removed from the tank and dried when the solvent front reached within one cm of the top. The matrix was autoradiographed at room temperature. CAT activity was calculated by measuring the radioactivity in a scintillation counter of 3-acetate chloramphenicol and expressing this as a percentage of the combined activity of 3-acetate chloramphenicol and chloramphenicol. Assays with conversion values greater than 55% were repeated using less extract (with the missing volume replaced by

0.25M Tris pH7.8), as values greater than 55% conversion do not correspond to the amount of CAT enzyme.

2.3.3.9 Assaying β-galactosidase Activity in Tissue Culture Extracts: (based on Sun *et al.*, 1988)

10μl of extract was mixed with 200μl of Z buffer (60mM Na_2HPO_4; 40mM NaH_2PO_4; 10mM KCl; 1mM $MgSO_4$; 50mM β-mercaptoethanol), 100μl of ONPG (4mg/ml in 0.1M KPO_4 pH7) and incubated at 37°C for 10-30 minutes (in any one experiment the reaction time was the same for each sample). The reaction was stopped by addition of 200μl of 1M Na_2CO_3 and the OD_{420nm} read to estimate β-galactosidase activity. Preliminary experiments using different dilutions of particular extracts for the same length of time, showed β-galactosidase activity (as measured by absorbance at 420nm) to vary linearly with the quantity of extract used up to at least 0.700 OD_{420nm} units. Thus for any experiment on a particular day (which comprised several extracts being assayed) the same volume of extract was assayed for the same length of time and the OD_{420nm} value used directly to normalise CAT activity values.

2.3.4 MANIPULATIONS OF *DROSOPHILA* STOCKS:

2.3.4.1 Establishment of Transgenic Lines by P Element Transformation: (Rubin and Spradling, 1982)

DNAs for transformation were prepared as described in section 2.3.1.4. The test plasmid was mixed with helper plasmid (pΠ25.7wc or phsp70Δ2-3wc) at a ratio of 5:1. The DNAs were ethanol precipitated and dissolved in injection buffer (5mM KCl; 0.1mM$NaPO_4$ pH6.8) to give a final concentration of 500μg/ml of test plasmid and 100μg/ml of helper. The DNA was spun for two minutes in a microfuge to pellet any debris, and sucked into the injection needle which was then stored under halocarbon oil to prevent drying. Eggs were collected every 30 minutes from 4-10 day old W^K flies kept in a chamber with a sliding food drawer to allow easy removal

of eggs. Eggs were dechorionated manually by gently rolling across double sided tape (Scotch brand 3M No.666) on a microscope slide. The naked eggs were lined up on a narrow strip of tape with their posterior ends all pointing in the same direction. Embryos were desiccated in petri dish containing silica gel for 3-6 minutes (depending on the humidity and temperature of the room) and then covered with halocarbon oil. DNA was injected gently into the posterior pole of preblastoderm eggs - older eggs were killed by tearing the vitelline membrane. Eggs were incubated at 20°C for 36-60 hours, larvae collected as they emerged, placed on food, and incubated at 25°C. Adults were mated with W^K virgins and their progeny examined for a red eye phenotype. Transformants were crossed with W^K virgins and homozygous lines established by sibling matings.

2.3.4.2 Collection of *Drosophila* Developmental Stages:

All *Drosophila* stages were collected from vials kept at 25°C. For each stage, ten females and ten males were frozen in Eppendorf tubes according to their sex.

30-35 female and male third instar larvae were collected and ten of each sex frozen to -70°C. The remaining 20-25 larvae were placed in separate female and male vials. After incubation at 25°C for two days, ten females and ten male pupae were collected and frozen as early pupae. The remaining pupae were returned to 25°C for a further two days (by which time the majority of the pupae were black) and then collected as late pupae and frozen to -70°C. 0-1 day old adults were collected by emptying parents from two week old vials. Newly eclosed adults were gathered 24 hours later and frozen. For 3-4 day old adults, 0-1 day adults were placed in a fresh vial, incubated for three days, and then frozen. 0-2 hour old embryos (females and males together) were collected as in the section 2.3.4.1, except the sliding food tray was changed every 90 minutes. Eggs were washed from the food with 1xPBS, rinsed twice with 1x PBS in a fine mesh basket, and frozen. Ovary and carcass tissues were collected by dissection from ten 3-4 day old adults in Ringers solution.

2.3.4.3 Preparation of CAT Extracts from *Drosophila* Tissues: (Mullins, 1990)

200μl of 0.25M Tris.HCl pH7.8 was added to thawed *Drosophila* stages on ice (only 100μl of 0.25M Tris.HCl pH7.8 was added to ovary, carcass and embryo tissues). Homogenisation was done in an Eppendorf tube using a motorised homogeniser with pestle. The homogenate was frozen in liquid nitrogen and thawed in a 37°C water bath five times to break open the cell membranes. The homogenate was then heated to 65°C for five minutes to precipitate most proteins which interfere with the CAT assay (CAT can withstand this temperature). The extract was then spun in a microfuge to pellet cell debris and some denatured proteins. The supernatant was transferred to a fresh tube as the CAT extract, and stored at -70°C.

2.3.4.4 Assaying *Drosophila* Extracts for CAT Activity:

CAT assays were performed as described for tissue culture extracts (2.3.3.8) except incubation at 37°C was for one hour. CAT activity was expressed as pmoles of ^{14}C-chloramphenicol acetylated per mg of protein per minute.

2.3.4.5 Estimation of Protein Concentration in *Drosophila* Extracts:

Protein concentrations of 1-25μg/ml were estimated using the Bradford method (Bradford, 1976; Read and Northcote, 1981), which utilises the protein binding properties of Coomassie blue G-250. Binding causes a shift in colour from red/brown to blue corresponding to an absorbance shift from 465 to 595nm. 5μl of protein extract was made up to 1ml with water and 1ml of Pierce Coomassie protein assay reagent (contains Coomassie blue G-250, phosphoric acid, methanol, and water) added and mixed. The OD_{595nm} was immediately measured and the concentration of protein estimated from a standard graph of BSA (0-25μg/ml on 0.00125M Tris.HCl, pH7.8), versus OD_{595nm}. A standard graph was prepared each time the protein concentration of extracts was measured as the Bradford assay sensitive to fluctuations in temperature.

2.3.4.6 Whole Mount *in situ* Hybridisation of Digoxygenin Labelled DNA to Ovary Transcripts:

The method used for *in situ* hybridisation to ovaries was adapted from the method described by Tautz and Pfeifle (1989) for *Drosophila* embryos. Ovaries were dissected from 1-4 day old flies (kept at 25°C) in 4% paraformaldehyde(w/v) in 1xPBS (PP), then placed on ice. Ovaries were washed in 0.1% triton X-100(v/v) and then with water. They were fixed in a solution of 0.8mls of 100mM HEPES (pH 6.9), 2mM $MgSO_4$, 1mM EGTA, and 0.2mls of 20% paraformaldehyde(w/v) for 20 minutes, on a rotating wheel (the paraformaldehyde solution was prepared by heating at 65°C, and neutralising with NaOH). 5mls of methanol were added and shaken for 10 seconds. The ovaries were transferred to an Eppendorf tube and washed once in ME (9 parts methanol to one part EGTA, pH8). At this stage ovaries could be stored at -20°C for up to two weeks. The ovaries were then fixed and rehydrated by the following mixing series:

5 minutes in 700μl of ME; 300μl of PP.
" " " 500μl " " 500μl " "
" " " 300μl " " 700μl " "
20 " " 1ml of PP.

The ovaries were washed three times in 1ml of PBT (1x PBS, in 0.1% tween-20(v/v)) for five minutes each time. To this 5μl of 10mg/ml of proteinase K was added and left for 3.5 minutes at room temperature. Proteolysis was stopped by washing the ovaries for 2 minutes in 2mg/ml glycine in PBT, followed by two more washes in PBT for five minutes each. The ovaries were refixed in PP for 20 minutes, and then washed five times for five minutes in PBT. Equilibration with hybridisation solution (50% formamide in 5xSSC; 100μg/ml of denatured sonicated salmon sperm DNA; 50μg/ml of heparin; 0.1% tween-20(v/v)) was done by washing the ovaries for 10 minutes in a 1:1 mix of PBT and hybridisation solution, followed by a ten minute wash in hybridisation solution. Prehybridisation was in 1ml of hybridisation solution in an Eppendorf tube, for one hour at 45°C in a shaking water bath. 500-700μl of the hybridisation solution was removed and 10μl of heat denatured DNA probe added to

give a final concentration of 0.5-1μg/ml (see 2.3.2.10 for probe preparation). Hybridisation occurred overnight in a shaking water bath at 45°C. The hybridisation solution was removed and the ovaries washed in serial mixtures as follows:

1ml of hybridisation solution for 20 minutes.

1ml of 4:1 hybridisation solution:PBT for 20 minutes.

" " 3:2 " " " " " "

" " 2:3 " " " " " "

" " 1:4 " " " " " "

" " PBT for 20 minutes.

Non specific antibody binding sites were saturated by spinning the ovaries on a rotating wheel in 10% sheep serum(v/v) in PBT (Hemmati-Brivanlou *et al.*, 1990). While washes and blocking were in progress, 1ml of a 1 in 1000 dilution of anti-dig antibody (polyclonal sheep digoxigenin F_{ab} fragment conjugated to alkaline phosphatase, 750u/ml, from Boehringer Mannheim) was made up in PBT and pre adsorbed to 30-50 fixed dissected W^K female flies (both ovaries and carcass) for 2 hours at room temperature on a rotating wheel. The preabsorbed antibodies were then diluted 2 fold in PBT with sheep serum to a final concentration of 10%, and adsorbed to the blocked ovaries for one hour. Ovaries were washed four times in PBT for 20 minutes and then three times in washing buffer (100mM NaCl; 50mM $MgCl_2$: 100mM Tris.HCl, pH9.5; 1mM levamisol; 0.1% tween-20(v/v)) five minutes each time. To the last wash 4.5μl of NBT (75mg/ml nitroblue tetrazolium salt in dimethyl formamide, 70% (v/v)) and 3.5μl of X-phosphate (50mg/ml 5-bromo-4-chloro-3-indoyl phosphate, toluidinium salt in dimethyl formamide), both from Boehringer Mannheim were added to the ovaries to allow staining. The ovaries, in this solution, were transferred to a microtitre plate, placing 4-6 ovaries in each well, and covered in aluminium foil. The ovaries were examined under a dissection microscope every 10-15 minutes and staining stopped by replacing the solution with PBT. Ovaries were mounted on microscope slides in DPX medium and spread overnight by placing weights on the coverslip, and photographed the next day.

2.3.4.7 Assaying Sterility of SF and RSF Female Flies:

Progeny of dysgenic and reciprocal crosses were collected 0-4 days after eclosion and placed in a cylindrical egg collection chamber. The chamber was open at one end over which a petri dish containing French fly food, blackened with charcoal powder, was placed. SF or RSF females were allowed to mate with their brothers and lay eggs over a 24 hour period at 25°C. Food dishes were then replaced with fresh ones and the extracted dish incubated for two days at 25°C. Sterility was measured as the percentage of eggs which had hatched - scored by counting empty and full chorions.

2.3.4.8 Conversion of Reactive Lines to Isogenic Inducer Lines:

The following matings converted reactive to inducer flies, an example with the transgene on the third chromosome is given. All flies were M cytotype and therefore unaffected by P-M dysgenesis.

125 I **126 I**

G_0 ♂ $\frac{+;\ Cy;\ Dcxf}{1\ \ Sm6\ SbIn(Mo)}$ x ♀ $\frac{w^aM5B;\ +;\ +}{w^aM5B\ +\ +}$

↓

I **R**

G_1 ♂ $\frac{w^aM5B;\ Cy;\ Dcxf}{1\ \ +\ \ +}$ or $\frac{w^aM5B;\ Cy;\ SbIn(Mo)}{1\ \ +\ \ +}$ x ♀ $\frac{w^-;\ +;\ w^+CAT}{w^-\ +\ w^+CAT}$

↓

W^K **SF**

G_2 ♂ $\frac{w^-;\ +;\ +}{1\ +\ +}$ x ♀ $\frac{w^aM5B;\ Cy;\ Dcxf}{w^-\ +\ (w^+CAT)}$ or $\frac{w^aM5B;\ Cy;\ SbIn(Mo)}{w^-\ +\ (w^+CAT)}$

↓

G_3 ♂ $\frac{w^-;\ +;\ +}{1\ +\ w^+CAT}$ or ♀ $\frac{w^-;\ +;\ +}{w^-\ +\ w^+CAT}$ (heterozygous for i^+ and w^+CAT)

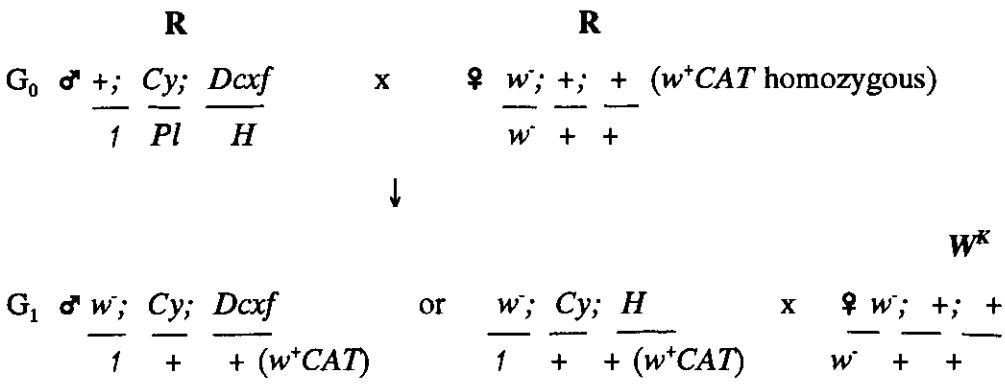
↓

Sibling matings to establish homozygous line.

Examination of the G_3 progeny allowed mapping of the w^+CAT transgene i.e. the marker not inherited with w^+ flies maps to the same chromosome as the transgenome. Conversion of W^K to an isogenic inducer strain was done by crossing W^K females directly with line 125 males, as the X chromosome of W^K is already marked with w^- and therefore could be followed through the dysgenic cross.

2.3.4.9 Mapping a Transgene to a Chromosome:

The chromosome to which w^+CAT mapped for lines not made inducer was determined by the following cross:



G_2 EXAMINE PROGENY

The marker not inherited with w^+ flies in the G_2 generation maps to the same chromosome as the transgene. Transgenes mapping to the X chromosome gave only female w^+ flies in the G_2 generation.

2.3.5 MISCELLANEOUS:

2.3.5.1 Autoradiography:

Radioactive signals were detected on Dupont Cronex-4 X-ray film with intensifying screens at -70°C for ^{32}P, and room temperature for ^{35}S and ^{14}C. The film was preflashed at an OD_{540} of 0.15 for ^{32}P detection.

RESULTS

CHAPTER 3

3.1 Introduction to the Concept of a Promoter Within *I*:

I factor transposition is mediated through an RNA intermediate, a candidate for which initiates at or very near nucleotide +1 (Pélisson *et al.*, 1991; Chaboissier *et al.*, 1990). The model for transposition proposes an internal promoter directs precise transcription initiation at base +1. Fawcett *et al.*(1986) proposed that the *I* factor contains an internal RNA polymerase II promoter. The suggestion that a promoter lies completely within the transcription unit is contrary to the classical RNA polymerase II promoters but more typical of an RNA polymerase III promoters. However Fawcett *et al.* pointed out that polymerase III is unlikely to transcribe the *I* factor as both strands of the element contain several oligo-T sequences which serve as RNA polymerase III terminators, and the first 200bp lack any regions similar to polymerase III box A or box B promoter consensus sequences.

I would like to discuss current understanding of RNA polymerase II promoters in general and then collate the evidence that has accumulated in the past three or four years which supports the existence of a new group of RNA polymerase II promoters that comprises mainly, and in some cases exclusively, of sequences at and downstream of the transcription initiation site.

3.2 RNA Polymerase II Promoters:

RNA polymerase II is a multisubunit protein catalysing transcription of genes encoding proteins and some snRNAs (small nuclear RNAs). Three groups of so called class II promoters shall be discussed:

Class II.1 - TATA containing promoters.

Class II.2 - CpG rich promoters.

Class II.3 - TATA-less, non-CpG rich promoters.

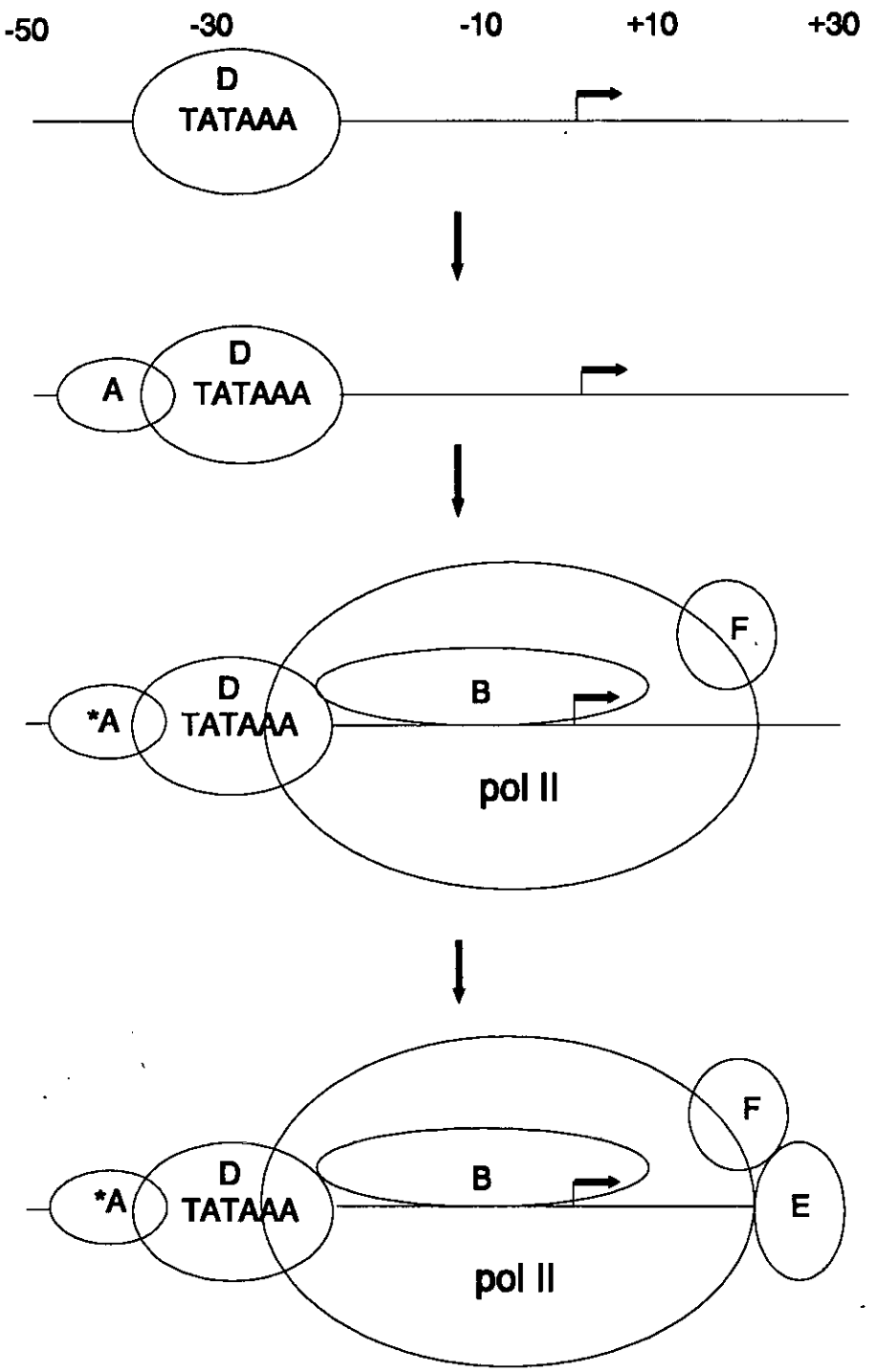
3.2.1 Class II.1 - TATA Promoters:

By far the best characterised group of class II promoters are those with the sequence TATA$^{A}_{T}$A$^{T}_{A}$ at position -20 to -30 (Corden *et al.*, 1980) which together with the RNA start site (at position +1) form the minimal promoter. Much of our

understanding of transcription initiation has been gained from kinetic studies of transcription initiation using reconstituted *in vitro* systems with the TATA containing adenovirus-2 major late promoter (Ad-MLP) and fractionated Hela cell extracts. The order of events is depicted in figure 3.1 and has been reviewed by Swadadogo and Sentenac (1990), and Saltzman and Weinmann (1989). The first step in assembly of the active transcription complex is binding of TFIID to the TATA box (Fire *et al.*,1984; Buratowski *et al.*, 1989) stabilisation of which requires TFIIA (Buratowski *et al.*, 1989; Maldonado *et al.*, 1990). Factors -IID and -IIA form a preinitiation complex committed to transcription of the bound template DNA (Fire *et al.*, 1984). This DA complex is specifically recognised by TFIIB, association of which requires sequences downstream of the RNA start site possibly to hold it in place (Maldonado *et al.*, 1990). The next step is recruitment of RNA polymerase II via the RAP30 subunit of TFIIF (Flores *et al.*, 1991) that specifically interacts with RNA polymerase II (Flores *et al*, 1988; Sopta *et al*, 1985). TFIIA may be dispensable for this, and subsequent steps as complexes that form after TFIIB binding can do so in the absence of TFIIA (Buratowski *et al.*, 1989). TFIIB is thought to acts as a bridge between TFIID and RNA polymerase II/RAP30 facilitating stable polymerase binding (Buratowski *et al.*, 1989). TFIIE then complexes with RNA polymerase II and in so doing protects downstream sequences from +20 to +30. The preinitiation complex spans the promoter from -40 to +30 relative to the RNA start site. Addition of dNTPs allows accurate transcription initiation. Together with RNA polymerase II, these five transcription factors are sufficient for basal transcription initiation from a minimal promoter. Other transcription factors may bind the promoter to invoke a positive or negative effect by interacting or competing with general transcription factors. For example yeast GAL4 transcription factor interacts directly with TFIID to stimulates transcription *in vitro* from a hybrid adenovirus E4 promoter with GAL4 binding sites (Horikoshi *et al.*, 1988); Sp1 activates transcription by association with TFIID indirectly via 'coactivators' (Pugh and Tjian, 1990); the *Drosophila* protein Engrailed represses transcription from the *hsp70*, Ad-ML, and *fushi -tarazu* promoters *in vitro* by competing with TFIID for the TATA sequence (Ohkuma *et al.*, 1990).

Figure 3.1 : **A model of the complexes that lead to transcription initiation of Class II.1 promoters** (adapted from Buratowski *et al.*, 1989, and Flores *et al.*, 1991). The numbers represent the nucleotide position relative to the transcription start site at +1. The arrow at position +1 indicates the direction of transcription. The asterisk in TFIIA indicates this factor may be dispensable for complex formation.

Figure 3.1


-50
-30
-10
+10
+30
D
TATAAA
A
D
TATAAA
F
*A
D
TATAAA
B
pol II
F
E
*A
D
TATAAA
B
pol II

Although assembly of the transcriptional machinery has been elucidated in a mammalian system, the events in *Drosophila* are likely to be similar as there is sequence conservation between general transcription factors of mammals and *Drosophila* (Ahearn *et al.*, 1987; Kao *et al.*, 1990). Functional conservation has also been demonstrated in that mammalian promoters are transcribed accurately and efficiently by *Drosophila* transcription systems (Heiermann and Pongs, 1985; Kamakaka *et al.*, 1991) and the *Drosophila* transcription machinery can be activated by mammalian specific transcription factors (Santoro *et al.*, 1988; Courey and Tjian, 1988).

The only general transcription factor with specific DNA binding activity is TFIID that binds the TATA box. In the promoters of higher eukaryotes the TATA box functions in transcription start site selection 30 or so nucleotides downstream (Grosveld *et al.*, 1981), and mutation of the TATA box can cause initiation heterogeneity (Grosschedl and Birnsteil, 1980; Grosveld *et al.*, 1982). However in *S. cerevisiae* the properties and function of the TATA box differ in that there can be more than one functional TATA box, the TATA box(es) are situated 60-120bp upstream of the mRNA start site, and sequences near the start site rather than the TATA box are required for correct initiation (Hahn *et al.*, 1985; Nagawa and Fink, 1985; Chen and Struhl. 1985). Moreover several sequences that bear no, or very little, resemblance to the consensus can substitute the TATA box in *S. cerevisiae* without loss of function (Singer *et al.*, 1990). The TATA box in some higher eukaryotic promoters (including the Ad-MLP) has been shown to play a role in initiation frequency rather than start site selection as shown by reduction in transcription levels following its mutation (Wasyly *et al.*, 1980; Carcamo *et al.*, 1991). Reduction in transcription frequency following mutation of the TATA box of the Ad-ML promoter correlates with loss of TFIID binding (Carcamo *et al.*, 1991).

As well as activating basal transcription, TFIID can mediate transcription activation by interacting with the activation domains of promoter specific transcription factors (reviewed by Greenblatt, 1991).

TFIID is highly conserved throughout eukaryotes. Partially purified yeast and human TFIID (yTFIID and hTFIID respectively) are functionally interchangeable *in vitro* (Buratowski *et al.*, 1988; Cavallini *et al.*, 1988; Horikoshi *et al.*, 1989; Flanagan *et al.*, 1990) and a cell fraction containing *Drosophila* TFIID (dTFIID) can substitute hTFIID *in vitro* (Hoey *et al.*, 1990). The TFIID genes from several species, including human (Kao *et al*, 1990), *Drosophila* (Hoey *et al.*, 1990), and *S. cerevisiae* (e.g. Hahn *et al.*, 1989a) have been cloned. The TFIID proteins of all species characterised to date share a highly conserved carboxyl terminus domain sufficient for TATA box binding (although purified yTFIID can bind non-consensus TATA sequences, Hahn *et al.*, 1989b) and can interact with TFIIA and TFIIB to activate basal transcription levels (Horikoshi *et al.*, 1990; Hoey *et al.*, 1990; Peterson *et al.*, 1990). In contrast to the *in vitro* situation hTFIID and dTFIID fail to complement yTFIID *in vivo*. This species specificity also maps to the conserved carboxyl terminus (Gill and Tjian, 1991; Cormack *et al.*, 1991).

The amino terminal domain of TFIID diverges from one species to another in both sequence and length. Its function is thought to be in 'coactivator' association (i.e. proteins contained in TFIID cell fractions that facilitate productive interaction between the TATA-box binding protein and transcription activators) as transcription activation by Sp1 is abolished when the N terminal domain of TFIID is removal (Pugh and Tjian, 1990).

3.2.2 Class II.2 - GC Rich Promoters:

Less is known of the mechanism of transcription initiation from GC rich promoters - they lack TATA box homology and have GC rich sequences typically at position -40 to -70. Transcription initiation is heterogenous but often has one or two major start sites and can be bidirectional (Melton *et al.*, 1984; Melton *et al.*, 1987; Crouse *et al.*, 1985).

This class of promoter is mainly associated with housekeeping genes, that is genes expressed at a low constitutive level in all tissues, such as hypoxanthine-guanine phosphoribosyltransferase (HPRT) (Melton *et al.*, 1984). The promoters of some

growth control genes e.g. c-Harvey *ras* (Ishii *et al.*, 1985), a few tissue specific genes such as α^D-globin (Kemper *et al.*, 1987) and some viruses (reviewed in McKnight and Tjian, 1986) also have GC rich motifs.

The sequence $\frac{G}{T}GGGCGG\frac{GGC}{AAT}$, which is the consensus binding site for the mammalian transcription activator Sp1 (Kadanoga *et al.*, 1986), is found repeated several times within this class of promoter. Sp1 was first identified as a host protein required for SV40 gene expression. It acts by binding 'GC' motifs within the 21bp repeats of the SV40 early promoter and activates transcription from both SV40 early and late promoters (Dynan and Tjian, 1983a; Dynan and Tjian, 1983b). Sp1's role in promoter binding and transcription activation for housekeeping genes has been shown *in vitro* for mouse dihydrofolate reductase (DHFR) (Dynan *et al.*, 1986). The mechanism of Sp1 activation has not yet been elucidated class II.2 promoters. However it has been demonstrated that transcription activation by Sp1 from a synthetic TATA-less promoter containing the 'GC' motifs form the SV40 early promoter upstream from the Ad-ML 'initiator' (see section 3.2.3) requires TFIID, as well as TFIIB, -IIE, and -IIF (Smale *et al.*, 1990; Pugh and Tjian, 1990; Pugh and Tjian, 1991). Interaction of the TATA binding protein and Sp1 is not direct but requires accessory factors termed TATA-box binding protein associated factors (TAFs).

The dinucleotide CpG is rare in the genomes of vertebrates occurring at about 20% the expected frequency with between 60 and 90% of CpGs methylated at position 5 on the cytosine ring (5mC). Rarity is thought to be a consequence of deamination of 5mC to thymidine which is overlooked by repair mechanisms and upon replication C-G basepairs would become T-A (Bird, 1980). GC rich promoters lack methylation and are enriched for CpG (about 65% G+C). As such they appear CpG rich and are termed methylation free islands (reviewed in Bird, 1986).

The genome of *D. melanogaster* lacks methylation hence CpG is not under represented. In addition *Drosophila* cells lack an Sp1 homolog (Courey and Tjian, 1988) making it unlikely that class II.2 promoters, as defined above, exist in

Drosophila. The sequence of three *D. melanogaster* housekeeping genes, glucose-6-phosphate dehydrogenase (G6PD) (Fouts *et al.*, 1988), and glyceraldehyde-3-phosphate dehydrogenase-1 and -2 (Gapdh-1 and Gapdh-2) (Tso *et al.*, 1985) has been determined. All three genes lack a TATA box. The Gapdh-2 promoter has the sequence TTTTCGCCC directly repeated twice that regulates expression during development (Sun *et al.*, 1988). G6PD has the sequence GCGGCG and the inverse complement, CGCCGC, upstream of the transcription initiation site but their function, if any, has not been tested (Fouts *et al.*, 1988). These novel 'GC' motifs might represent binding sites for transcription factors that regulate expression of *Drosophila* housekeeping genes.

3.2.3 Class II.3 - TATA-less, Non-CpG Rich Promoters:

A number of RNA polymerase promoters have emerged recently that lack either a TATA box or a GC rich region. Many such promoters direct expression of developmentally regulated genes e.g. *terminal deoxynucleotydyl transferase* (TdT) expressed during T and B lymphocyte differentiation (Smale and Baltimore, 1989), *porphobilinogen deaminase* (PBGD) expressed during erythroid differentiation (Beaupain *et al.*, 1990), and *D. melanogaster Ultrabithorax* (*Ubx*) (Biggin and Tjian, 1988), *engrailed* (*en*) (Soeller *et al.*, 1988), and *Antennapedia* (*Antp*) (Perkins *et al.*, 1988) all regulated during *Drosophila* development. In addition the adenovirus IVa2 (Carcamo *et al.*, 1990), the SV40 major late (Ayers and Dynan, 1988) and some retrovirus-like transposon (Arkhipova and Ilyin, 1991; Jarrel and Meselson, 1991) and LINE-like transposon (Mizrokhi *et al.*, 1988; Minchiotti and Di Nocera, 1991) promoters fall into this class. Transcription initiation is at a single or a few clustered sites. In all promoters of this type tested to date accurate basal transcription initiation *in vitro* has limited or no dependence on upstream sequences but requires sequences at and downstream of the start site.

Detailed analysis of the murine TdT promoter identified a sequence termed the 'initiator' essential for correct transcription initiation which surrounds the mRNA start site (Smale and Baltimore, 1989). The functional initiator is CTCATTCT with the A

being the transcription start site (+1) (Smale *et al.*, 1990). Similar motifs with the core consensus TCANT (with a preference of N=G or T) are present at or near the transcription initiation site of many TATA-less (Arkhipova and Ilyin, 1991) and TATA containing promoters including the Ad-ML promoter (Hultmark *et al.*, 1986; Smale and Baltimore, 1989; Hoopes and Rohrmann, 1991). The function of this sequence is to defines the precise nucleotide at which transcription starts (Smale *et al.*, 1990; Beaupain *et al.*, 1990). In promoters with both the initiator core TCAG/TT and TATA elements the initiator core defines the start site as mutation of the TATA box does not cause heterogeneity (Hirsh *et al.*, 1986; Thummel, 1990; Hen *et al.*, 1982; Carcamo *et al.*, 1991).

Sequences downstream of the initiator are mandatory in several promoters of this type for basal transcription. Comparison of internal sequences has identified conserved motifs between bases +30 and +40 (Arkhipova *et al.*, 1991, and this thesis) that have been shown in some promoters to be important for transcription. The investigations carried out in this chapter establish that the *I* factor promoter falls into this group of class II promoters.

3.3 Strategy for *I* Factor Promoter Detection:

The primary aim of this thesis is to identify those sequences which direct transcription of the *I* factor. Detection of DNA sequences encoding a promoter is most conveniently accomplished by linking the putative promoter to a reporter gene on a plasmid and assaying activity by transient expression in tissue culture. This method is not suitable for all promoters as *trans* acting factors essential for transcription, particularly of tissue specific and/or developmentally regulated genes, may be absent in tissue culture cells. Full length transcripts of *I* are confined to the ovaries of SF and RSF females. Transcription is therefore tightly controlled. Regulation may be at the level of the promoter in which case results gained from transient assays may be unobtainable, misleading, or transcription may proceed at basal levels. Nevertheless, due to the ease of the assay, this method was employed to test *I* factor promoter activity prior to more involved techniques, but caution

exercised when interpreting results.

3.4 RESULTS and CONCLUSIONS:

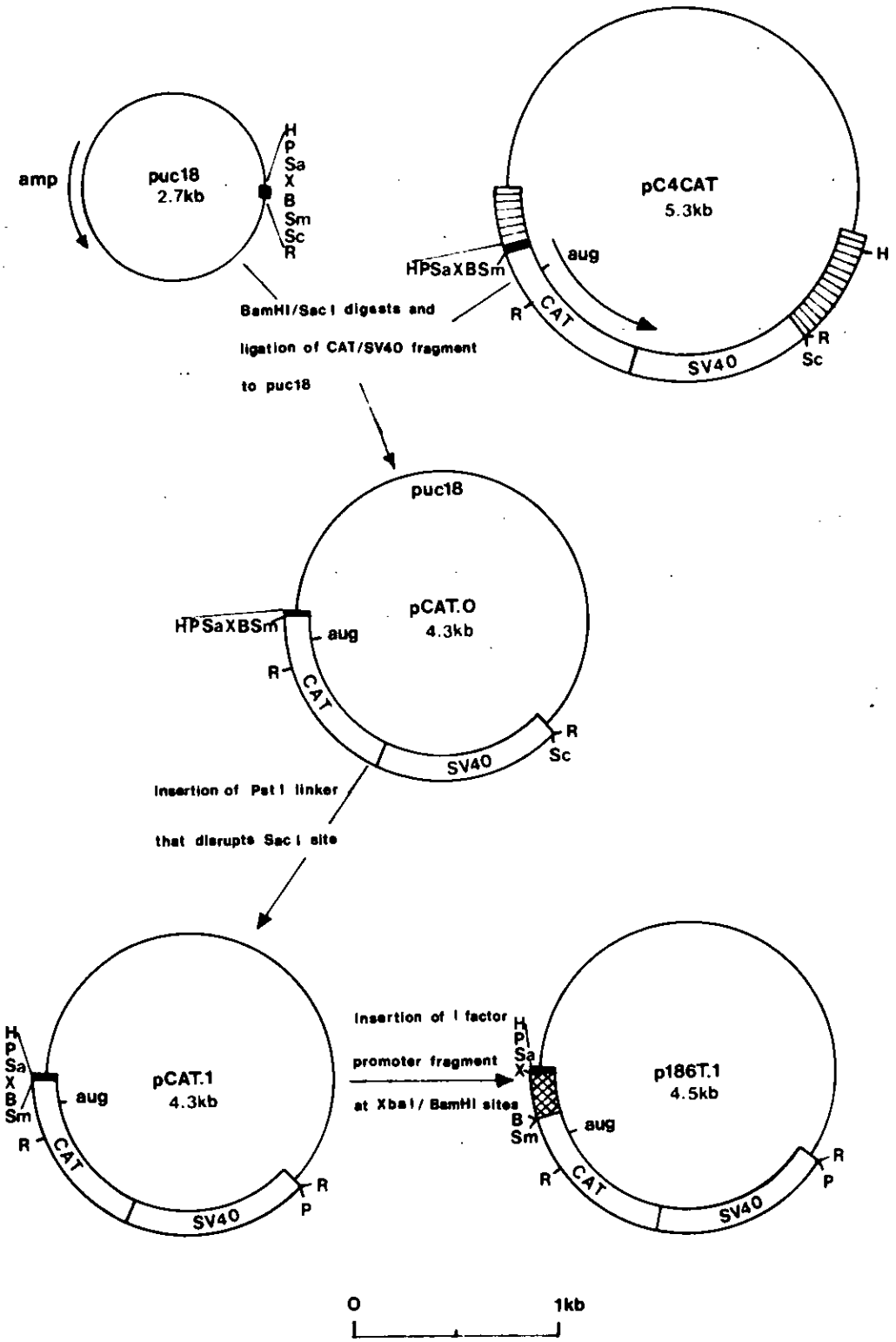
3.4.1 Construction of Plasmids for Transient Assays:

The *I* factor promoter may be rather weak as full length transcripts are rare or undetectable in SF females. To optimise detection of promoter activity in transient assays the reporter gene chloramphenicol acetyl transferase (CAT) was used as the assay for this enzyme is reputed to be sensitive. The plasmid pC4CAT (Thummel *et al.*, 1988) was available, however it produced measurable conversion of ^{14}C-Chloramphenicol to its acetylated forms (CAT activity) in the absence of a promoter. It was feared this would mask *I* promoter activity, therefore a plasmid named pCAT.1 was constructed as shown in figure 3.2, tailored to the specific requirements of this thesis.

All promoter fragments were synthesised by the polymerase chain reaction (PCR) using pI954 (Pritchard *et al.*, 1988) as a template. pI954 was chosen because it contains a full length *I* factor which can transpose and induce hybrid dysgenesis when introduced into reactive flies by P-element transformation. It therefore has a functional promoter. The PCR primers were designed with an *Xba*I site (for the plus strand primer) or a *Bam*HI site (for the minus strand) at their 5' end. The PCR products therefore terminate with these sites and could be directionally cloned (figure 3.2). *Xba*I and *Bam*HI were chosen because double restriction digestion proceeds to completion when the sites are adjacent (Crouse and Amores, 1986). Restriction should therefore also be efficient when the sites are at the ends of a DNA molecule as in the PCR products. The DNA sequence of the PCR products was checked prior to the promoter assay either by cloning into M13 mp18 or mp19 and sequencing, or by plasmid sequencing once in pCAT.1. The recipient cell line for DNA transfection was *Drosophila* Schneider Line 2 (Schneider, 1972), which is reactive (Lynch, 1989). Transfection efficiency using the calcium phosphate method is influenced by several factors including the integrity of the plasmid DNA and the condition of the recipient

Figure 3.2 : **Construction of the expression vector (pCAT.1) and insertion of *I* factor promoter fragments.** pC4CAT is described in Thummel *et al.* (1988). Abbreviations for restriction sites are: B - *Bam*HI; H - *Hind*III; P - *Pst*I; R - *Eco*RI; Sa - *Sal*I; Sc - *Sac*I; Sm - *Sma*I; X - *Xba*I. The sequence of *Pst*I linker used in construction of pCAT.1 (oligonucleotide 663C, see section 2.2.9) is given in the materials section. aug represents the methionine codon of the CAT gene and the arrow represents the direction of CAT transcription. The SV40 sequences contain the polyadenylation site and small t intron. *P* element sequences are shown as the vertically stripped boxes, and *I* factor sequences as the crossed box. Construction of p186T.1 is depicted here, all other constructs were made by the same procedure.

Figure 3.2


amp
puc18
2.7kb
H
P
Sa
X
B
Sm
Sc
R
pC4CAT
5.3kb
HPSaXBSm
aug
R
CAT
SV40
R
Sc
H
BamHI/Sac I digests and
ligation of CAT/SV40 fragment
to puc18
puc18
pCAT.O
4.3kb
HPSaXBSm
aug
R
CAT
SV40
R
Sc
Insertion of Pst I linker
that disrupts Sac I site
H
P
Sa
X
B
Sm
R
pCAT.1
4.3kb
aug
CAT
SV40
R
P
Insertion of I factor
promoter fragment
at Xba I / BamHI sites
H
P
Sa
X
B
Sm
p186T.1
4.5kb
aug
R
CAT
SV40
R
P
0
1kb

cells. Except for pCAT.2, p186T.2, pΔ38, pΔ100, p186.29T, p186.29C, p186.34C, p40G*w*, and p50Grp, at least two different plasmid preparations were assayed to control for DNA integrity. By cotransfecting with p-194.70Z (Xiao and Lis, 1988), transfection efficiency could be measured and the CAT activity of each promoter construct corrected accordingly.

3.4.2 Detection of *I* Factor Promoter Activity:

Nucleotides +1 to +186 of the *I* factor sequence (as published by Fawcett *et al.*, 1986) were selected as a putative promoter region as they make up the entire 5' untranslated region (see appendix A for sequence). This sequence was amplified by PCR and cloned into pCAT.1 to form p186T.1 (figure 3.2). To measure the influence of upstream puc18 sequences on promoter activity, the *Pst*I fragment containing the CAT gene was inverted in pCAT.1 and p186T.1 to make pCAT.2 and p186T.2 respectively (figure 3.3a). An autoradiograph of a TLC plate on which the acetylated forms of ^{14}C-chloramphenicol have been separated from the unacetylated form is shown in figure 3.3b. Cells transfected with p186T.1 and p186T.2 exhibit high levels of conversion in contrast to pCAT.1 and pCAT.2 that have little and no detectable CAT activity respectively. The *I* factor therefore has an internal promoter within its first 186 nucleotides. A weak cryptic promoter is present in the puc18 sequences giving low, but measurable CAT activity in pCAT.1. All other constructs assayed in this thesis have the promoter/CAT fusion in the same orientation, with respect to the puc18 sequences, as in p186T.1.

In each experiment duplicate transfections were done for each construct, and the mean value for the two p186T.1 transfections (once corrected for the transfection efficiency) expressed as 100%. The corrected value for all other constructs is expressed as a percentage relative to 100%. Normally, two independent duplicate transfections were done for each construct (i.e. four plates), and the standard deviation of the four individual transfection calculated, rather than the standard deviation of the mean value for each experiment, which would be smaller, but less representative.

Figure 3.3a : **Construction of pCAT.2 and p186T.2.** p186T.2 was made by K. Hatrick. Abbreviations for restriction sites are as in figure 3.2.

Figure 3.3b : **Transient assay of pCAT.1, pCAT.2, p186T.1, and p186T.2.** The acetylation products are labelled at the side of the autoradiograph.

Figure 3.3a

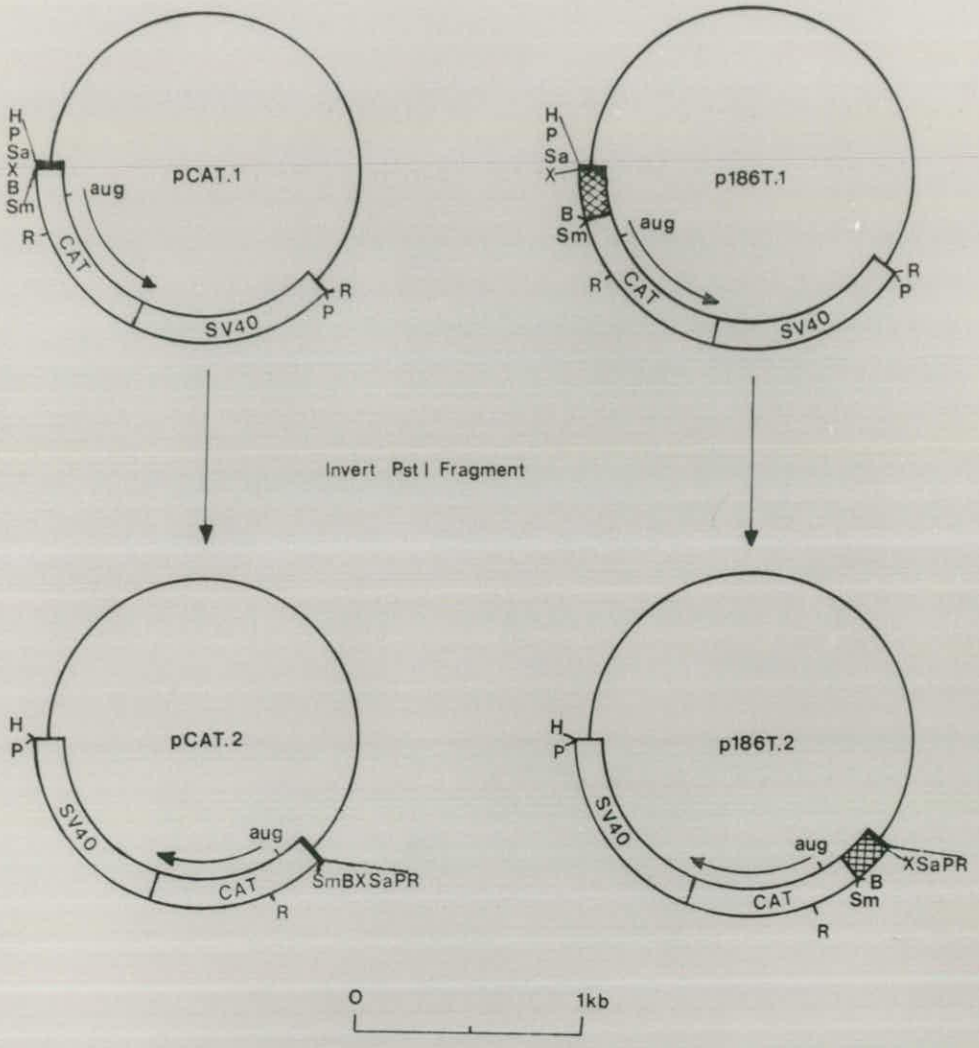
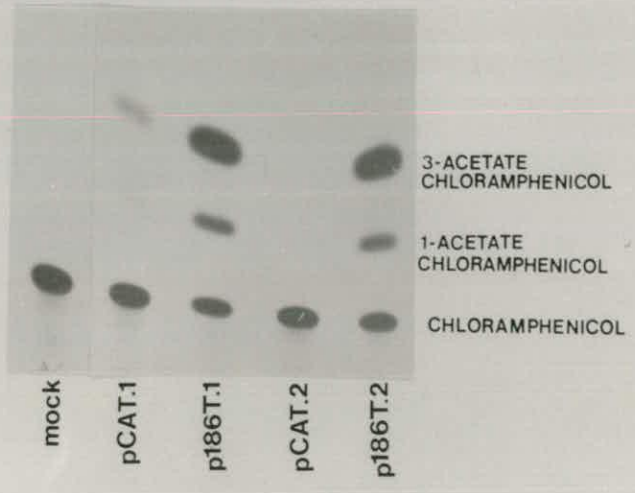


Figure 3.3b



3.4.3 Significance of the Polymorphism in the *I* Factor Promoter:

The 5' end of all *I* factors sequenced to date are highly conserved with only a single nucleotide difference at position +3 which can be a T or a G. To test the effect, if any of this polymorphism, p186G was constructed, which is identical to p186T.1 except nucleotide +3 is G instead of T. Table 3.1 shows p186G directs on average 60% more CAT activity than p186T.1 although the standard deviation is high. A G at position +3 might therefore confers greater promoter strength in tissue culture. The same may be true in flies as eight of the 11 elements sequenced at their 5' end have a G at this position (Bucheton, 1990). The higher copy number may be a reflection of elevated transcription levels and consequently a greater transposition rate of these elements. It is therefore conceivable that the identity of position +3 mediates or contributes to the intrinsic ability of different *I* factors, to induce chromosome contamination to varying degrees (Pélisson, 1978).

TABLE 3.1:

CONSTRUCT	MEAN % CAT ACTIVITY	n	σ(n-1)
p186T.1	100	12	-
p186G	161	4	64
pCAT.1	11	11	3

n = the number of transfections. Transfections were done in duplicate in each experiment, therefore the number of experiments = n/2.

3.4.4: 3' Deletion Analysis:

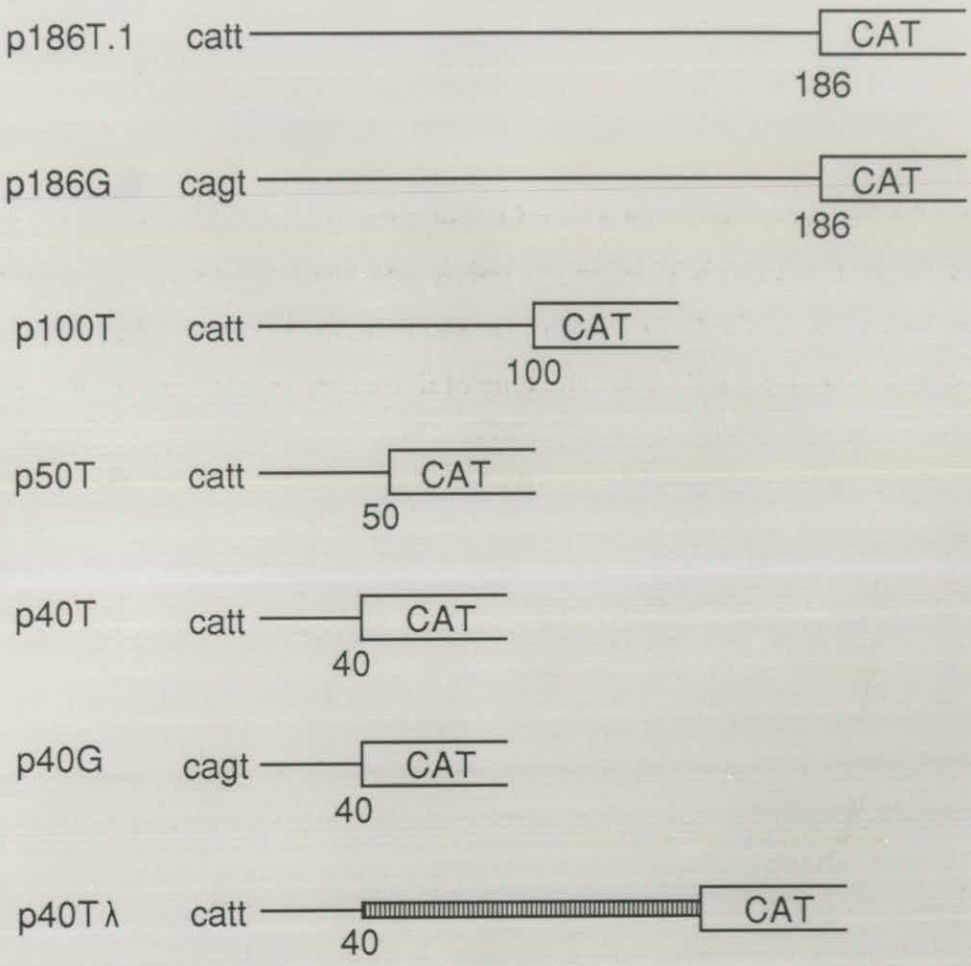
The previous results demonstrate that nucleotides 1 to 186 of the *I* factor have promoter activity. In an attempt to define the 3' limit of this promoter, several deletion derivatives were constructed and tested for their ability to direct CAT activity

in transient assays. The fragments assayed for promoter activity and CAT assays of a selection of these are illustrated in figure 3.4 with an autoradiograph showing the CAT activity of some of these constructs. Constructs were named according to their 3' end point and the identity of nucleotide +3. Table 3.2 shows deletion of sequences to position +100 has very little affect on CAT activity. However deletion to +50 increases activity by around 50%, and removal of the next 10 nucleotides to +40 increases activity to twice that of p186T.1. This increase is apparent whether position +3 is a T or a G. These results suggest sequences between position +40 and +100 act to suppress promoter activity. However it is not known if the CAT activity is an accurate reflection of promoter strength or transcript levels as no RNA analysis was done. The increased CAT activity may therefore be due to a translational affect, or RNA stability. One can imagine several ways by which removal of sequences could influence these parameters e.g. sequences between +40 and +100 may have RNA secondary structure potential that results in masking of the CAT methionine codon, or sequences that signal RNA degradation. Alternatively, altering the spacing of the cryptic puc18 promoter from the CAT gene might increase its strength, or the promoter/CAT junction in p40T and p40G may enhance the transcription rate, RNA stability and/or translation rate. The last two possibilities are unlikely because p40Tλ, which is identical to p40T except a 118bp λcI857 *Sau*3A fragment has been inserted into the *Bam*HI (see appendix B for sequence), has CAT activity almost identical to p40T (see table 3.2 and figure 3.4).

Because removal of sequences between position +40 and +50 produce the greatest increase in CAT activity, whether this be at the transcription or post transcription level, one might expect this region to play an similar role in flies. If so one would predict this region to be conserved between *I* factors of different species. Comparison of the first 50bp in the *I* factor of *D. melanogaster* and *D. teissieri* (Abad *et al.*, 1989) reveals only five base differences - three of which lie between position +40 and +50 rendering this 10bp stretch only 70% homologous, compared with 82.8% homology over the entire 5'UTR. This suggests sequences +41 to +50 do not play an important functional role in flies.

Figure 3.4 : **3' deletion analysis of the *I* factor promoter.** The line drawings show the extent of the various deletion endpoints and the identity of the first four nucleotides are given. λcI857 sequences are shown as the vertical stripped box. The autoradiograph shows the CAT activity only of those constructs fragments that create the greatest effect on promoter activity.

Figure 3.4


p186T.1
catt
CAT
186
p186G
cagt
CAT
186
p100T
catt
CAT
100
p50T
catt
CAT
50
p40T
catt
CAT
40
p40G
cagt
CAT
40
p40Tλ
catt
CAT
40

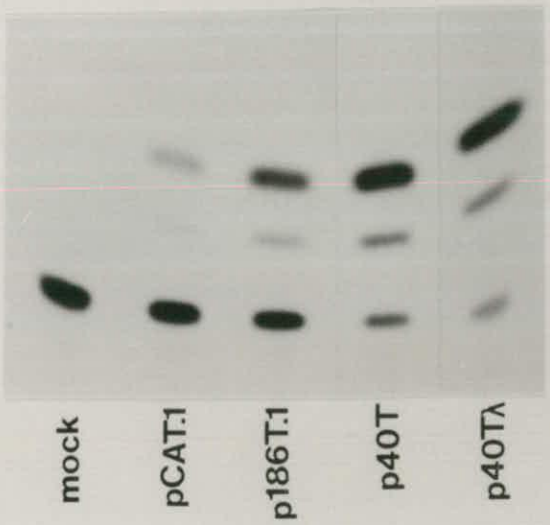

mock
pCAT:1
p186T:1
p40T
p40Tλ

TABLE 3.2:

CONSTRUCT	MEAN % CAT ACTIVITY	n	σ(n-1)
p186T.1	100	12	-
p100T	112	4	21
p50T	155	4	24
p40T	217	4	81
p40G	397	4	155
p40Tλ	226	4	35
pCAT.1	11	12	3

3.4.5 Comparison of the *I* Factor Promoter with TATA-less promoters and LINEs:

3' deletion analysis established that the first 40 nucleotides of *I* are sufficient for promoter activity in *Drosophila* Schneider Line 2 cells. This, in combination with the observation that nucleotides 1 to 37 displaying only one difference between *D. melanogaster* and *D. teissieri* (97.3% homologous compared to 79.2% for position +38 to +186) (Abad *et al.*, 1989) suggest that this region contains motifs involved in transcription initiation. It is possible that such motifs are shared with other promoters, in particular those which lack TATA or GC rich sequences (class II.3), and promoters of other LINE-like elements. Figure 3.5a shows a comparison of nucleotides +1 to +40 of *I* with the corresponding region in other genes with this type of promoter. Two conserved motifs exist - one at the transcription initiation site or start of the element with the consensus CAG/TT, and one centred around position +30 with the consensus CGTG that can be expanded in most cases to AGACGTGPyPyT. This sequence has not been noted before in transcription factor binding site compilations

(Wingender, 1988; Biggin and Tjian, 1989). The more proximal location of this motif in *R1Dm* is probably because this element is a 5' truncated copy which would also explain the absence of the CAG/TT motif. *F* has in addition to the +30 consensus, two other very similar flanking motifs (underlined in figure 3.5a).

The consensus CAG/TT is found at or near the transcription start site in several other class II.3 and TATA containing promoters (Arkhipova and Ilyin, 1991; Hultmark *et al.*, 1986), with transcription often initiating at the C. The motif is usually preceded by a T. In eight of the 13 *I* factors sequenced at their 5' end the nucleotide at the 5' insertion site is a T (Bucheton, 1990; Jensen and Heidmann, 1991). This apparent target site preference may be because TCAG/TT confers greater promoter strength and consequently a higher transposition frequency. However the *I* factor does not depend on insertion downstream of a T for promoter activity as the functional promoters used in this thesis are preceded by an A.

The conserved sequence at position +30 is found exclusively in classII.3 promoters. This element in *engrailed* (*en*) has previously been noticed by Kassis *et al.* (1989) by virtue of its conservation in the *D. melanogaster* and *D. virilus en* genes. Comparison of classII.3 promoters lacking the AGACGTGPyPyT motif identified the consensus ACAG again situated approximately 30 nucleotides downstream of the initiation site (Arkhipova and Ilyin, 1991) (figure 3.5b). This consensus is found particularly in retrovirus-like transposons of *D. melanogaster* and some classII.3 developmentally regulated genes. The *E74* gene of *D. melanogaster* and *Ddc* of both *D. melanogaster* and *D. virilus* have TATA like sequences at an appropriate upstream position. However removal of this sequence has no affect on expression *in vitro* for *E74*, or *in vivo* for *D. melanogaster Ddc* (Thummel, 1988; Hirsh *et al.*, 1986).

Conservation of both the sequence an location of motifs between genes within, and across species strongly suggests functional significance. Their importance in *I* factor promoter activity in transient tissue culture assays was therefore tested.

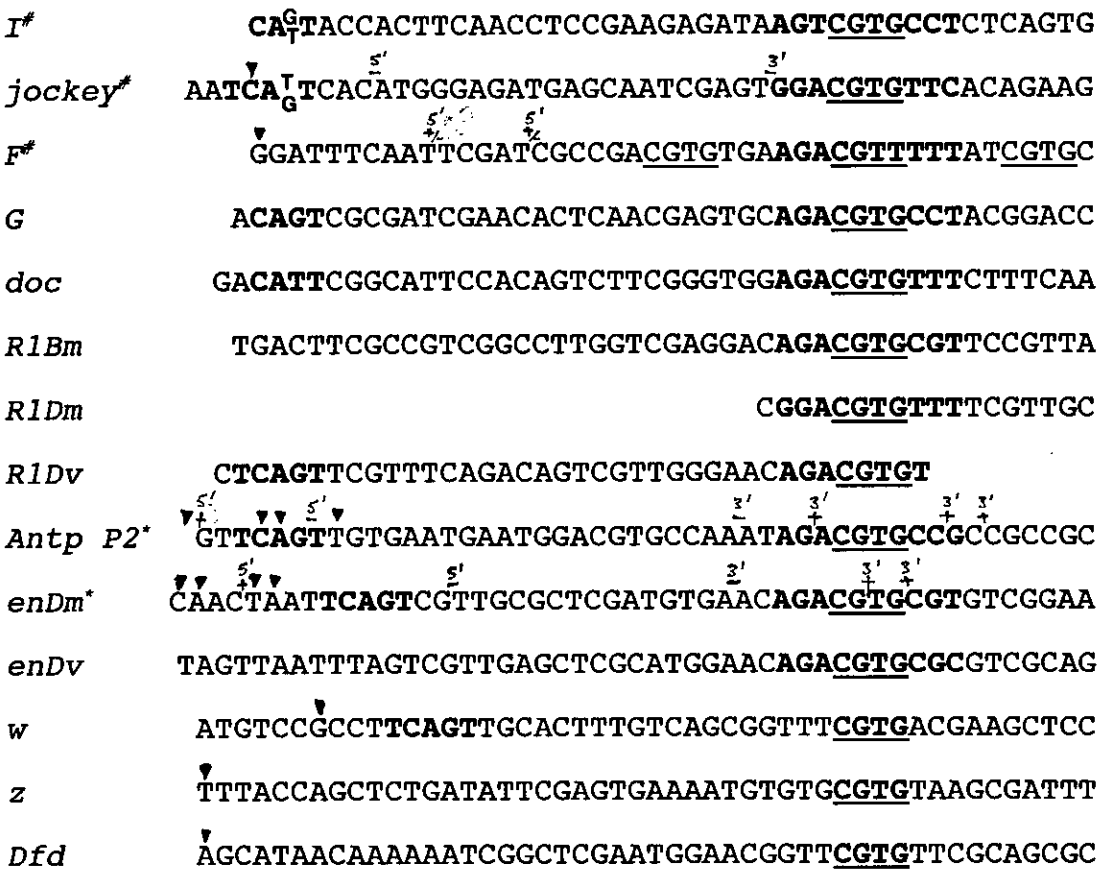


Figure 3.5a: Comparison of the first 45 or so nucleotides of some LINE-like elements and class II.3 promoters. The conserved start site and +30 motifs are shown in bold typeface, and the +30 core sequence is underlined. + and - symbols denote limits that have been tested for promoter function with + signifying promoter function and -, loss of promoter activity; 5' and 3'indicates from which direction these deletions were made. # means that the assay was done *in vivo*, and *, *in vitro*. ▼ is the transcription initiation site(s). The references are: *I*, Fawcett *et al.*, 1986; *jockey*, Priimagi *et al.*, 1988; *F*, Minchiotti and Di Nocera 1991; *G*, Di Nocera, 1988; *doc*, Driver *et al.*, 1989; *R1Bm*, Xiong and Eickbush, 1988b; *R1Dm*, Jakubczak *et al.*, 1990; *R1Dv* (*D. virilis*), Rae *et al.*, 1980; *Antp P2*, Perkins *et al.*, 1988; *enDm*, Soeller *et al.*, 1988; *enDv*, Kassis *et al.*, 1989; *w*, O'Hare *et al.*, 1984; *z*, Pirrotta *et al.*, 1987; *Dfd*, Regulski *et al.*, 1987.

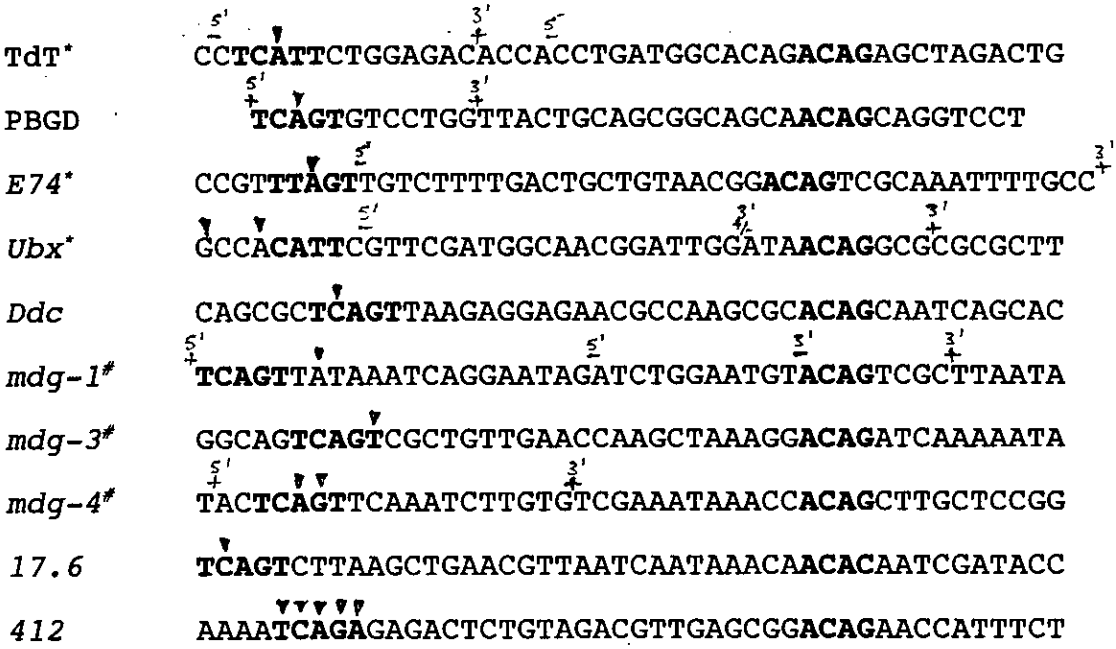


Figure 3.5b: Sequence comparison of classII.3 promoter with the +30 ACAG motif (adapted from Arkhipova and Ilyin, 1991). Both the +30 motif and the start site consensus are shown in bold typeface. Symbols are as described in figure 3.5a legend. References are: TdT, Smale and Baltimore, 1988; PBGD, Chreitien *et al.*, 1988, Beaupain *et al.*, 1990; *E74*, Thummel, 1989; *Ubx*, Biggin and Tjian, 1988; *Ddc*, Hirsh *et al.*, 1986; *mdg-1*, and *mdg-3*, Arkhipova *et al.*, 1986, Arkhipova and Ilyin, 1991; *mdg-4*, Jarrel and Meselson, 1991; *17.6*, Inouye *et al.*, 1986; *412*, Yuki *et al.*, 1986.

3.4.6 Mutational Analysis of the CAG/TT Motif:

Three mutants, illustrated in figure 3.6a, were assayed to determine the importance of the CAG/TT motif. Constructs p186Δ4.2 and p40Δ4 have nucleotides 1 to 4 of the *I* factor replaced by ACCG, effectively deleting the first four bases. Table 3.3 and figure 3.6b shows that both constructs direct CAT activity at a similar level to that in pCAT.1 indicating that the sequence CAG/TT is essential for promoter activity. This agrees with deletion analysis of other promoters of this type (see figure 3.5a and 3.5b). A third promoter, p186Δ4.1, has the first four nucleotides replaced by TAGA, effectively substituting T and G for C and T at positions 1 and 4 respectively. CAT activity of p186Δ4.1 is reduced by only about one third relative to p186G (table 3.3). Positions 1 and 4 can therefore be replace by T and G without severe reduction in promoter strength. Detailed mutational analysis of the CAG/T motif has been done only for the murine TdT and human PBGD genes in which it is a central component of their initiator (Smale and Baltimore, 1989; Smale *et al.*, 1990; Beaupain *et al.*,1990). Point mutation of the C to a G in PBGD alters accuracy but not frequency of transcription initiation, whereas in TdT a C to G transversion moves the initiation site one position upstream and reduces the initiation frequency slightly. Substitution of a C for T at position 4 has no affect on accuracy or strength of initiation by the TdT initiator. In contrast mutation of the A to a T affects accuracy of initiation in both TdT and PBGD, and severely reduces the level of transcription in TdT. However position 3 can be a C, T, or A with no effect. Taken together, these analyses demonstrate the A in CAG/TT to be the most important base, with C second, and the G/TT bases last. One would predict that mutation of this motif in *I* to TAGA would reduce transcription frequency slightly and alter accuracy of start site selection. The CAT activity results are consistent with slight reduction in initiation frequency. RNA analysis is lacking therefore no conclusions can be drawn regarding the effect on initiation accuracy.

Three other 5' deletion constructs, p186Δ17, p186Δ38, and p186Δ100, which lack the first 17, 38, and 100bp of *I* respectively, were assayed for promoter activity in tissue culture (figure 3.6a). Table 3.3 and figure 3.6b show, not surprisingly, that

none of them direct CAT activity greater than pCAT.1.

TABLE 3.3:

CONSTRUCT	MEAN % CAT ACTIVITY	n	σ(n-1)
p186T.1	100	12	-
p186Δ4.2	6	4	3
p40Δ4	7	4	5
p186Δ4.1	126	2	8
p186Δ17	6	3	5
p186Δ38	1	1	-
p186Δ100	3	1	-
pCAT.1	11	12	3

Figure 3.6a : **5' deletion analysis of the *I* factor promoter.** The line drawings indicate the extent of the deletions, p186G and p40G are included for comparison. The identity of the first four nucleotides are indicated where necessary.

Figure 3.6a

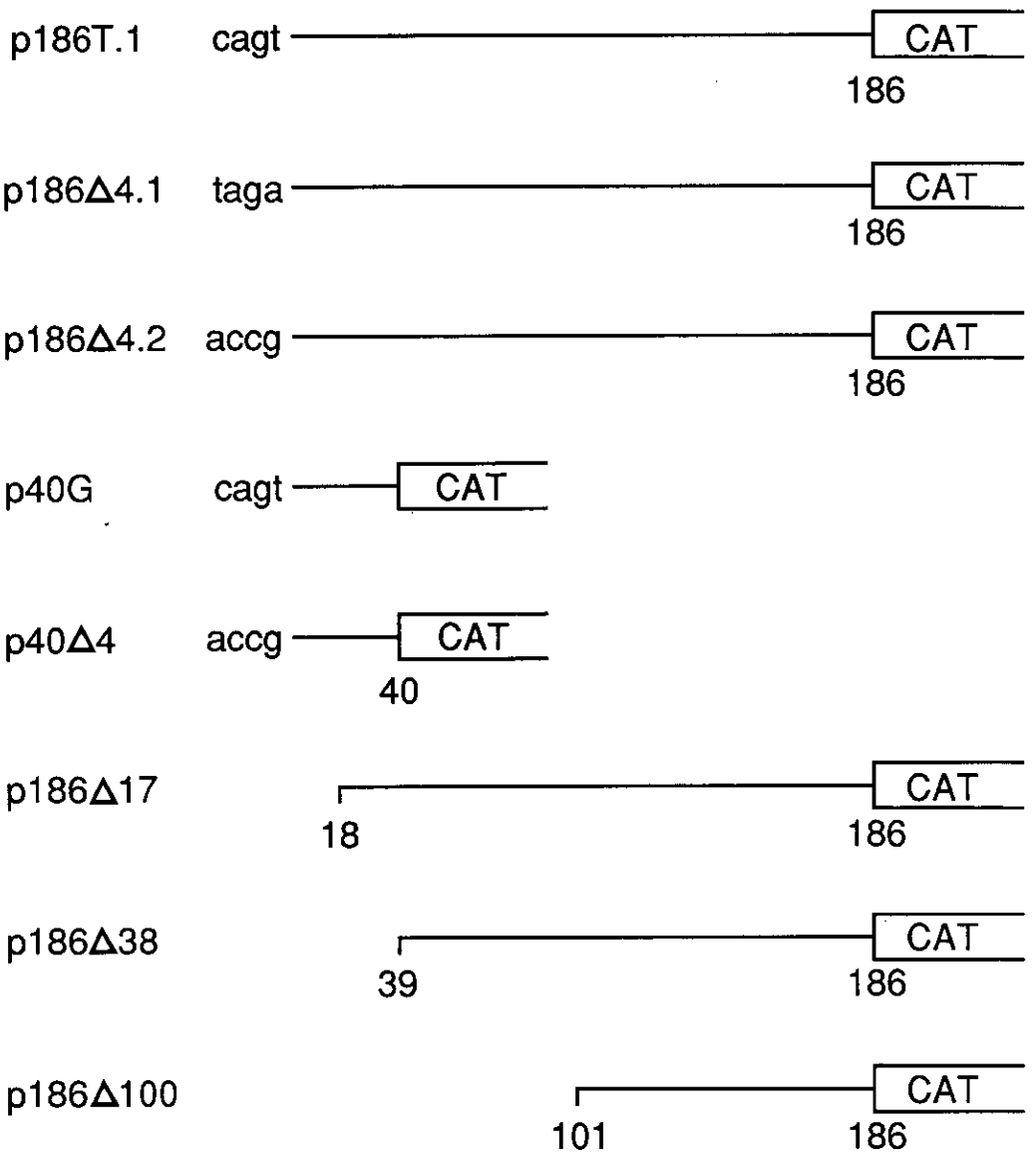

p186T.1
cagt
CAT
186
p186Δ4.1
taga
CAT
186
p186Δ4.2
accg
CAT
186
p40G
cagt
CAT
p40Δ4
accg
CAT
40
p186Δ17
CAT
18
186
p186Δ38
CAT
39
186
p186Δ100
CAT
101
186

Figure 3.6b : **5' deletion analysis of the *I* factor promoter.** CAT assays of various intact and 5' deleted promoters are shown. The difference in intensity of the two autoradiographs is due to variation in exposure time.

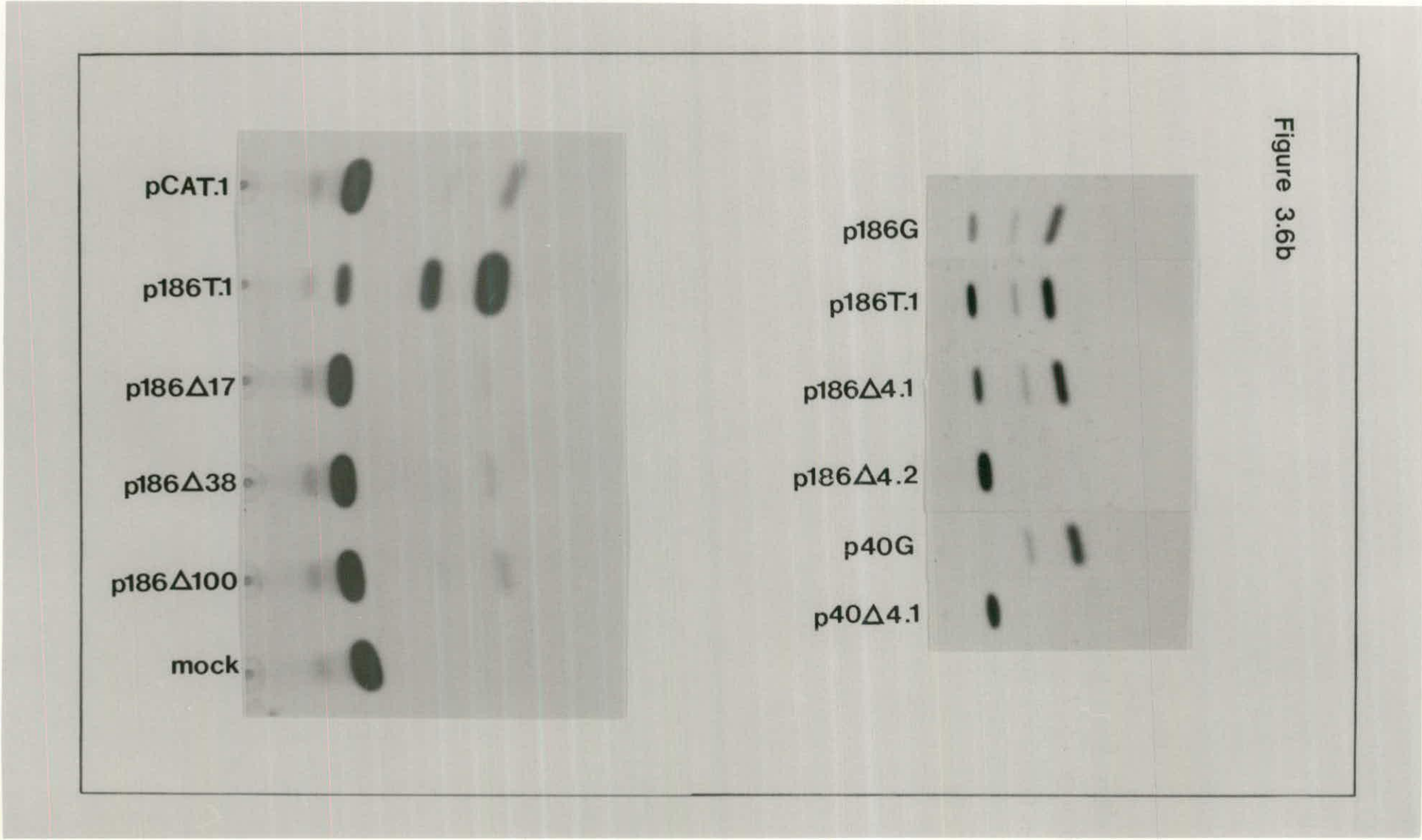

pCAT.1
p186T.1
p186Δ17
p186Δ38
p186Δ100
mock
p186G
p186T.1
p186Δ4.1
p186Δ4.2
p40G
p40Δ4.1

Figure 3.6b

3.4.7 Mutational Analysis of the +30 Motif:

Five mutant promoter constructs were tested to elucidate the importance of the AGTCGTGPyPy motif in *I*. The promoter fragments are depicted in figure 3.7a with CAT assays of each in figure 3.7b and their respective CAT activity values in table 3.4. The three mutants with a single base change have little or no affect on CAT activity. This was most surprising for construct p186T.34C which converts the T in the core consensus CGTG to a C. However, substitution of positions +30 to +37 from GTCGTGCC to CCTCGAGG in p186T#8 decreases CAT activity only 1.7 fold, therefore it is not surprising that the point mutations have very little affect on promoter activity. p28T directs CAT activity 2.4 fold lower than p40T, therefore removal of the conserved motif in p28T has a greater affect than the substitution in p186T#8. A contributing factor to this may be that the substitution in p186T#8 includes the sequence TCGAG, very similar to bases +31 to +35 (TCGTG, that include the core consensus), but displaced one position downstream. The DNA sequence immediately downstream of the promoter fragment in p28T bears no resemblance to the consensus. Accordingly, p28T should be representative of the function of the +30 motif i.e. it is a positively acting, but non essential promoter element in tissue culture. Retention of promoter activity despite removal of this sequence was somewhat surprising in light of deletion analysis of this motif in the *D. melanogaster Antp P2* and *en* promoters (Perkins *et al.*, 1988; Soeller *et al.*, 1988). Complete and partial removal of the core consensus, CGTG from *Antp* and *en* respectively, reduces efficiency but not accuracy of initiation. However removal of a few more 5' bases abolishes transcription *in vitro* (see figure 3.5a for deletion limits) suggesting the core may not be essential for transcription but sequences immediately upstream are. *Jockey* also has this consensus, deletion of which abolishes transcription as measured by CAT activity in transient tissue culture assays (Mizrokhi and Mazo, 1990). Why then does the *I* factor promoter retain high levels of activity in the absence of this motif? Two possibilities exist:

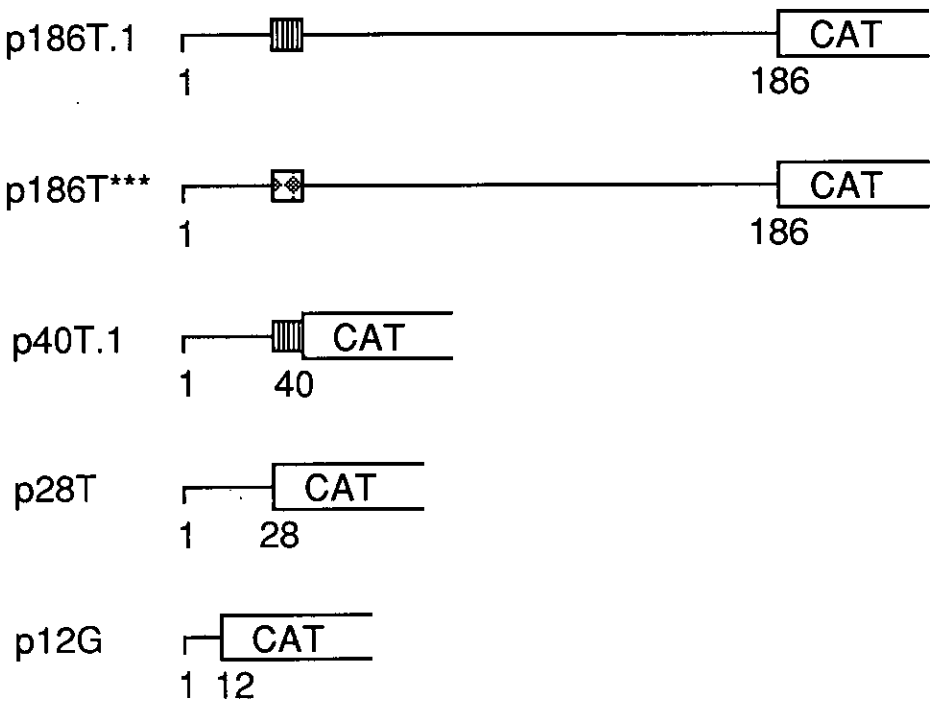
1. The CAT activity of p28T is directed entirely from the cryptic promoter within pCAT.1 - the elevated activity in p28T being a

TABLE 3.4:

CONSTRUCT	MEAN % CAT ACTIVITY	n	σ(n-1)
p186T.1	100	12	-
p186T.29T	117	2	17
p186T.29C	81	2	26
p186T.34C	116	2	31
p186T#8	58	4	20
p28T	102	4	36
p12G	14	4	2
pCAT.1	11	12	3

Figure 3.7a : **Mutational analysis of the conserved +30 motif.** The line drawing illustrates each promoter fragment assayed. *** means that the motif at position +30 can be one of the four substitutions below. The precise substitution in each of the mutations is shown in bold.

Figure 3.7a



29 37

▨ – AGTCGTGCC

▨ – **T**GTCGTGCC for p186T.29T

– **C**GTCGTGCC for p186T.29C

– AGTCG**C**GCC for p186T.34C

– A**CCTCGAGG** for p186T#8

Figure 3.7b : **Mutational analysis of the conserved +30 motif.** CAT assays are shown for each mutant promoter fragment and the appropriate wildtype promoter. Slight variation in apparent CAT activity relative to the values given in table 3.4 are due to differences in transfection efficiency.

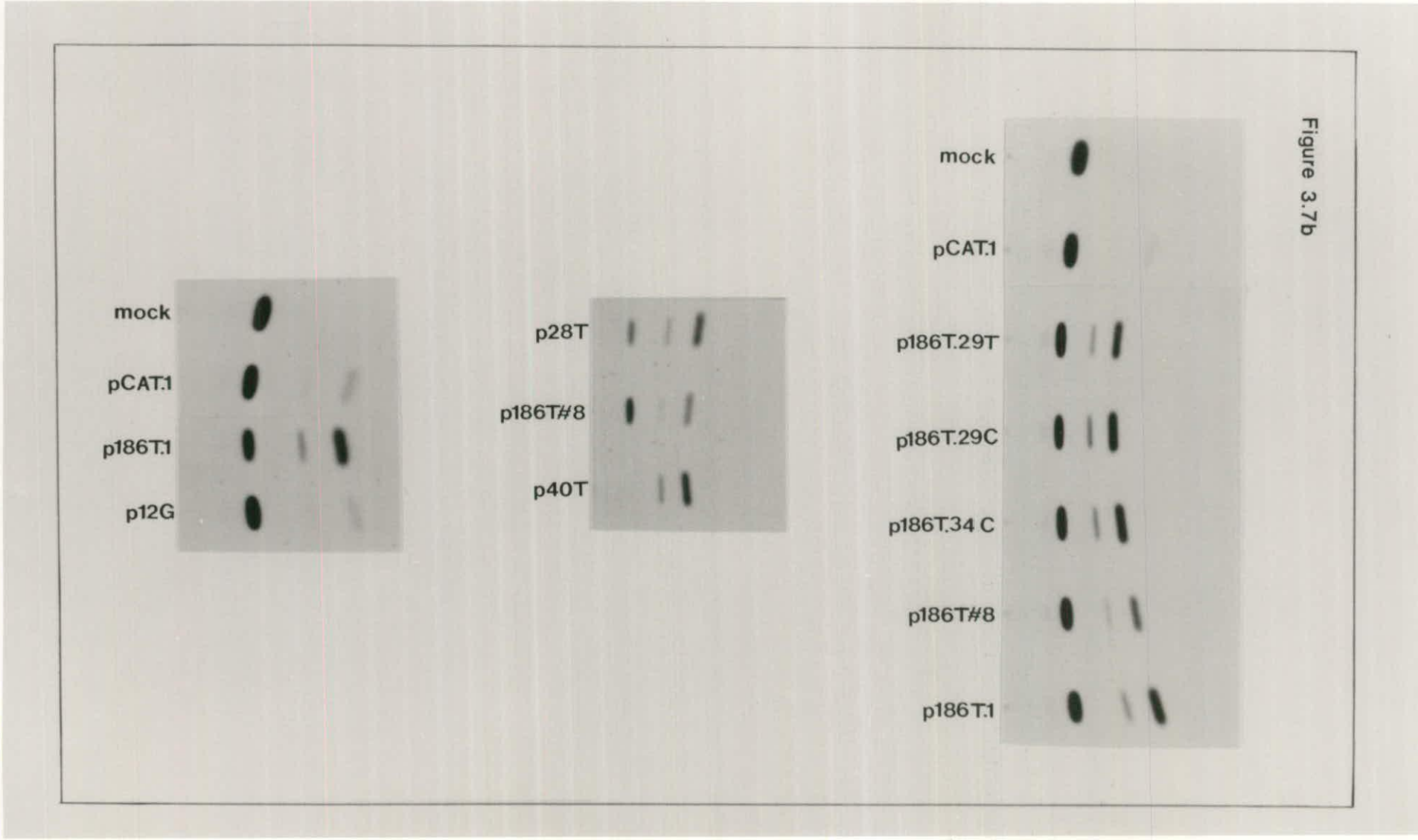

mock
pCAT.1
p186T.1
p12G
p28T
p186T#8
p40T
mock
pCAT.1
p186T.29T
p186T.29C
p186T.34 C
p186T#8
p186T.1

Figure 3.7b

consequence of altered spacing between the cryptic promoter and the CAT gene.

2. Not all promoters of this class require the +30 element.

Possibility 1. is under test by inverting the promoter/CAT *PstI* fragment of p28T. The second possibility is supported by mutational analysis of the +30 motif ACAG. When deleted from the *D. melanogaster E74, Ubx*, and *mdg-1* promoters transcription is abolished *in vitro* for *E74* and *Ubx*, and in transient transcription assays for *mdg-1* (Thummel, 1989; Biggin and Tjian, 1988; Arkhipova and Ilyin, 1991). However deletion of this motif and sequences to position +13 in the *mdg-4* promoter reduces promoter strength four fold (as measured by CAT activity from transient assays) but does not abolish activity (Jarrel and Meselson, 1991). In addition deletion of the ACAG motif and sequences to positions +10 and +11 does not alter transcription from the PBGD and TdT initiators respectively *in vitro* (Beaupain *et al.*, 1990; Smale and Baltimore, 1989). This type of internal motif is therefore essential in some, but not all promoters. The same may be true of the CGTG consensus.

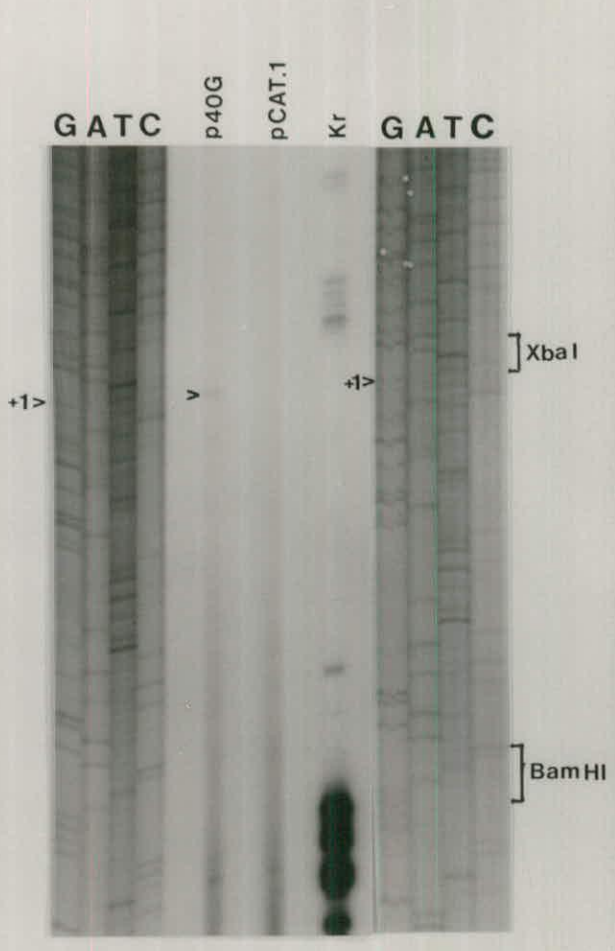
A further promoter construct, p12G which has bases +1 to +12 of the *I* factor was tested for CAT activity. Table 3.4 shows its activity to be similar to that of pCAT.1. Assuming possibility 2. above to be correct, the 3' limit of the *I* factor promoter lies between position +12 and +28.

3.4.8 Transcription of the *I* Factor α-amanatin sensitive and Initiates at Nucleotide 1:

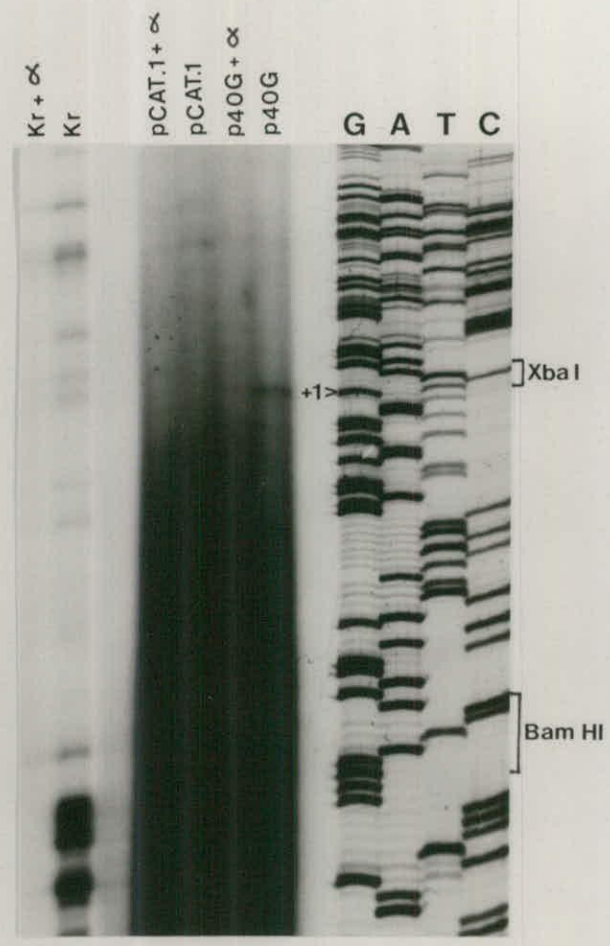
Chaboissier *et al.* (1990) demonstrated the 5' end of the *I* factor RNA postulated to be the transposition intermediate is at a single site at, or very near, position +1. In flies this transcript is most likely initiated from an internal promoter. To support the idea that the internal promoter activity detected in tissue culture is that used in flies, and determine the precise initiation site, primer extension analysis was done on transcripts synthesised *in vitro*, using a *Drosophila* embryo transcription extract. The template chosen for *in vitro* transcription was p40G because it produces the highest level of CAT activity in transient assays. The primer extension analysis

Figure 3.8 : Figure 3.8a shows primer extension analysis of transcripts initiating from p40G in a *Drosophila in vitro* transcription extract. The transcription start site maps to position +1 that lies immediately downstream of the *Xba*I cloning site indicated on the markers. The size markers are plasmid sequencing products of p40G primed from the same oligonucleotide as was used for primer extension (oligonucleotide 332G - see materials and methods). pCAT.1 was used as a negative control, and a construct containing the *Kruppel* (Kr) promoter (Stratagene) as a positive control. *Kr* transcription initiates at several sites over a 10bp region (Kamakaka *et al.*, 1991). The bands seen in the *Kr* lane are of the expected size (70-80bp).

Figure 3.8b illustrates the same reactions and parallel reactions that include 4μg/ml α-amanatin (indicated as +α). The size markers are sequence products of an M13 clone that contains the *Xba*I-*Eco*RI (promoter/CAT 5' end) fragment of p40T. The primer for both primer extension of p40G and sequence analysis was 332G.


G A T C
p40G
pCAT.1
Kr
G A T C
+1>
>
+1>
Xba I
Bam HI

3.8a


Kr + ∝
Kr
pCAT.1 + ∝
pCAT.1
p40G + ∝
p40G
G A T C
+1>
Xba I
Bam HI

3.8b

is presented in figure 3.8a and shows transcription initiates homogenously at position +1. Figure 3.8b shows synthesis of these transcripts is sensitive to α-amanatin at a concentration of 4μg/ml indicative of a class II promoter. Inclusion of a class III promoter in this experiment would be conclusive proof that the polymerase catalysing *I* transcription is RNA polymerase II. These results demonstrate all sequences for accurate initiation lie within the first 40bp of *I*.

3.4.9 Promoter Activity is Influenced by Upstream Sequences:

Pélisson and Bregliano (1981) noticed that the inducer strength of some *I* factors can increase following transposition to a new chromosomal location. The increase in inducer strength may be a reflection of higher transcription activity. The *I* factor promoter may therefore be influenced by its chromosomal environment. To test if the upstream DNA sequence can affect promoter strength constructs p186G.*w* and p40G.*w* were tested for their ability to direct CAT activity (see figure 3.9). These constructs include 99bp of *white* DNA found upstream of the *I* factors in and W^{IR3} and are included in plasmid pI954. Table 3.5 shows CAT activity is several fold lower in p186G.*w* and p40G.*w* relative to equivalent constructs lacking these sequences (p186G and p40G respectively). Despite the apparent suppression, promoter strength of *I* factors with the same *white* sequences as in p186G.*w* must be sufficient for transposition as flies harbouring the pI954 transgene evolve into strong inducer strains (Pritchard *et al.*, 1988).

TABLE 3.5:

CONSTRUCT	MEAN % CAT ACTIVITY	n	σ(n-1)
p186T.1	100	12	-
p186G	161	4	64
p186G.*w*	34	4	16
p40G	397	4	155
p40G.*w*	47	2	13
pCAT.1	11	12	3

Figure 3.9 : An illustration of the promoter fragments used to test the influence of upstream sequences. p186G and p40G are included for comparison.

Figure 3.10 : The promoter region of the constructs tested is shown above a CAT assay of the results. The trinucleotide CTG is written at the end of the promoter to emphasise that the fragment inverted starts with CAG rather than CAT (the inverse complement of which is a methionine codon).

Figure 3.9

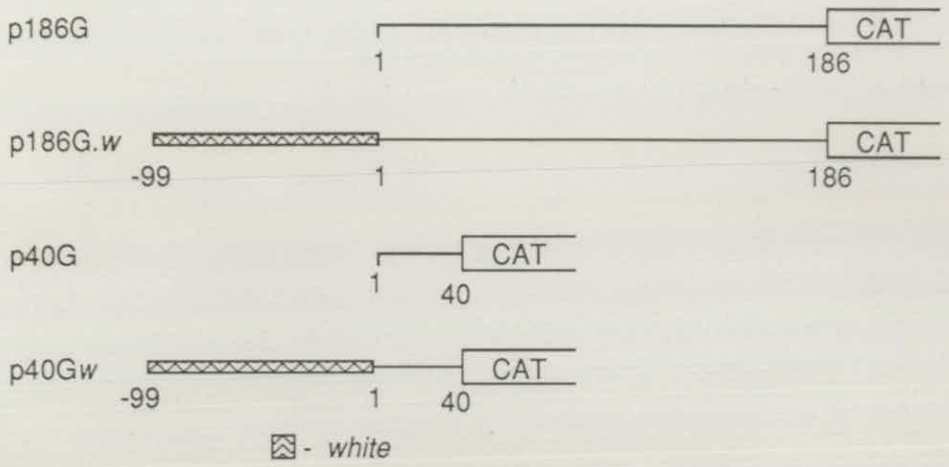
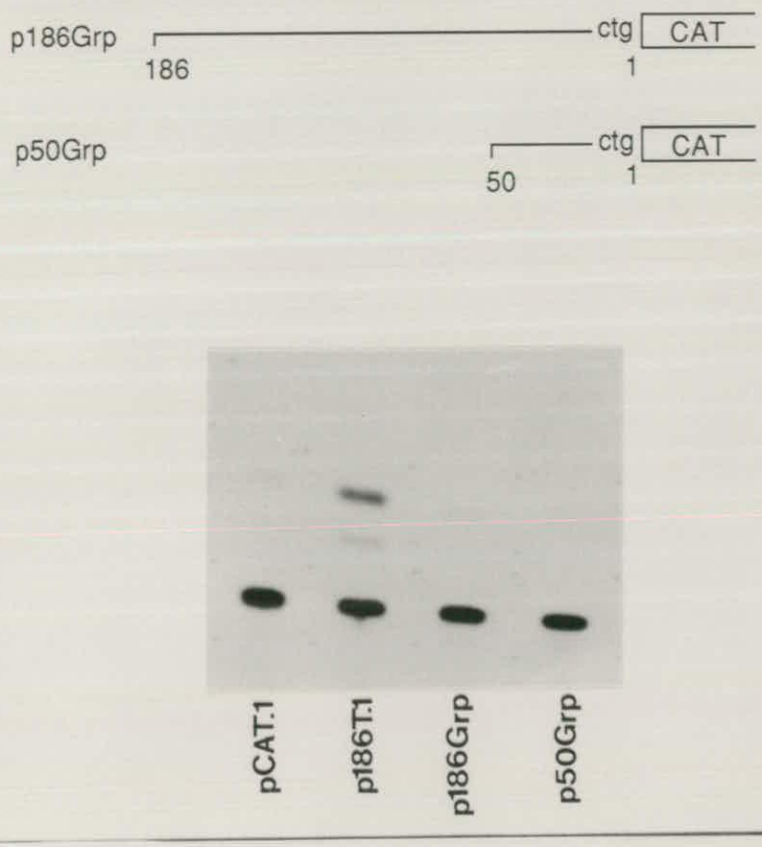


Figure 3.10



3.4.10 The *I* Factor 5'UTR does not Contain Convergent Promoters:

It was of interest to test whether the 5'UTR of the *I* factor contains oppositely orientated promoters as synthesis of transcripts complementary to the 5' end of the transposition intermediate may function in regulation by, for example, inhibition of translation or reverse transcription. Indeed, Minchiotti and Di Nocera (1991) have reported that the 5' end of *F* can direct convergent transcription from oppositely orientated promoters. Two constructs, illustrated in figure 3.10, were made to test bidirectionality if the promoter. Neither *I* factor fragment under assay included the trinucleotide CAT, which when reversed becomes ATG that could act as an initiation codon and produce a negative result if out of frame with the CAT gene even if the promoter were active. The results presented in figure 3.10 and table 3.6 show that neither construct has CAT activity making bidirectionality of the promoter unlikely.

TABLE 3.6:

CONSTRUCT	MEAN % CAT ACTIVITY	n	σ(n-1)
p186T.1	100	12	-
p186Grp	1	4	1
p50Grp	8	2	0
pCAT.1	11	12	3

3.4.11 The *I* Factor Promoter is Not Active in Mammalian Cells:

Constructs p186T.1, p40G, and pCAT.1 were tested for promoter activity in the chinese hamster lung fibroblast cells RJK88 (Fuscoe *et al.*, 1983). All three constructs direct similar levels of low CAT activity indicating weak promoter activity from vector sequences and none, or very little from *I*. The values for CAT activity are presented in table 3.7 and are expressed as a percentage of the CAT activity directed from the SV40 early promoter in pSV2CAT, after correction to transfection efficiency measured by cotransfection with pCMVβgal.

TABLE 3.7:

CONSTRUCT	MEAN % CAT ACTIVITY	n	σ(n-1)
pSV2CAT	100	4	-
p186T.1	2	4	1
p40G	8	4	6
pCAT.1	5	4	3

3.5 SUMMARY and DISCUSSION:

This chapter has established that the *I* factor has a promoter that is active in Schneider Line 2 tissue culture cells. It lies within the first 186 nucleotides of *I* and contains sequences that are essential, stimulate and suppress CAT activity. Transcripts initiate at position +1 and are sensitive to α-amanatin, consistent with RNA polymerase II catalysed transcription of an promoter that is entirely internal. The 5' limit of the promoter is at position +2 and the 3' limit is between positions +12 and +28, as measured by CAT activity. However it is not known whether these end points permit accurate transcription initiation. These results categorise the *I* factor promoter with TATA-less, non GC-rich promoters i.e. classII.3. This classification is further

supported by conservation, in sequence and location, of DNA motifs which play a positive role in promoter activity.

The mechanism of transcription initiation from classII.3 promoters has not yet been elucidated. However work has been done to determine which sequences and protein factors are involved. Proteins have been identified that bind near the initiator of PBGD and across the transcription start sites of *en*. The PBGD initiator binding factor is present in erythroid and non erythroid nuclei. However it is unlikely to be a general initiator factor as it requires sequences downstream of the TCANT core, and does not bind the TdT initiator (Beaupain *et al.* 1990). The *en* binding factor associates with the transcription start site (located upstream of the TCAGT sequence) (Soeller *et al.*, 1988) but is thought not to be RNA polymerase II as purified fractions of this protein do not bind this region. Carcamo *et al.* (1991) have however demonstrated that highly purified RNA polymerase II initiates transcription from a wildtype (but not a mutated Ad-IVa2 initiator) in the absence of other transcription factors, albeit very weakly and rather heterogeneously. Carcamo *et al.* concluded that RNA polymerase II recognises the initiator.

A protein (or two different proteins) present in *Drosophila* embryo extracts binds the +30 consensus of *Ubx* and *Antp* promoters. Association of this factor to template DNA may be essential for transcription initiation and/or activation of at least some members of classII.3 promoters.

As mentioned at the beginning of this chapter, binding of TFIID to the TATA box is the first step in formation of an active transcription initiation complex. The *I* factor promoter, and the majority of the other promoters of this type lack a TATA box, so is TFIID necessary for transcription initiation? Carcamo *et al.* (1990) have shown that TFIID binds the Ad-IVa2 promoter at a poor TATA consensus downstream of the initiation site on the opposite strand. Mutation of this motif decreases initiation frequency but not accuracy which is affected by mutation of its initiator element.

Not all classII.3 promoters have TATA like sequences at unconventional sites in the vicinity of the initiator. To test whether TFIID is required for basal transcription initiation from a genuine TATA less promoter, Smale *et al.* (1990) tested

whether transcription could be initiated from the TdT initiator in an *in vitro* transcription assay following mild heat treatment of the transcription extract (which disrupts TFIID activity). Transcription was eliminated by heating but failed to recover after a TFIID cell fraction was added suggesting a requirement for another heat sensitive transcription factor for initiator activity. In parallel experiments with a promoter comprising an initiator and upstream Sp1 binding sites, accurate initiation was recovered following addition of the TFIID fraction. This implies that the initiator sequence can intrinsically define the transcription start site in the absence of the heat sensitive initiator factor.

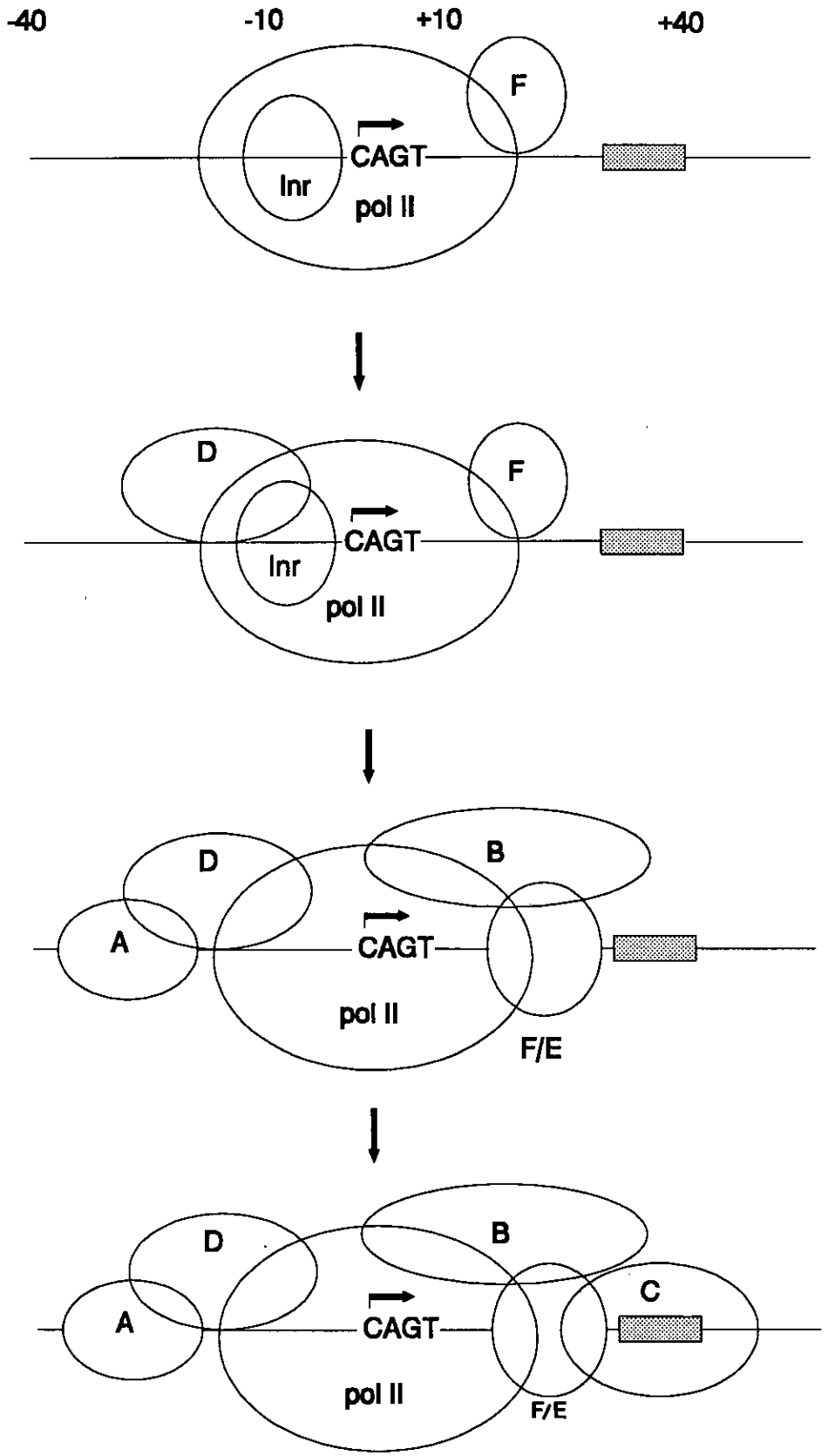
Carcamo *et al.* (1991) have demonstrated that TFIID is required for transcription from the Ad-MLP (which contains an initiator sequence) when the TATA box is mutated such that TFIID does not bind. Involvement of TFIID in transcription can therefore be independent of its consensus binding site. In such circumstances TFIID must acts through protein/protein interactions, or by binding non-TATA sequences, or to TATA-like sequences elsewhere in the gene. In all promoters tested to date with initiator sequences, it appears the initiator defines the start site and TFIID functions in transcription activation.

Formation of an transcription preinitiation complex at promoters that have only an initiator involves cooperative assembly of the general transcription factors TFIIB, -IID, and -IIF. These factor build on RNA polymerase II bound at the initiator, rather than TFIID bound at the TATA box (Carcamo *et al.*, 1991).

From the information discussed regarding the DNA sequences and *trans* acting factors that are involved in transcription initiation and activation of classII.3 promoters, a model for transcription initiation of the minimal *I* factor promoter can be predicted and is illustrated in figure 3.11. The model envisages a protein termed the Inr factor recognises and binds adjacent to the CAG/TT core, and facilitates strong association of RNA polymerase II (possibly associated with TFIIF) with the transcription start site. TFIID then associates with the complex by protein/protein or protein/non-TATA box DNA interactions. The remaining essential transcription

Figure 3.11 : **A model for assembly of a transcription initiation complex at class II.3 promoters.**

Figure 3.11


-40
-10
+10
+40
F
CAGT
Inr
pol II
D
F
CAGT
Inr
pol II
D
B
A
CAGT
pol II
F/E
D
B
C
A
CAGT
pol II
F/E

factors - TFIIA, -IIB, and -IIE then assemble to form the initiation complex. Factor C (the name given to the +30 *Antp P2* promoter binding factor (Perkins *et al.*, 1988) then binds the +30 motif and possibly activates transcription, or increases fidelity of start site selection by RNA polymerase II.

Isolation of the genes encoding *trans* acting factors used for transcription initiation of the *I* factor promoter may be possible by making use of mammalian cells, in which the *I* factor promoter is apparently inactive. Identification of those fractions from Schneider line 2 transcription extracts that facilitate transcription when added to a mammalian extract, followed by further purification of the active component, could lead to isolation of the genes by immunological or protein sequencing techniques. Alternatively, the genes could be isolated by transfecting a *Drosophila* expression library into a mammalian cell line in which the only copy of a selectable gene (e.g. HPRT) is transcribed from the *I* factor promoter. By enforcing selection, only those cells with an active *I* promoter would survive.

CHAPTER 4

4.1 INTRODUCTION:

The preceding chapter established the that first 186 nucleotides of the *I* factor contain a promoter that is active in Schneider line 2 tissue culture cells and in a *D. melanogaster* embryo nuclear extract. This chapter sets out to determine whether this promoter is active in flies, and if so what role it might play in the I-R dysgenic phenotype.

Chaboissier *et al.* (1990) provided evidence that *I* transposition is controlled at the level of transcription by demonstrating that full length transcripts of *I* are confined to the tissue in which transposition occurs, that is the ovaries of SF and RSF females. The amount of this transcript correlates with transposition frequency, and is highest in SF ovaries, lower in RSF ovaries, barely detectable in the ovaries of inducer females, and absent from female somatic tissues and males. Transcripts also become less abundant in the ovaries of SF flies as they age. Chaboissier *et al.* concluded that transposition is controlled at the level of transcription or RNA stability. If transcription dictates when and where transposition occurs, one would predict that the *I* factor promoter might contain sequences controlling the sex specificity, tissue specificity, developmental regulation, maternal effect, and sensitivity to suppression in inducer stocks.

To address the questions of whether the *I* factor promoter is active in flies and what aspects of transposition it controls, *I* promoter CAT fusions were introduced into flies by P-mediated transformation and the pattern of CAT expression recorded. This chapter deals with the function of the first 186 nucleotides in a reactive background.

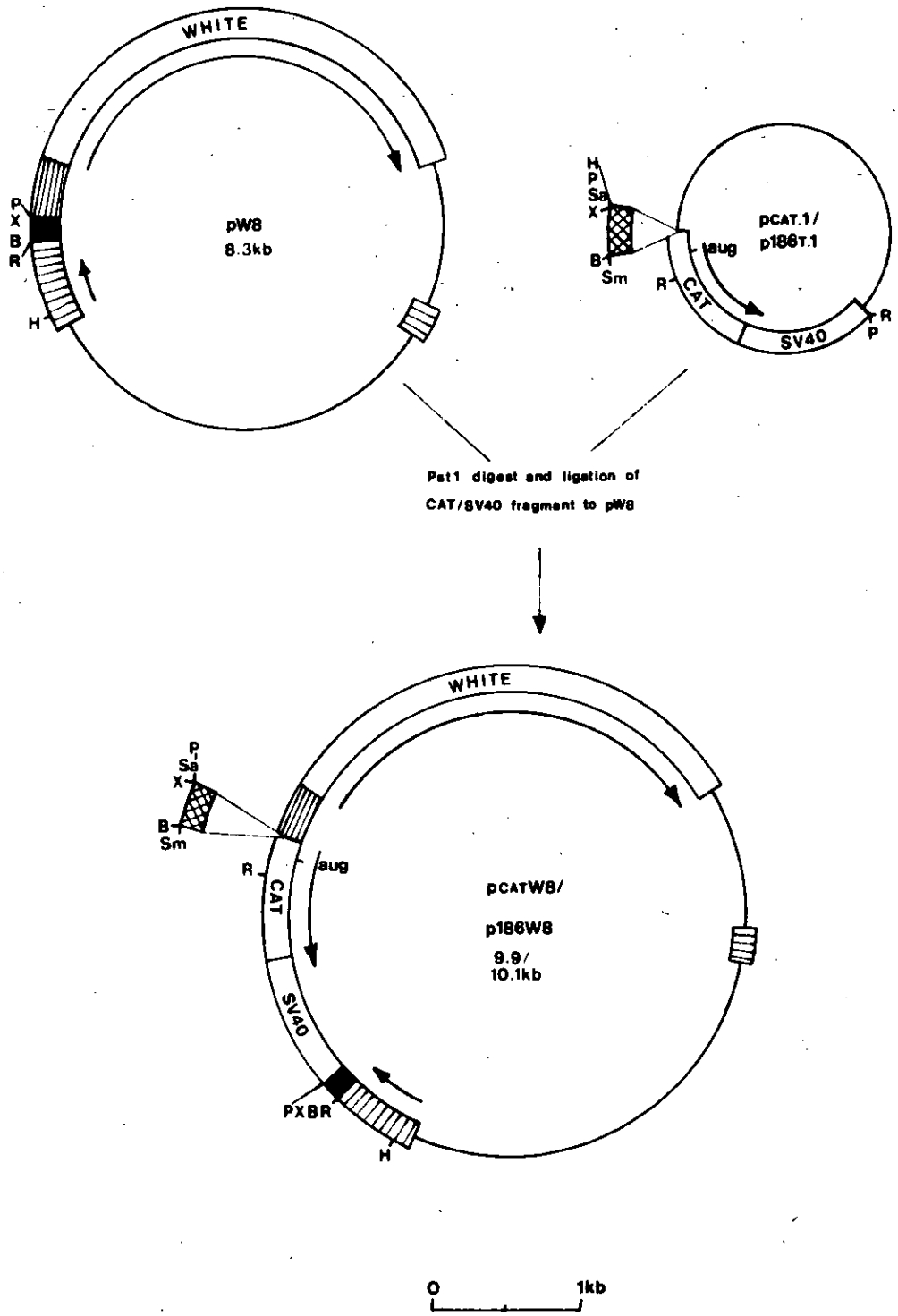
4.2 RESULTS:

4.2.1 Construction of Plasmids for P-Mediated Transformation:

The plasmids used for *I* promoter assay in flies (see figure 4.1) were constructed by inserting the *Pst*I fragment of pCAT.1 and p186T.1 containing the *I* promoter fragment (in p186T.1), CAT gene, and the SV40 small t intron and polyadenylation signals, into the *Pst*I site of pW8 (Klemenz *et al.*, 1987). This gave plasmids pCATW8 and p186W8 respectively. The orientation of their inserts was

Figure 4.1 : **Construction of plasmids for P-element transformation.** Abbreviations for restriction sites are: B - *Bam*HI; H - *Hind*III; P - *Pst*I; R - *Eco*RI; Sa - *Sal*I; Sm - *Sma*I; X - *Xba*I. Not all the restriction sites in the pW8 polylinker are shown (see Klemenz *et al.* (1987) for a complete polylinker restriction map). *I* factor sequences are shown in the crossed box, the *hsp70* promoter that directs transcription of the *white* gene is shown as the stripped box with lines parallel to the circumference of the plasmids, and *P* element sequences are shown as the stripped boxes with lines perpendicular to the circumference of the plasmid. The arrows indicate the direction of transcription. aug is the start codon of the CAT gene.

Figure 4.1


WHITE
pW8
8.3kb
P
X
B
R
H
H
P
Sa
X
B
Sm
aug
pCAT.1/
p186T.1
R
CAT
SV40
R
P
Pst1 digest and ligation of
CAT/SV40 fragment to pW8
WHITE
P
Sa
X
B
Sm
R
CAT
aug
pCATW8/
p186W8
9.9/
10.1kb
SV40
PXBR
H
0
1kb

such that *I* transcription would be in the opposite direction to transcription from both the *P* and *hsp70* promoter as a precaution to avoid readthrough transcription from upstream genes. The constructs were injected into the *D. melanogaster* strain W^K (Lüning, 1981) which is reactive. The *I* factor promoter behaviour should therefore be similar to that in a dysgenic environment.

4.2.2 Vector Sequences do not Direct Promoter Activity:

In order to determine what contribution (if any) vector sequences upstream of the insert make to CAT activity, pCATW8 was introduced into W^K and 13 lines homozygous for the transgene were established. The lines were known to be independent because each originated from a different injected embryo. Five pCATW8 transformants were analyzed in detail. The chromosome to which the transgene maps was determined (see appendix C), and each line was shown to have only one copy of the transgene by Southern blot analysis (data not shown).

CAT activity was assayed in extracts of developmental stages during which the sexes can be distinguished, comprising third instar larvae, early pupae, late pupae, 0-1 day old adults, and 3-4 day old adults. In addition, separated ovary and carcass tissues of 3-4 day old females were tested. Tables 4.1 lists the CAT activity, normalised to the protein concentration in the extract, for the five lines and W^K. All five lines gave similar results with conversion values very similar to those of W^K. Only female late pupae of line CAT.192 have measurable activity most probably directed by flanking genomic sequences. The overall negligible influence of upstream vector sequences permit CAT activity from the *I* promoter to be interpreted directly.

TABLE 4.1

LINE	CAT.187	CAT.188	CAT.192	CAT.197	CAT.198	W^K
♀L3	0	0	0	0	0	0
♀ E.P	0.34	0.20	0	0.48	0	ND
♀ L.P	0	0	6.05	0	0	0
♀0-1dy	0	0	0	0	0	ND
♀3-4dy	0	0	0	0	0	0
ov	0	0	0	0	0	ND
car	0	0	0	0	0	ND
♂L3	0	0	0	0	0	0
♂ E.P	0	0	0	0	0	ND
♂ L.P	0	0	0	0	0	0
♂0-1dy	0	0	0	0	0	ND
♂3-4dy	0	0	0	0	0	0

Values are given in pmoles ^{14}C-chloramphenicol acetylated/min/mg protein.

L3 = wandering third instar larvae; **E.P**= early pupae; **L.P**= late pupae; **0-1dy**= 0-1 day after eclosion; **3-4dy**= 3-4 days after eclosion; **ov**= ovaries from 3-4 day old female; **car**= carcass from 3-4 day old female; **ND**= not done.

4.2.3 Nucleotides 1-186 of *I* have Promoter Activity in Flies:

Ten independent lines, homozygous for the transgene p186W8 were established. Figure 4.2 illustrated Southern blot hybridisation showing that each line has only one copy of the transgene. Six of these lines were selected for further analysis and CAT assays performed on the same developmental stages as described for the pCATW8 transformants. The graphs in figures 4.3a-f illustrate the developmental profile of the CAT expression in each line normalised for the amount of protein in each sample, and examples of TLC separated CAT conversion products for females and males of line 186.137 and 186.148 are shown in figure 4.4. Each line has measurable CAT activity in all stages indicating that the *I* factor promoter functions in reactive flies. All six lines display a similar pattern of expression, with very low levels in third instar larvae, steadily increasing throughout ontogenesis. However, while the rate of increase in CAT activity is steady throughout males development, in females the rate rises after eclosion.

4.2.4 The *I* Factor Promoter is Most Active in Ovaries:

The fact that full length *I* transcripts are confined to ovaries suggests that the high level of CAT activity particular to female is due to high ovary expression. Ovaries were separated from carcass tissue in 3-4 day old female transformants, and CAT activity measured in the two tissues. Figure 4.5 shows that CAT activity in the ovary extract exceeds that of carcass tissue by several fold despite the concentration of protein in the ovary sample being less than in the carcass. Table 4.2 lists the CAT activities corrected for the protein concentration. CAT activity in the range of 200-1300 pmoles ^{14}C-chloramphenicol acetylated per minute per mg of protein is reasonably high, being comparable to that of the *Drosophila E74* promoter during larval development (Thummel, 1988), and the promoter of the 35S transcript of cauliflower mosaic virus in transformed tobacco plants (Grant *et al.*, 1990).

Where indicated values are given for assays done 20 and 36 generations following establishment of the line.

Figure 4.2 : **Southern blot analysis of lines transformed with p186W8.** Approximately 3μg of DNA from the 10 independent transformed lines were digested with *Eco*RI and following electrophoresis and Southern transfer, hybridised with the 250bp *Bam*HI-*Eco*RI fragment from the CAT gene in pCAT.1 (see figure 4.1). Only one band of hybridisation can be seen for each line indicating one integration event. p(a-87) and W^K are included as positive and negative controls respectively. The prefix '186' is omitted from the labelling above each p186W8 transformant track.

Figure 4.2

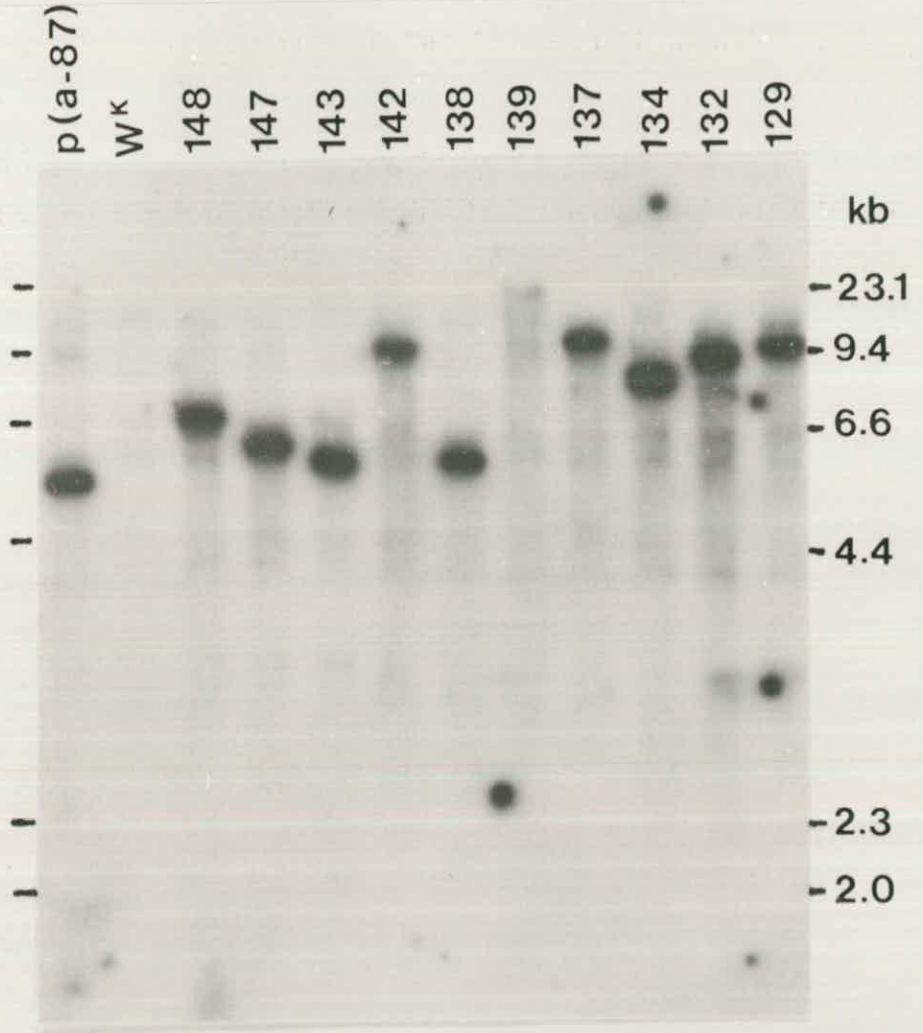

p(a-87)
W^K
148
147
143
142
138
139
137
134
132
129
kb
23.1
9.4
6.6
4.4
2.3
2.0

Figure 4.3 : **Developmental profile of CAT activity of p186W8 transformants.** Graphs a-f show the CAT conversion values of females and males from lines 186.132, 186.137, 186.138, 186.142, 186.143, and 186.148 respectively, normalised to the amount of protein used in the assay. Note that each graph has the same y axis scale i.e. 0-240 pmoles ^{14}C-chloramphenicol acetylated/min/mg protein, to emphasise variation between lines.

Figure 4.3a

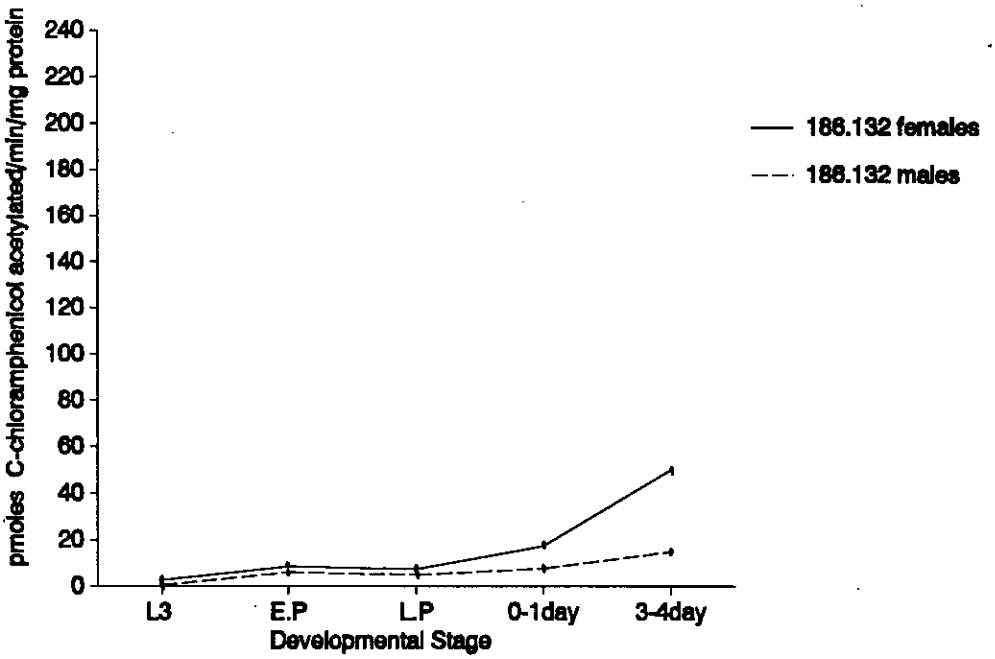

pmoles C-chloramphenicol acetylated/min/mg protein
240
220
200
180
160
140
120
100
80
60
40
20
0
L3
E.P
L.P
0-1day
3-4day
Developmental Stage
186.132 females
186.132 males

Figure 4.3b

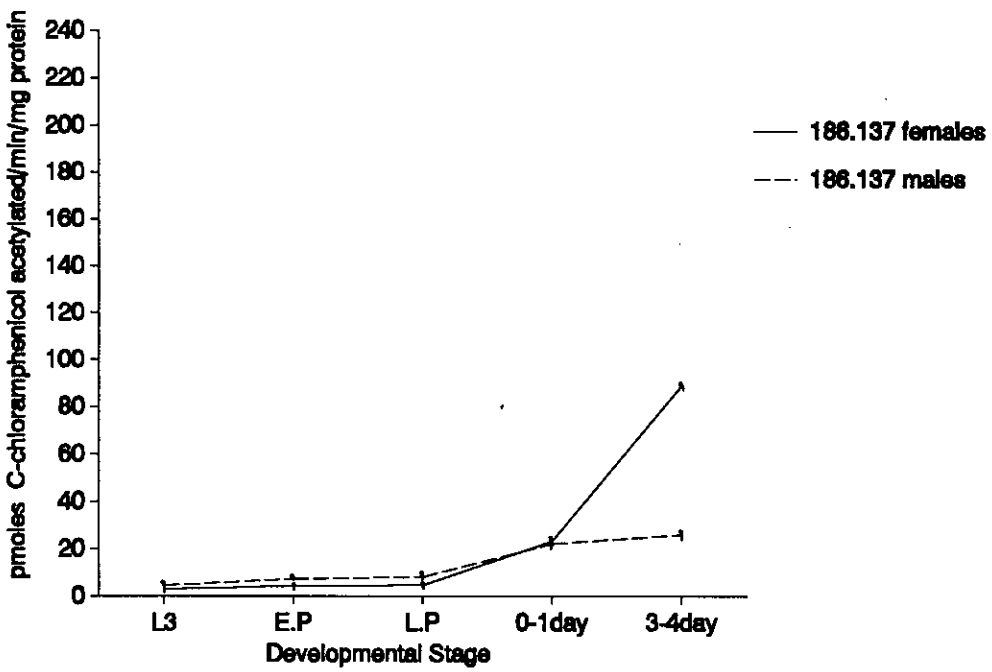

pmoles C-chloramphenicol acetylated/min/mg protein
240
220
200
180
160
140
120
100
80
60
40
20
0
L3
E.P
L.P
0-1day
3-4day
Developmental Stage
186.137 females
186.137 males

Figure 4.3c

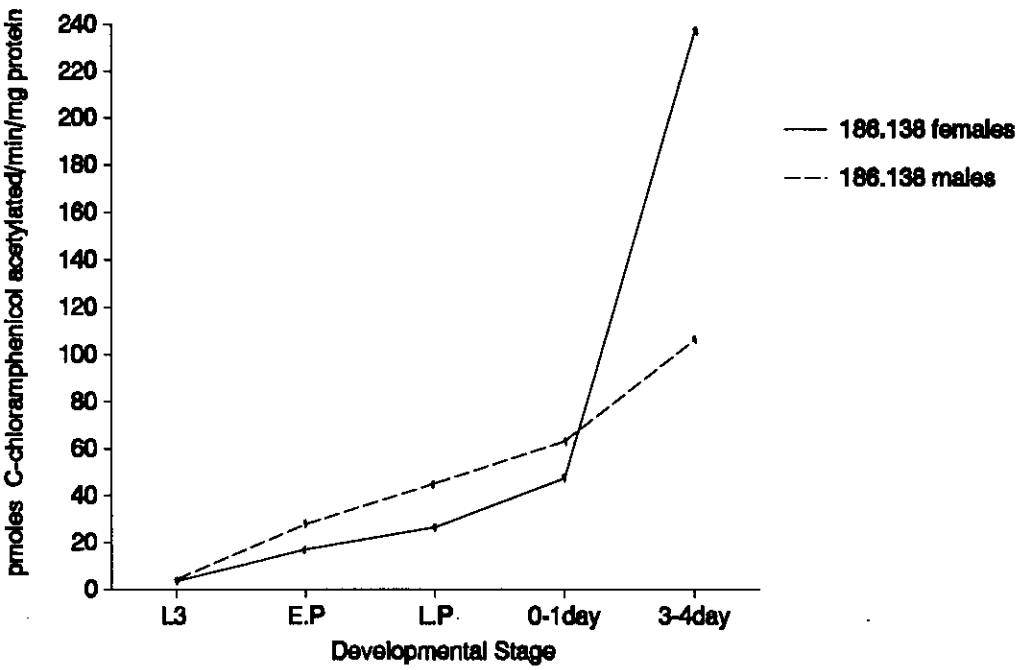

pmoles C-chloramphenicol acetylated/min/mg protein
240
220
200
180
160
140
120
100
80
60
40
20
0
L3
E.P
L.P
0-1day
3-4day
Developmental Stage
186.138 females
186.138 males

Figure 4.3d

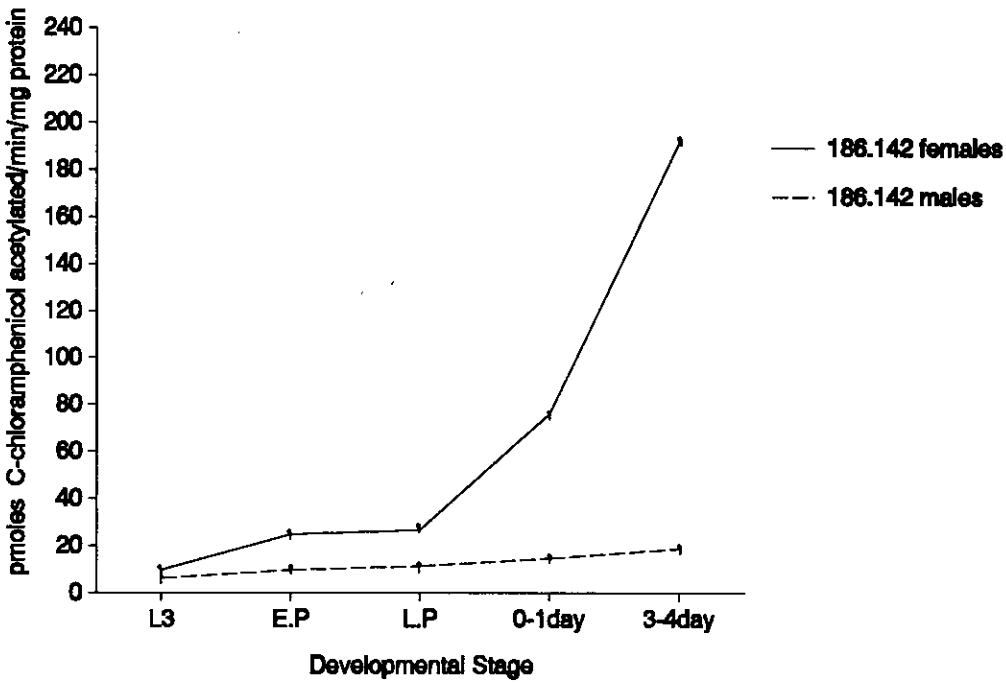

pmoles C-chloramphenicol acetylated/min/mg protein
240
220
200
180
160
140
120
100
80
60
40
20
0
L3
E.P
L.P
0-1day
3-4day
Developmental Stage
186.142 females
186.142 males

Figure 4.3e

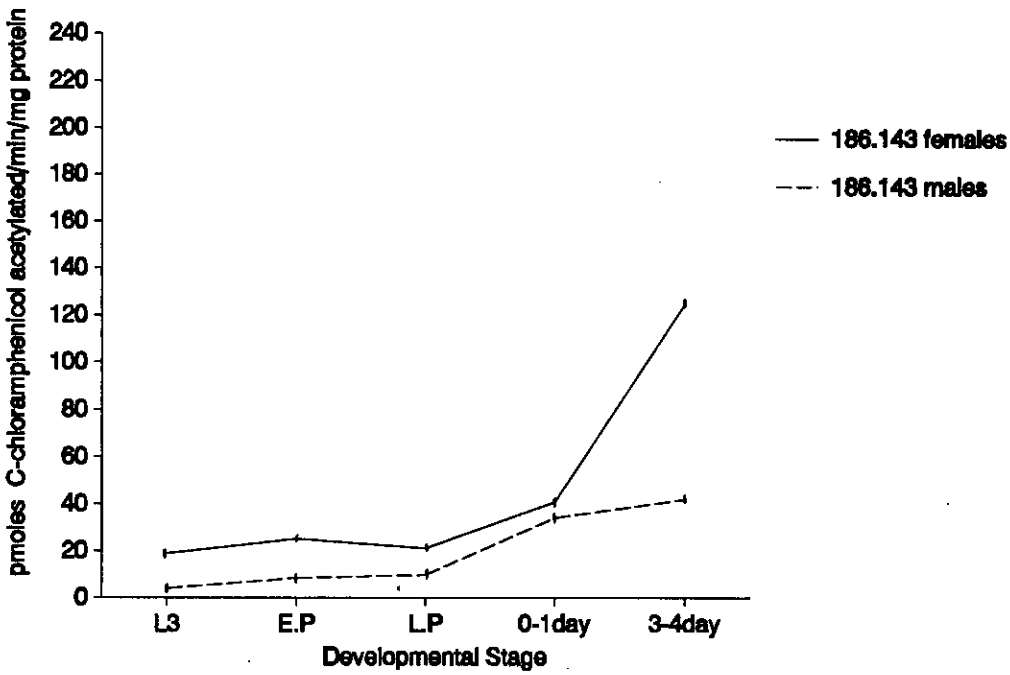

pmoles C-chloramphenicol acetylated/min/mg protein
240
220
200
180
160
140
120
100
80
60
40
20
0
L3
E.P
L.P
0-1day
3-4day
Developmental Stage
186.143 females
186.143 males

Figure 4.3f

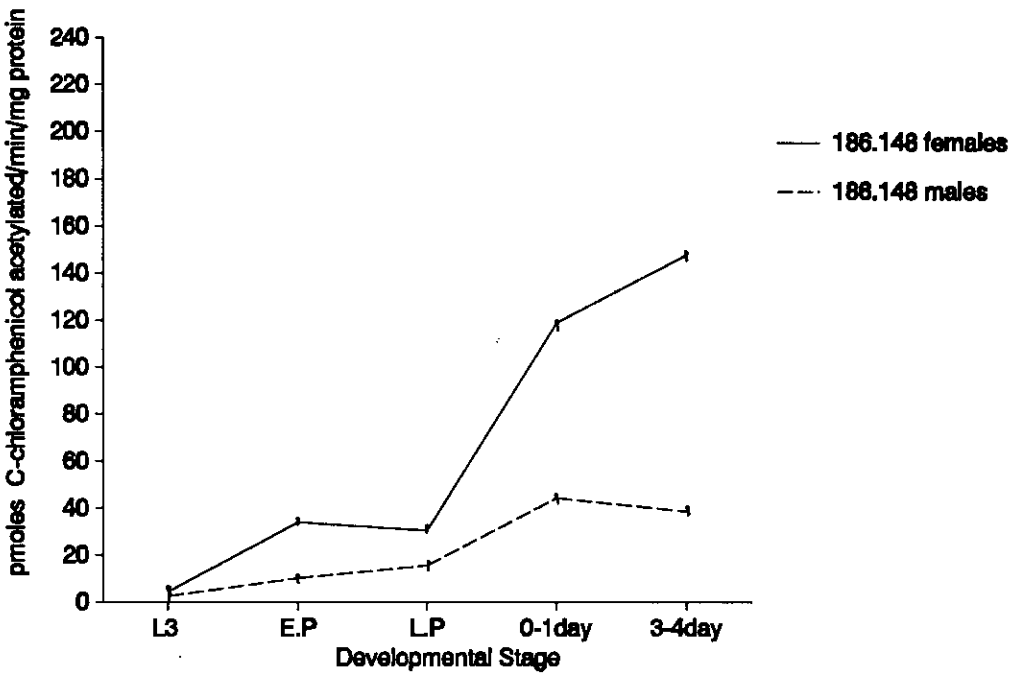

pmoles C-chloramphenicol acetylated/min/mg protein
240
220
200
180
160
140
120
100
80
60
40
20
0
L3
E.P
L.P
0-1day
3-4day
Developmental Stage
186.148 females
186.148 males

Figure 4.4 : **CAT assays of p186W8 transformants during development.** CAT assays of males and females from lines 186.137 and 186.148 are illustrated. L3 - wandering third instar larvae; ep - early pupae; lp - late pupae; 0-1 day - 0-1 day after eclosion; 3-4 day - 3-4 days after eclosion.

Figure 4.5 : **CAT assays of ovary and carcass tissues from p186W8 transformants.** CAT assays are illustrated for lines 186.137 and 186.143. ♀ - females 3-4 days after eclosion; ♂ - males 3-4 days after eclosion; car - carcass tissue; ov - ovary tissues.

Figure 4.4

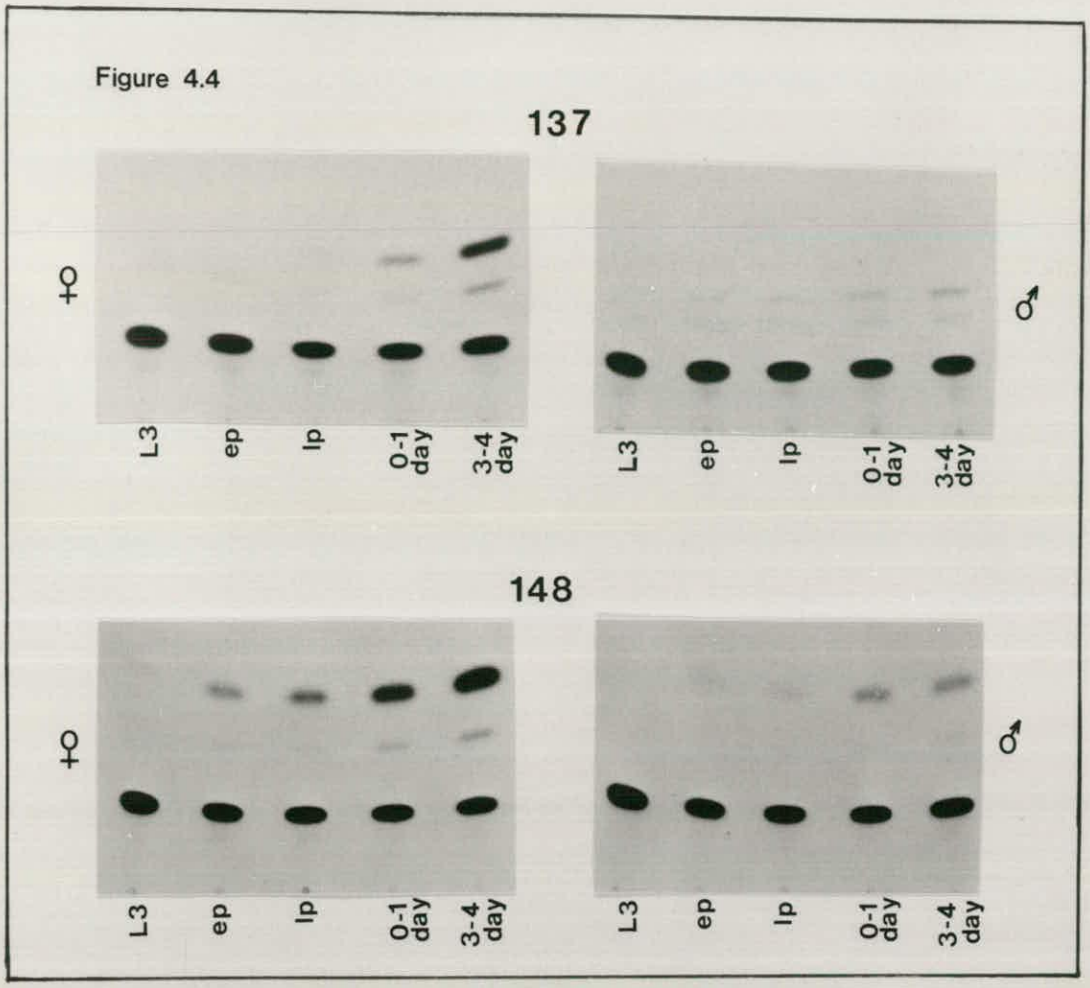


Figure 4.5

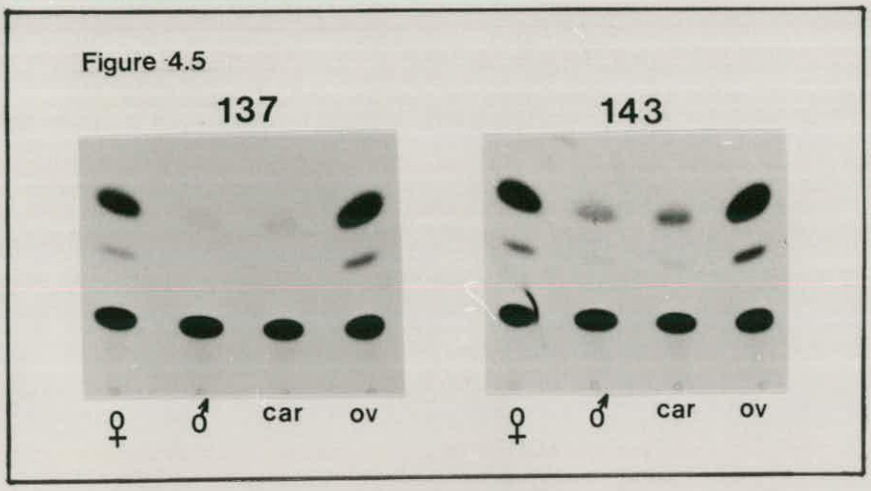


TABLE 4.2

	♀3-4dy	♂3-4dy	carcass	ovaries	ov:car
186.132	50*	15*	15*	160*	11:1*
186.137	89 55*	26 7*	19 1*	445 536*	23:1 536:1*
186.138	137*	78*	50*	229*	4.6:1*
186.142	192 143*	17 20*	50 26*	734 890*	15:1 34:1*
186.143	125 111*	42 36*	61 55*	707 928*	12:1 17:1*
186.148	148 129*	39 29*	109 72*	684 1259*	6:1 17:1*

Values are expressed in pmoles ^{14}C-chloramphenicol acetylated/min/mg protein.
* = assays done approximately 36 generations after line established, all others done 20 generations after establishment.

In most lines, the activity from female carcass is of the same order as in males. In contrast, the CAT activity in ovaries is up to several hundred times greater than in carcass, but in general is around 15 fold more. The increase in CAT activity particular to females after eclosion is therefore due to elevated promoter activity in ovarian tissues.

The first 186 nucleotides of *I* therefore contain sequences that have promoter activity in several developmental stages of males and females, but allow enhanced expression in ovaries.

4.2.5 The *I* Promoter is Most Active in Nurse Cells:

Prior to construction of transformed lines harbouring the *I* promoter directing transcription of a reporter gene, localisation of the cell type and developmental stage of *I* expression by standard *in situ* hybridisation techniques was made impossible by the numerous transcription and translation products originating from pericentromeric *I* elements. However, by using DNA probes specific to CAT RNA these problems were overcome.

Unlike *P* which transposes very early in germline development, before meiosis, transposition of *I* is thought to be at or shortly before meiosis, since clusters of I-R dysgenically induced mutations are rare and small (Picard *et al.*, 1978). In adult *D. melanogaster* females, meiosis occurs across restricted stages of oocyte development within the ovary. Figure 4.6(B) illustrates schematically an ovariole (15-20 ovarioles make up the adult ovary (figure 4.6A)), extending from the earliest structure, the germarium, at the anterior end, to the mature oocyte at the posterior end. The germarium contains stem cells of germline lineage. These divide to give a daughter stem cell and another cell that undergoes four cell divisions to produce a 16 cell cyst. Fifteen of the cells form the nurse cells, and one the oocyte. These 16 cells are surrounded by somatically derived follicle cells that together make up the egg chamber. The egg chamber then pinches off the germarium and proceeds through the 14 stages of oocyte development that together make up the ovariole (King, 1970). Meiosis starts at stage 13, is arrested in metaphase-1 in the mature oocyte (stage 14), and is not complete until after fertilisation (Puro and Nokkala, 1977).

Figure 4.7 depicts *in situ* hybridisation of labelled CAT DNA to ovaries from 3-4 day old 186.137 flies. This line was chosen because it has the highest ratio of ovary:somatic tissue CAT expression. The strongest staining is in the nurse cell cytoplasm at stages 9 and 10 of oogenesis. The negative control, W^K (pCATW8 transformants were unavailable at the time) has some staining, however it was consistently weaker than in 186.137 ovaries. The origin of this background staining is unknown as W^K females have no CAT activity (section 4.2.2).

Nurse cells synthesise most of the transcripts required by the oocyte, and

Figure 4.6 : **Schematic drawing of the *Drosophila* adult ovary** (Fasano and Kerridge, 1988, after King, 1970). (A) The ovary with 15-20 ovarioles, two of which have been pulled away. (B) A single ovariole with some stages numbered. (C) Stage 13.

Figure 4.6

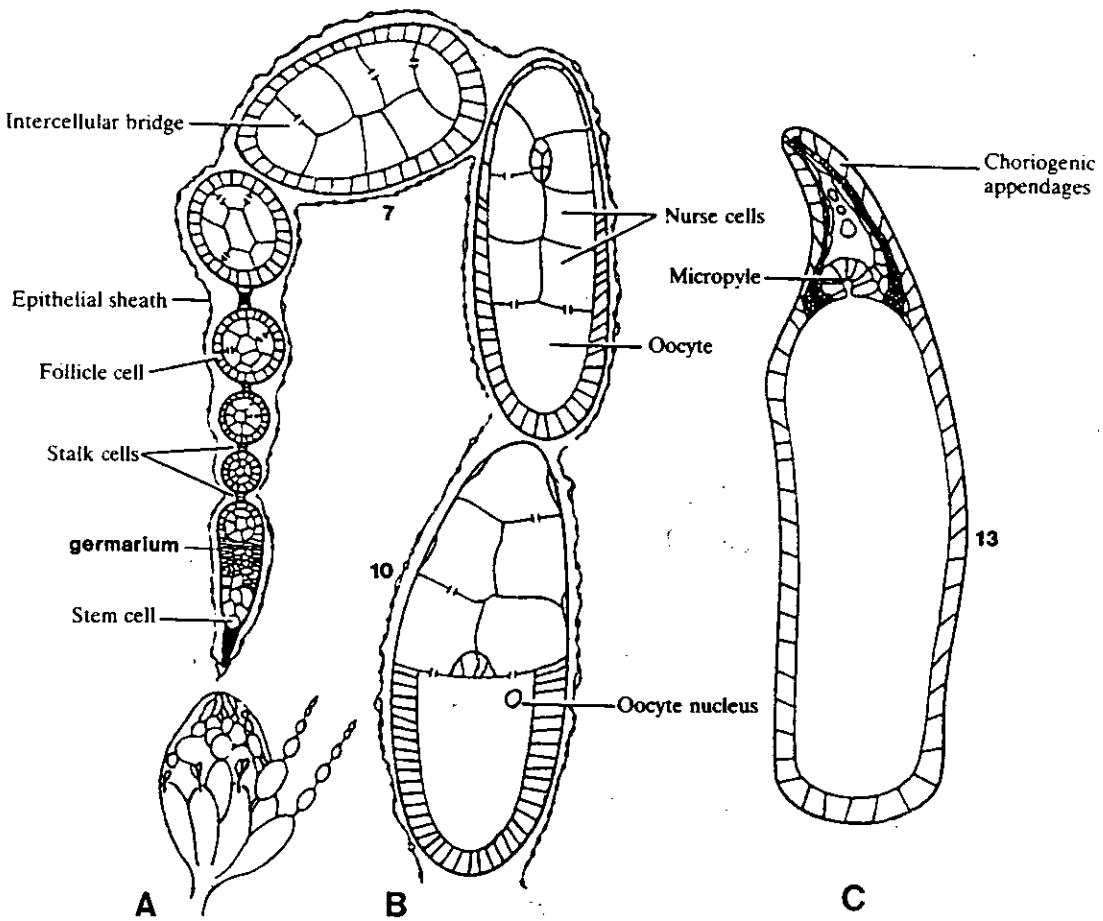
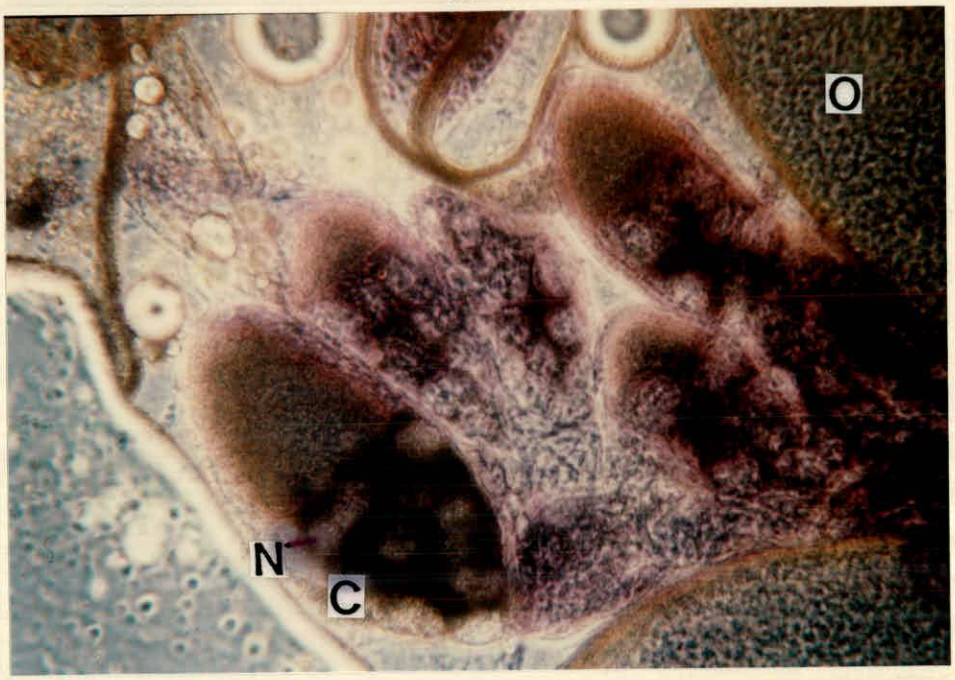

Intercellular bridge
Epithelial sheath
Follicle cell
Stalk cells
germarium
Stem cell
7
10
Nurse cells
Oocyte
Oocyte nucleus
Micropyle
Choriogenic appendages
13
A
B
C

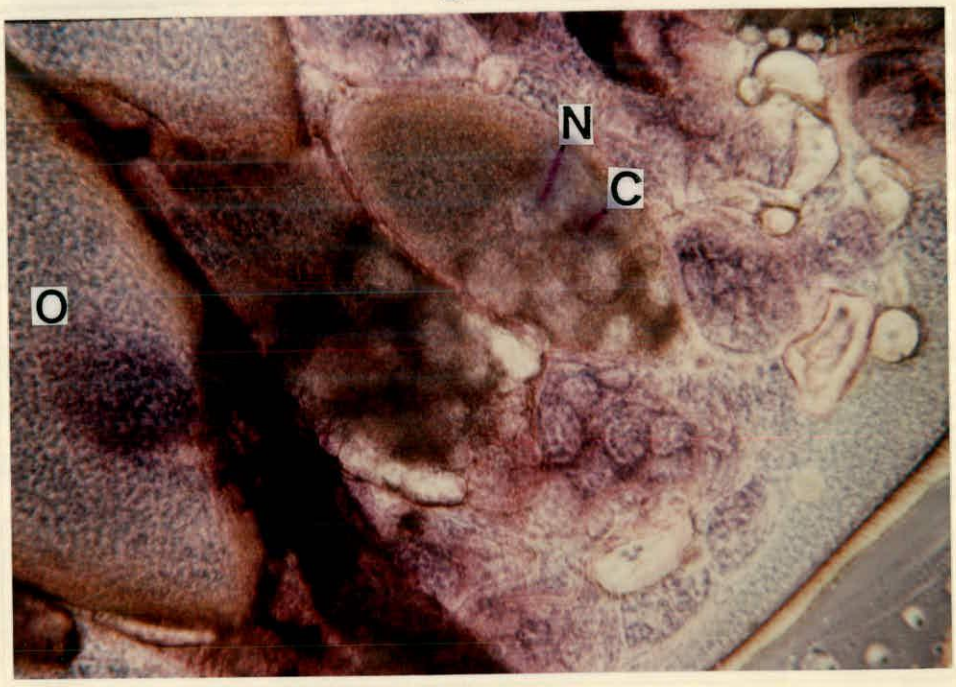
Figure 4.7 : ***In situ*** **hybridisation of digoxygenein labelled CAT DNA to ovary transcripts.** Magnification is x400. (A) Hybridisation of the CAT 250bp *Bam*HI-*Eco*RI fragment labelled with dig-UTP to ovaries from line 186.137. (B) Hybridisation of the same probe to ovaries form W^K. C - nurse cell cytoplasm; N - nurse cell nucleus; O - oocyte.

Figure 4.7

A



B



deliver them through intercellular bridges during stages 7-13, the majority being delivered at stage 11 (Mahowald and Kambysellis, 1980). Theoretically transcription of *I* at stages 9 and 10 could allow transposition in the nurse cell nuclei, that are lost, as well as the oocyte once the nurse cell delivers the *I* factor transcript or transposition particle. The oocytes of 186.137 stain weakly, however similar staining is apparent in W^K oocytes. It is therefore impossible to conclude that the *I*-CAT transcripts are maternally inherited. If, however *I*-CAT RNA does contribute the oocyte staining, the transcripts could be received from the nurse cells, or be synthesised in the oocyte nucleus, as a brief period of RNA synthesis takes place here between stages 9 and 10 (Mahowald and Tiefert, 1970). These results are consistent with *I* factor transposition occurring at, or shortly before meiosis.

4.2.6 The CAT Enzyme is Maternally Inherited:

To support the hypothesis that *I* factor transcripts and/or translation products enter or are synthesised in the oocyte, the CAT activity of 0-2 hour embryos was measured as transcription does not begin in the embryo until after this point. Table 4.3 lists the oocyte CAT activity together with the ovary activity at the same point in time for four lines. The oocyte activity in all lines is reasonably high indicating maternal inheritance, but in all cases is less than ovary activity. This difference may be due to instability if the transcript leading to a drop in CAT levels.

TABLE 4.3

LINE	0-2hour EMBRYO	OVARY
186.137	496	536
186.142	173	890
186.143	376	928
186.148	306	1259

Values are expressed in pmoles ^{14}C-chloramphenicol acetylated/min/mg protein.

DISCUSSION:

Transformation of reactive *D. melanogaster* flies with p186W8 demonstrated that nucleotides 1 to 186 of *I* contains a promoter, active in several developmental stages of males and females and directs around 15 fold higher activity in adult ovary tissue relative to female carcass and males of the same age. There is however variation in the levels of CAT between independent lines presumably because of position effects. Pélisson and Bregliano (1981) reported that the inducer potential of an *I* factor is influenced by its chromosomal environment as inducer strength can increase upon transposition. As argued in chapter 3, difference in inducer strength may be a reflection of position effect on promoter activity.

Enhanced expression in the ovaries strongly suggests that the promoter controls the tissue specificity, and consequently sex specificity of transposition in I-R hybrid dysgenesis. However, detectable promoter activity exists where no transposition and no full length transcripts have been reported. Sequences downstream of position 186 might therefore be required to suppress promoter activity, or cause RNA instability in tissues other than ovaries. Alternatively, transcription might occur in tissues other than ovaries but at a level too low to be detect by Northern analysis (Chaboissier *et al.*, 1990), and too low for mobility as transposition has never been reported in males.

The level of CAT activity during development could reflect an accumulation of CAT protein from a weak constitutive promoter, rather than a steady increase in promoter strength. RNA analysis and information on CAT stability in flies are lacking, therefore no conclusions can be drawn on this point. Nevertheless, even if CAT accumulation does occur, an increase in promoter activity in females following eclosion does take place, evident from the sudden increase in female CAT activity, absent in males.

In situ hybridisation analysis demonstrated that the *I* factor promoter is active in nurse cells and possibly the oocyte at stages 9 and 10 (figure 4.7). The daughters of SF females sometimes show SF sterility (Picard 1978a) implying that the *I* factor can transpose in her germline, however there is no evidence that *I* transposes to her paternally inherited chromosomes prior germline formation. It is therefore possible

that transposition is confined to the period from stage 9 to before fertilisation. At 25°C, the duration of this period is about 16 hours (David and Merle, 1968). Chromosomes condensation during meiosis may prevent transposition, restricting the permissive period further to 13 hours.

CHAPTER 5

5.1 INTRODUCTION:

Chapter 4 demonstrated that within the first 186 nucleotides of *I* lie sequences that direct promoter activity in reactive flies, with enhanced expression in ovaries. The experiments described in this chapter aim to identify those sequences that play a role in the pattern of *I* expression in reactive flies. This was done by assaying expression from a selection of 3' and 5' promoter deletion derivatives, through several developmental stages.

5.2 RESULTS

5.2.1 Construction of Plasmids for Promoter Assay in Flies:

Constructs for P mediated transformation of W^K were made as described in section 4.2.1 except the *PstI* fragment of p100T, p40T, p28T, and p186Δ17 were inserted into pW8 to form p100W8, p40W8, p28W8, and pΔ17W8 respectively. Again the orientation of these insert was such that putative *I* transcription is in the opposite direction to both *P* and *white*.

Developmental Expression Pattern in p100W8 Transformants:

Five independent homozygous lines, 100.178, 100.179, 100.182, 100.183, 100.184, transformed with p100W8 were established, four of which have one copy of the transgene, and one line (100.182) with two copies. The developmental profile of CAT expression was determined for the three lines, 100.178, 100.179, and 100.183, and is illustrated graphically in figures 5.1a-c. An autoradiograph of line 100.183 CAT activity is shown in figure 5.2.

Removal of sequences +101 to +186 has a dramatic effect on expression. This contrasts with the apparent unimportance of these sequences in tissue culture cells. All three lines show a burst in promoter activity during pupal development in both males and females although it is more pronounced in females. The CAT levels produced are comparable to those of ovary tissues in p186W8 transformants. However, CAT activity recedes after eclosion. Sequences between +100 and +187 therefore suppress high promoter activity during pupal development of male and female flies and following eclosion in males.

Figure 5.1 : **Developmental profile of CAT expression in lines transformed with p100W8.** Graphs a-c depict the CAT activity of males and females from lines 100.178, 100.179, and 100.183 respectively. The y axis scale (in pmoles ^{14}C-chloramphenicol acetylated/min/mg protein) is the same for each line to emphasise variation between lines. On the x axis: L3 - wandering third instar larvae; EP - early pupae; LP - late pupae; 0-1 day - adults 0-1 day after eclosion; 3-4 day - adults 3-4 days after eclosion.

Figure 5.1a

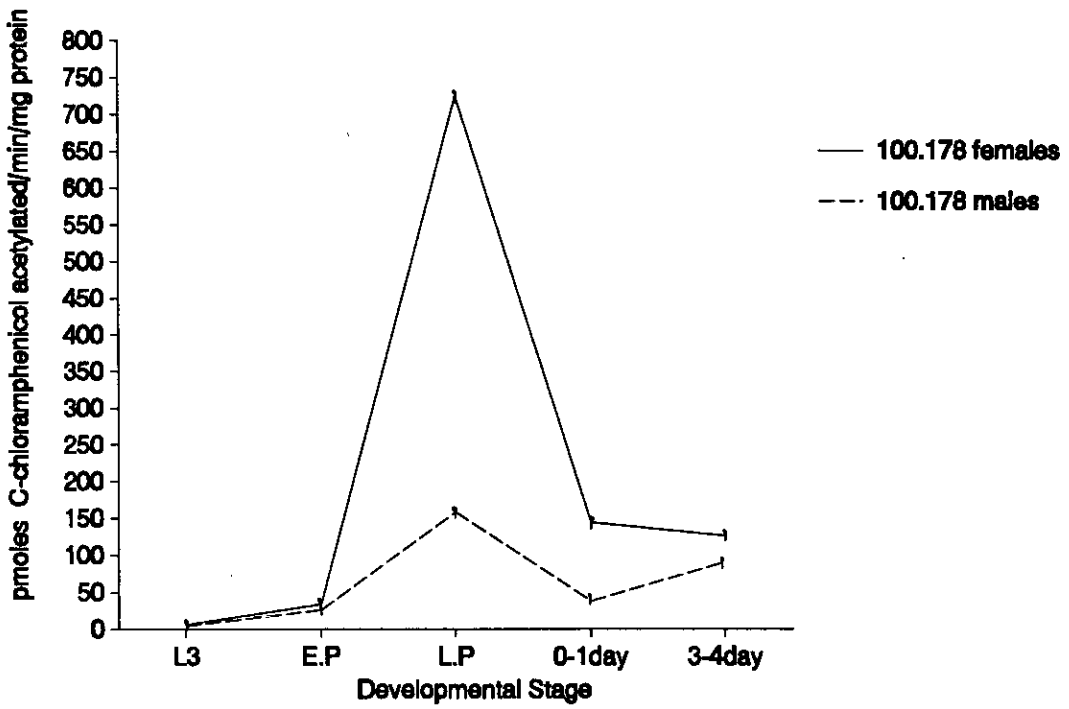

pmoles C-chloramphenicol acetylated/min/mg protein
800
750
700
650
600
550
500
450
400
350
300
250
200
150
100
50
0
L3
E.P
L.P
0-1day
3-4day
Developmental Stage
100.178 females
100.178 males

Figure 5.1b

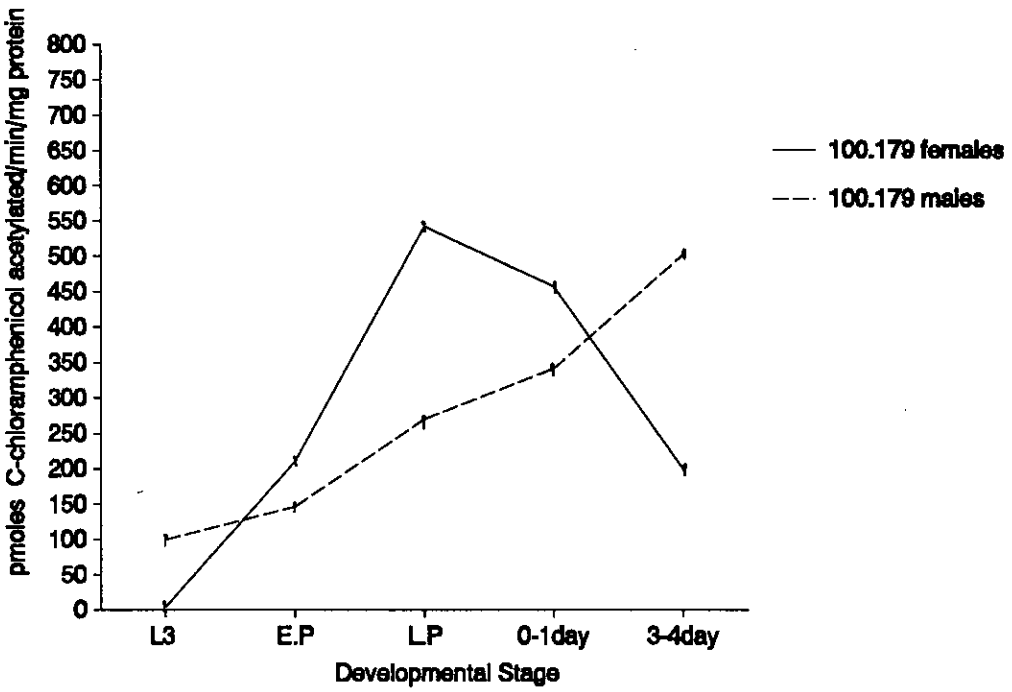

pmoles C-chloramphenicol acetylated/min/mg protein
800
750
700
650
600
550
500
450
400
350
300
250
200
150
100
50
0
L3
E.P
L.P
0-1day
3-4day
Developmental Stage
100.179 females
100.179 males

Figure 5.1c

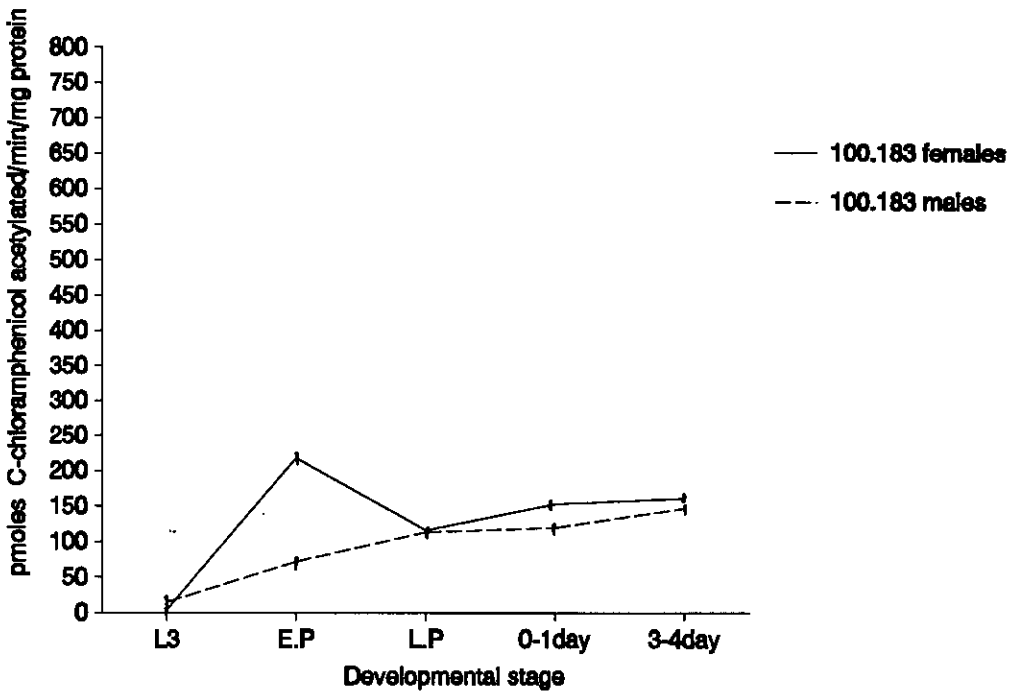

pmoles C-chloramphenicol acetylated/min/mg protein
800
750
700
650
600
550
500
450
400
350
300
250
200
150
100
50
0
L3
E.P
L.P
0-1day
3-4day
Developmental stage
100.183 females
100.183 males

Figure 5.2 : **CAT assays of developmental stages of line 100.183.** L3 - wandering third instar larvae; EP - early pupae; LP - late pupae; 0-1 day - adults 0-1 day after eclosion; 3-4 day - adults 3-4 days after eclosion.

Figure 5.3 : **CAT assays of ovary and carcass tissues of p100W8 lines.** Examples of CAT assays for line 100.178 and 100.183 are shown. ♀ - females 3-4 days after eclosion; ♂ - males 3-4 days after eclosion; car - carcass of 3-4 day old females; ov - ovaries from 3-4 day old females.

Figure 5.2

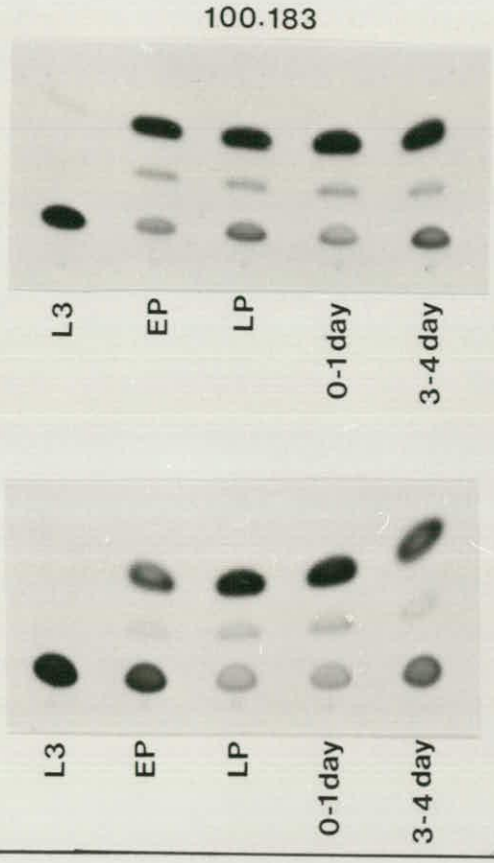
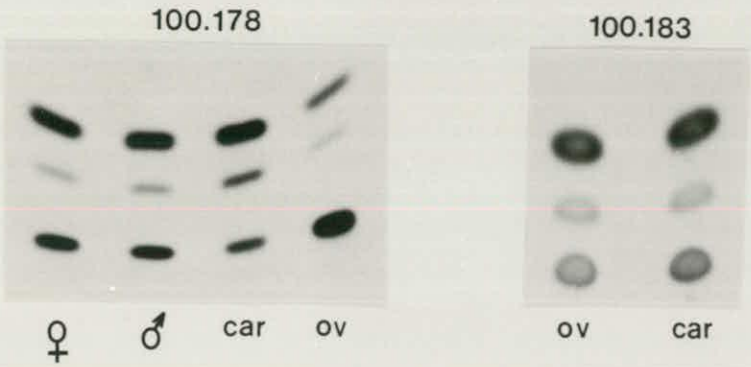


Figure 5.3



5.2.3 CAT Activity in Ovaries and Carcass of p100W8 Transformants:

In an attempt to determine whether the sequences responsible for ovary enhancement in p186W8 lines lie downstream of position +100, the ovary and carcass CAT activity was measured for all four lines. Figure 5.3 shows CAT assays for lines 100.178 and 100.183 and table 5.1 lists the values for all five lines normalised for the protein concentration in the sample, together with the ratio of CAT activity between ovaries and carcass tissues and CAT activity of adult males and females.

TABLE 5.1

	♀3-4dy	♂3-4dy	carcass	ovary	ov:car
100.178	143	244	63	91	1.5:1
100.179	55	49	30	395	13:1
100.182	5223	10506	7185	1675	1:4
100.183	161	147	82	246	3:1
100.184	149	47	33	125	4:1

Values for CAT activity are in pmoles ^{14}C-chloramphenicol acetylated/min/mg protein. Some ovary enhancement exists in all lines, except 100.182 that shows four fold carcass enhanced expression. CAT activity values for this line are exceedingly high suggesting that at least one of its transgenes is under a position affect. The results from line 100.182 will therefore be ignored. In general ovary enhancement in the remaining four lines is less marked than in p186W8 transformants with line 100.179 showing a 13 fold ovary enhancement, line 100.178 only 1.5 fold ovary enhancement and the others intermediate enhancement.

5.2.4 Developmental Expression Pattern of p40W8 Transformant:

Two independent lines, each with a single copy of the p40W8 transgene, were established but despite several attempts only one of the lines, 40.177, could be made

Figure 5.4 : **Developmental profile of CAT expression in lines transformed with p40W8.** The graph depicts the CAT activity of males and females from lines 40.177. On the x axis: L3 - wandering third instar larvae; EP - early pupae; LP - late pupae; 0-1 day - adults 0-1 day after eclosion; 3-4 day - adults 3-4 adys after eclosion.

Figure 5.4

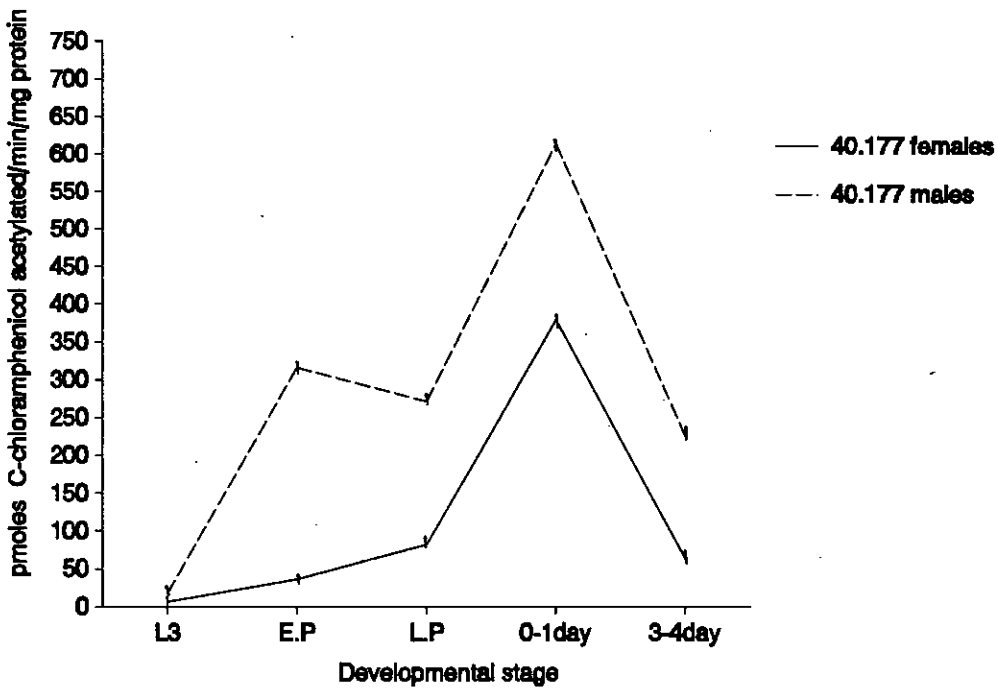

pmoles C-chloramphenicol acetylated/min/mg protein
750
700
650
600
550
500
450
400
350
300
250
200
150
100
50
0
L3
E.P
L.P
0-1day
3-4day
Developmental stage
40.177 females
40.177 males

Figure 5.5 : **CAT assays of developmental stages of line 40.177.** L3 - wandering third instar larvae; EP - early pupae; LP - late pupae; 0-1 day - adults 0-1 day after eclosion; 3-4 day - adults 3-4 days after eclosion.

Figure 5.6 : **CAT assays of ovary and carcass tissues of line 40.177.** ♀ - females 3-4 days after eclosion; ♂ - males 3-4 days after eclosion; car - carcass of 3-4 day old females; ov - ovaries of 3-4 day old females.

Figure 5.5

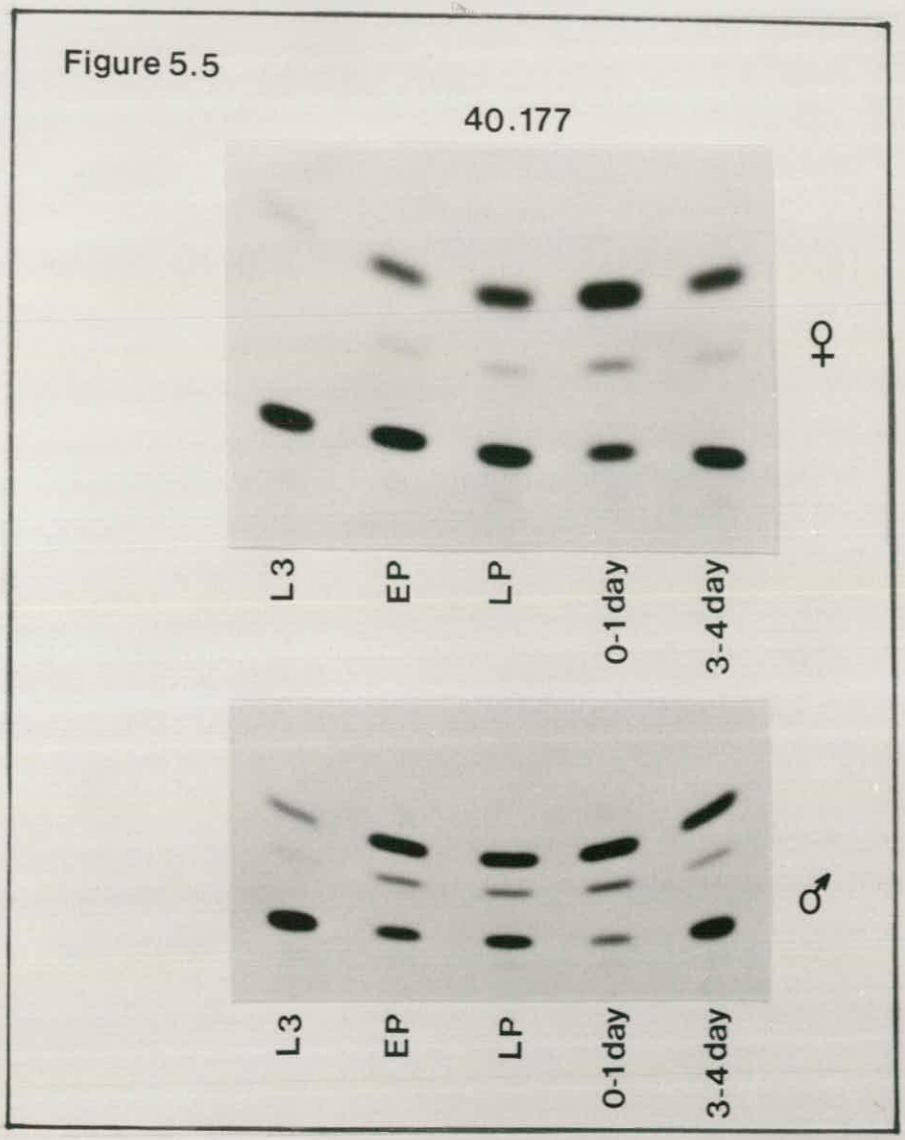
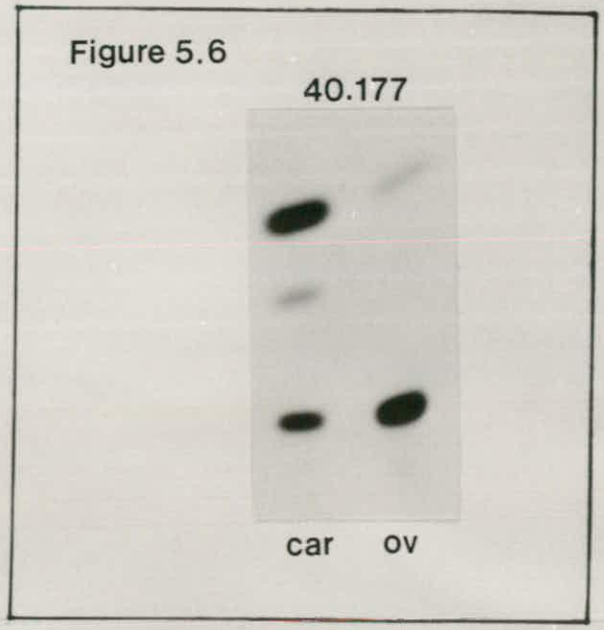


Figure 5.6



homozygous. The graph in figure 5.4 illustrates the CAT activity values of various male and female developmental stages normalised to protein concentrations, and figure 5.5 an autoradiograph of the CAT assays performed. The expression pattern is very similar to that of p100W8 transformants with a burst of activity in both males and females during pupal development. However, unlike p100W8 lines, male activity exceeds female and the burst peaks soon after eclosion and then falls off in 3-4 day adults.

5.2.5 CAT Activity in Ovaries and Carcass of p40W8 Transformants:

The CAT activity of ovary and female carcass tissues of both lines was measured and compared with male and female activity from flies of the same age. Figure 5.6 illustrates CAT assays of ovary and carcass tissues of the line 40.177 and table 5.2 summarises the data.

TABLE 5.2

LINE	♀3-4day	♂3-4day	carcass	ovary	ov:car
40.177	135	226	210	23	1:9
40.185	178	168	84	140	1.7:1

Values for CAT activity are in pmoles ^{14}C-chloramphenicol acetylated/min/mg protein.

The two lines are somewhat inconsistent, however there is no marked ovary enhancement, on the contrary line 40.177 displays suppression, indicating sequences for ovary enhancement lie downstream of position +40. Also carcass and male activities are high signifying loss of suppression in those tissues.

5.2.6 Developmental Expression Pattern in p28W8 Transformants:

Three independent lines harbouring a single copy of the p28W8 transgene were established, and two of them were made homozygous. The developmental profiles of

Figure 5.7 : **Developmental profile of CAT expression in lines transformed with p28W8.** Graphs a and b depict the CAT activity of males and females from lines 28.166, and 28.167 respectively. The y axis scale (in pmoles ^{14}C-chloramphenicol acetylated/min/mg protein) is the same for each line to emphasise the different level of expression in each line. On the x axis: L3 - wandering third instar larvae; EP - early pupae; LP - late pupae; 0-1 day - adults 0-1 day after eclosion; 3-4 day - adults 3-4 days after eclosion.

Figure 5.7a

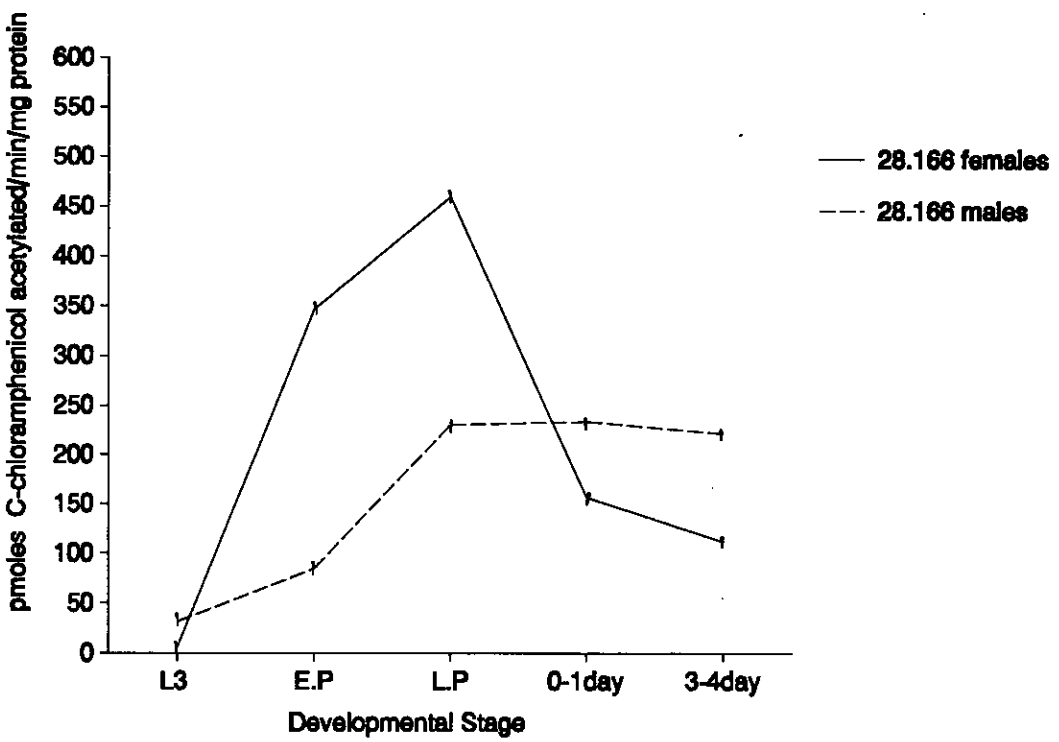

pmoles C-chloramphenicol acetylated/min/mg protein
600
550
500
450
400
350
300
250
200
150
100
50
0
L3
E.P
L.P
0-1day
3-4day
Developmental Stage
28.166 females
28.166 males

Figure 5.7b

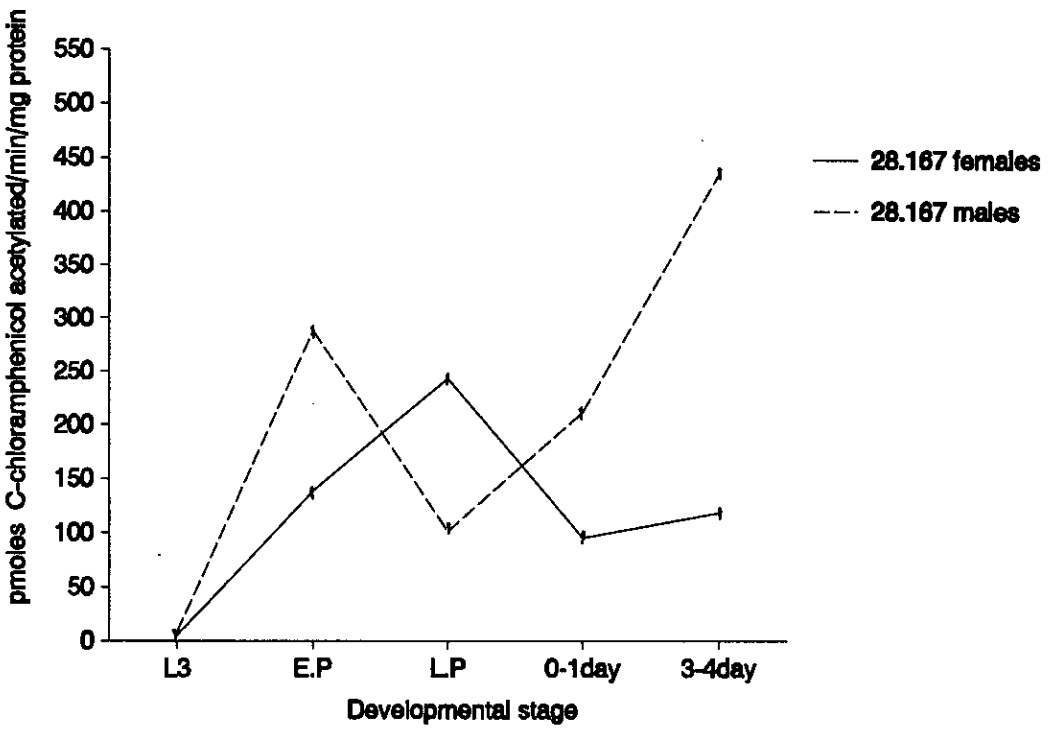

pmoles C-chloramphenicol acetylated/min/mg protein
550
500
450
400
350
300
250
200
150
100
50
0
L3
E.P
L.P
0-1day
3-4day
Developmental stage
28.167 females
28.167 males

Figure 5.8 : **CAT assays of developmental stages of line 28.166.** L3 - wandering third instar larvae; EP - early pupae; LP - late pupae; 0-1 day - adults 0-1 day after eclosion; 3-4 day - adults 3-4 days after eclosion.

Figure 5.9 : **CAT assays of ovary and carcass tissues of line 28.166.** ♀ - females 3-4 days after eclosion; ♂ - males 3-4 days after eclosion; car - carcass of 3-4 day old females; ov - ovaries of 3-4 day old females.

Figure 5.8

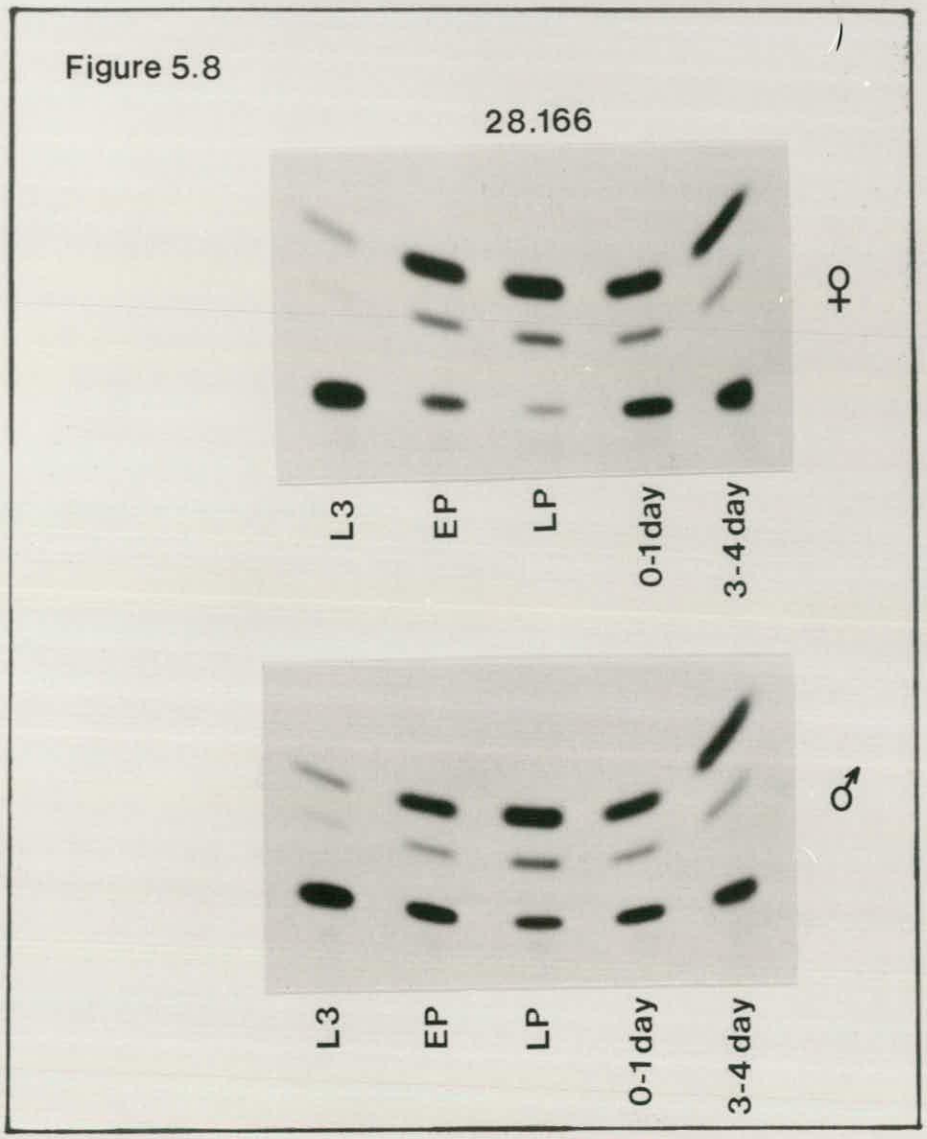
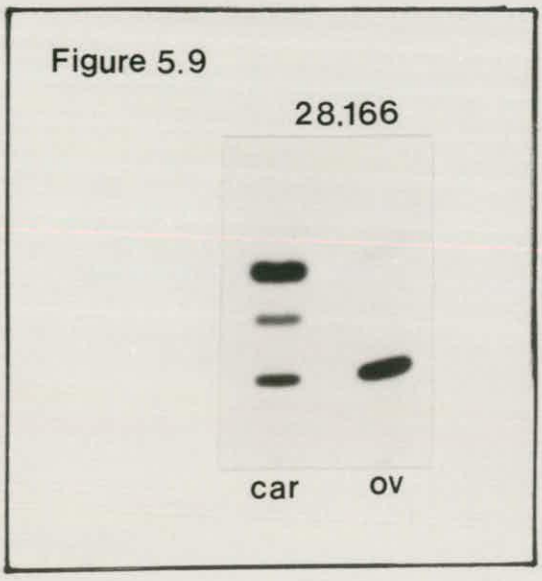


Figure 5.9



both, was determined and are shown graphically in figure 5.7a and b. A CAT assay of line 28.166 is shown in figure 5.8. Again the profile is similar to p100W8 transformants in that males and females display a burst in CAT activity during pupal development that in general recedes after eclosion.

5.2.7 CAT Activity in Ovaries and Carcass of p28W8 Transformants:

The CAT activities of all three p28W8 transformants were determined for ovary and carcass, and compared with the values for males and females of the same age. Figure 5.9 illustrates an autoradiograph of a CAT assay from line 28.166. Examination of the results in table 5.3 for all three lines show them to be very similar to line 40.177 with high male and female carcass CAT activity and some suppression of ovary expression. Removal of the conserved motif around position +30 does not, therefore, have a marked affect on the expression pattern or strength of the *I* factor promoter in flies.

TABLE 5.3

LINE	♀3-4day	♂3-4day	carcass	ovary	ov:car
28.166	112	221	161	15	1:11
28.167	118	435	178	55	1:3
28.165	151	205	270	136	1:2

Values are given in pmoles ^{14}C-chloramphenicol acetylated/min/mg protein.

5.2.8 Developmental Expression Pattern of pΔ17W8 Transformants:

Three independent lines homozygous for the pΔ17W8 transgene were established, each having a single copy of the transgene determined by Southern blot hybridisation (data not shown). The developmental profiles of CAT expression were determined for all three lines and are shown graphically in figures 5.10a-c. Figure

Figure 5.10 : **Developmental profile of CAT expression in lines transformed with pΔ17W8.** Graphs a-c depict the CAT activity of males and females from lines Δ17.162, Δ17.163, and Δ17.164 respectively. The y axis scale (in pmoles ^{14}C-chloramphenicol acetylated/min/mg protein) is constant for each graph to emphasise the different levels of expression in each line. On the x axis: L3 - wandering third instar larvae; EP - early pupae; LP - late pupae; 0-1 day - adults 0-1 day after eclosion; 3-4 day - adults 3-4 adys after eclosion.

Figure 5.10a

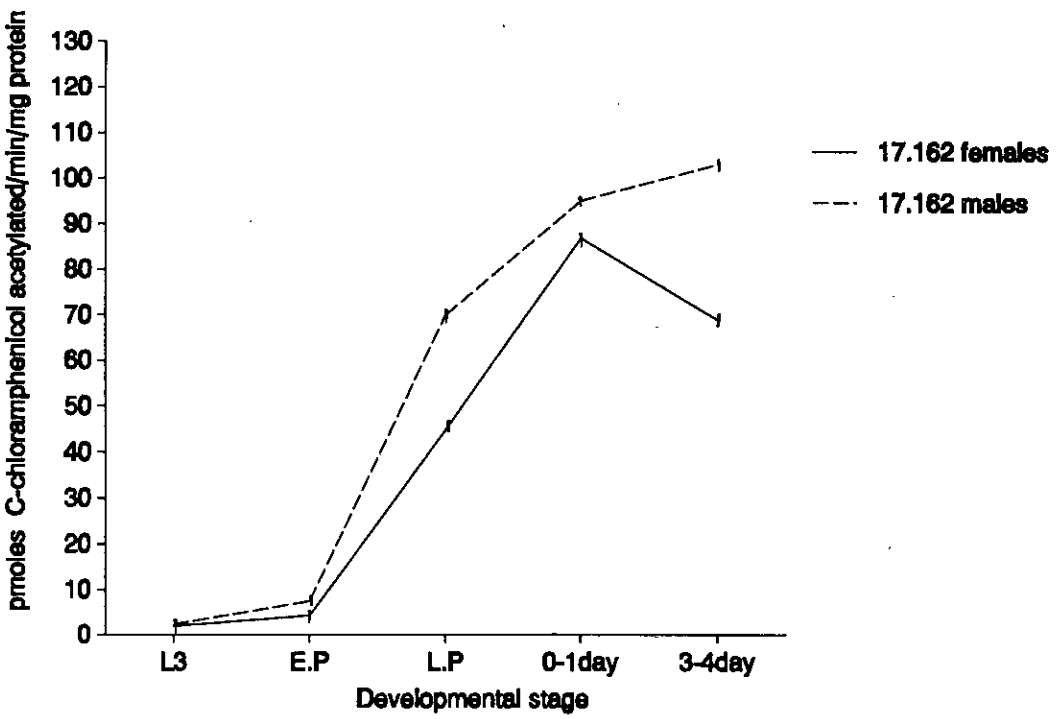

pmoles C-chloramphenicol acetylated/min/mg protein
130
120
110
100
90
80
70
60
50
40
30
20
10
0
L3
E.P
L.P
0-1day
3-4day
Developmental stage
17.162 females
17.162 males

Figure 5.10b

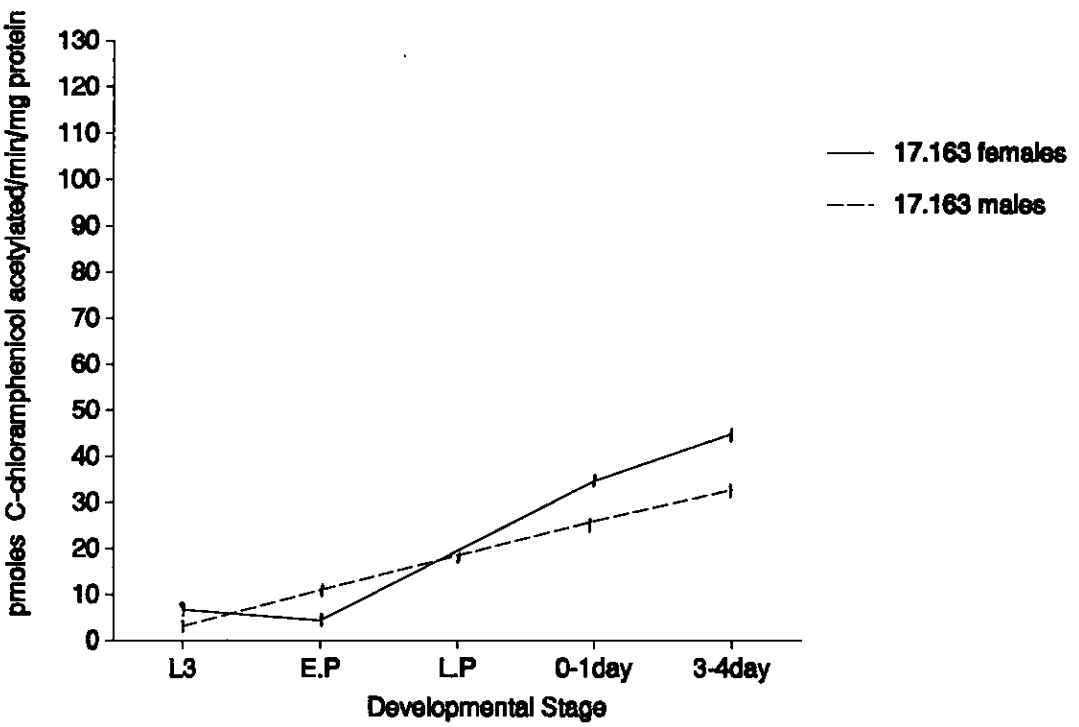

pmoles C-chloramphenicol acetylated/min/mg protein
130
120
110
100
90
80
70
60
50
40
30
20
10
0
L3
E.P
L.P
0-1day
3-4day
Developmental Stage
17.163 females
17.163 males

Figure 5.10c

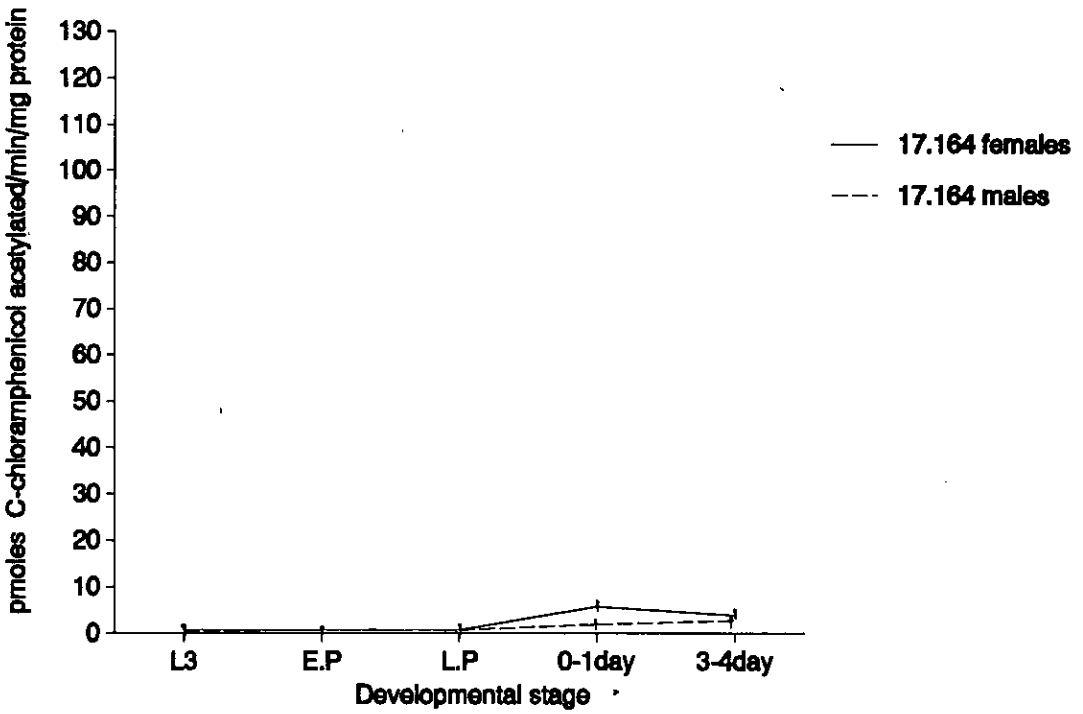

pmoles C-chloramphenicol acetylated/min/mg protein
130
120
110
100
90
80
70
60
50
40
30
20
10
0
L3
E.P
L.P
0-1day
3-4day
Developmental stage
17.164 females
17.164 males

Figure 5.11 : **CAT assays of developmental stages of line Δ17.163.** L3 - wandering third instar larvae; EP - early pupae; LP - late pupae; 0-1 day - adults 0-1 day after eclosion; 3-4 day - adults 3-4 days after eclosion.

Figure 5.12 : **CAT assays of ovary and carcass tissues of line Δ17.163.** ♀ - females 3-4 days after eclosion; ♂ - males 3-4 days after eclosion; car - carcass of 3-4 day old females; ov - ovaries of 3-4day old females.

Figure 5.11

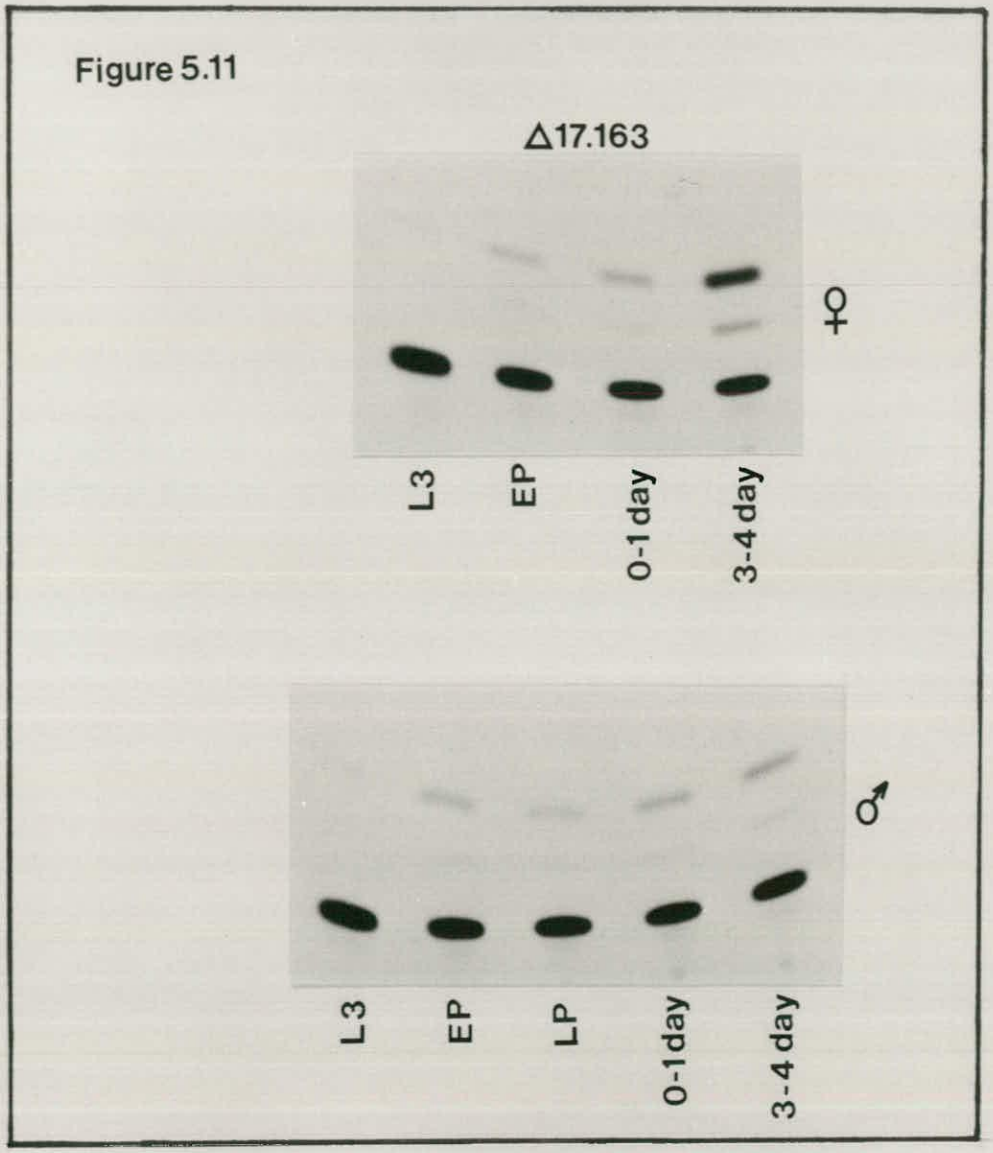
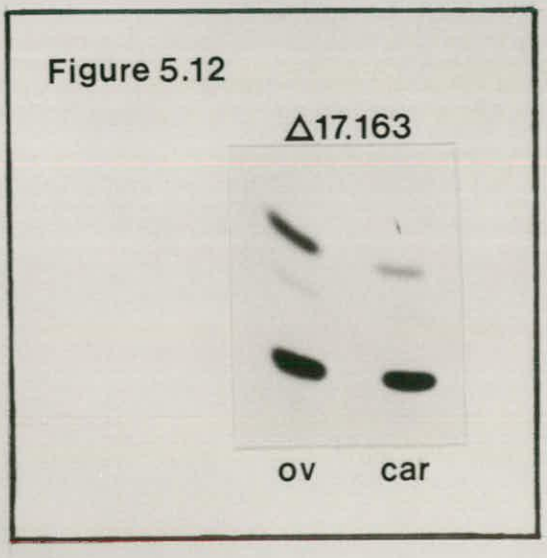


Figure 5.12



5.11 illustrates an autoradiograph of CAT activity for line Δ17.163. Surprisingly, two of the lines, Δ17.162 and Δ17.163, have CAT activity similar to p186W8 transformants, even though this promoter fragment directing no CAT activity in tissue culture. Examination of upstream vector sequences reveals the motif CAGG 12 to 15bp upstream from the first nucleotide of the pΔ17W8 promoter fragment, very similar in sequence and location to the CAGT motif at the start of the *I* factor. From tissue culture analysis it was shown that the T in the CAGT motif can be changed to an A with no major affect on promoter strength. CAGG might therefore be able to substitute for the wildtype sequence. However, this sequence exists at the same location upstream of the promoter fragment in construct p186Δ17 that showed no CAT activity in tissue culture. In addition sequences between the CAGG motif and the start of the Δ17 fragment bear no resemblance to corresponding *I* sequences. One would imagine that the sequences from +5 to +17 are important for function as they are highly conserved between *D. melanogaster* and *D.teissieri I* factors. Construct p186Δ17 has not been tested for CAT activity in tissue culture transfection experiments. It is therefore possible that the upstream vector sequences in pW8 and sequences +18 to +186 of *I* together form a promoter.

5.2.9 Δ17W8 Transformants Display Enhanced Ovary CAT Activity:

CAT assays of ovary and carcass tissues of line Δ27.163 are illustrated in figure 5.12 and table 5.4 lists the CAT activity values for ovary and carcass tissues of the three Δ17W8 lines compared with the values for females and males of the same age. The level of expression is higher in females than in males for two of the lines, and all three lines show enhanced CAT expression in the ovaries. These data support the conclusion that sequences controlling ovary enhancement lie downstream of position +40.

TABLE 5.4

LINE	♀3-4day	♂3-4day	carcass	ovary	ov:car
Δ17.162	69	103	26	159	6:1
Δ17.163	45	33	28	213	8:1
Δ17.164	4	3	5	81	16:1

Values given in pmoles ^{14}C-chloramphenicol acetylated/min/mg protein.

5.3 DISCUSSION:

Analysis of the CAT expression pattern of p186W8 transformants in chapter 4 could be simplistically interpreted as a weak promoter active in both males and females with sequences that allow enhanced expression in ovary tissues thus conferring, or at least contributing to, the tissue specificity and consequently the sex specificity of *I* transposition. However the results presented in this chapter uncover a far more complex scenario of promoter regulation.

Removal of 3' sequences to position +100 drastically alters both the developmental and sex specific *I* factor promoter activity, resulting in a burst of expression during pupal stages of both males and females that levels of in adults. Sequences between +100 and +187 therefore suppress *I* promoter activity in those tissues. Gel retardation and DNA footprinting experiments performed by Arkhipova (personal communication) identified a protein which binds the sequence CAACAAAAACAACAATAC that spans position +135 to +152 of the *I* factor. This binding activity is found in Schneider line 2 and K_c *Drosophila* tissue culture cells, pupae, and males and females of both inducer and reactive flies. Identical binding activity is also found in the ovaries making it unlikely that absence of this protein facilitates high expression only in ovaries, unless the protein is not in the nurse cells. No binding activity particular to the ovaries has been reported for nucleotides 1 to 186

(Arkhipova, personal communication). This suggests that the putative ovary specific *trans* acting factor is rare or competes with the suppressor for the same binding site and coincidentally retards promoter fragments by the same amount on bandshift assays. Arkhipova sometimes observed that the retarded complex with ovary extracts is stronger which might mean the binding activity is not a suppressor but a protein that enhances expression and is more abundant in ovaries. Alternatively a stronger signal may be because the ovaries are free of many of the proteases and nucleases present in the gut of the adult and pupae samples.

Transformants with nucleotides 1 to 100 show more CAT activity in ovaries than carcass, indicating sequences that regulate high expression in the ovaries are present. However the ratio of ovary:carcass CAT activity is lower than in p186W8 transformants. This could be because sequences downstream of position +100 also contribute to the ovary enhancement, or increased carcass expression is masking the enhancement effect. The CAT activities of carcass samples from p100W8 lines are rather similar to those from p186W8 lines (cf. table 4.6). However ovary CAT expression is somewhat lower indicating that sequences contributing to enhanced ovary expression have been removed. Sequences upstream of position +101 are however also involved in stimulating expression in the ovaries as four of the five lines retain some ovary enhancement. An anomaly of p100W8 lines is that male CAT activity is sometimes several fold greater than that of female carcass. Why then does female carcass activity appear suppressed but male activity not? The answer could involve sex and somatic specific control, or may be because the difference in promoter activity between males and female carcass is slight but the sample size (10 individuals for each category) is too small to give an accurate representation of this.

CAT expression during development from p40W8 transformants is similar to that of p100W8 transformants with high pupal expression that plateaus in 3-4 day old adults. However in contrast to the p100W8 promoter, female carcass CAT activity in p40W8 lines is similar to that of males and there is no marked ovary enhancement. Therefore the second region of the *I* factor promoter that contributes to elevated expression in the ovaries lies between position +40 and +101.

Many other genes are expressed in the nurse cell and their products inherited by the oocyte. Mutations in such maternally inherited genes are manifested, like I-R dysgenesis, as female sterility without affecting viability of either parent. Several of these defects have been characterisation at the genetic and morphological level (reviewed in Mahowald and Hardy, 1985) but relatively few have analysed at the molecular level. The possibility that the maternal effect genes share sequences with the *I* factor that direct nurse cell specific expression was investigated. The promoter of only one maternally inherited gene, *hunchback* that is expressed maternally and zygotically, has been analysed, however only those sequences controlling zygotic expression have been identified (Dreiver and Nüsslein-Volhard, 1989). Comparison of region +41 to +186 of *I* with the 5' region of the maternal affect genes for which there is sequence information (*Bicaudal D*, Wharton and Struhl, 1989; *torso*, Sprenger *et al.*, 1989; *hunchback*, Tautz *et al.*, 1987; *bicoid*, Berleth *et al.*, 1988; *vasa*, Lasko and Ashburner, 1988, Hay *et al.*, 1988; *oskar*, Ephrussi *et al.*, 1991; *staufen*, St. Johnston *et al.*, 1991), by dotplot analysis (UWGCG) failed to identify common sequence motifs, except A rich sequences that lie immediately 5' and 3' to position +100. The AT richness of the *Drosophila* genome means that the occurrence of such sequence in several maternally inherited genes is likely to be chance.

It was noted in chapter 3 that removal of the motif conserved between class II.3 promoters around position +30 reduced CAT expression two fold. The effect of removing this motif, is even less in flies than in tissue culture transfections, with no pronounced difference in the expression pattern or strength when compared with p40W8 lines. As argued in chapter 3, conservation of the +30 motif implies a function. As the sequence does not function in promoter strength it possibly functions in transcription start site selection which the model for transposition predicts to be crucial for *I* mobility. This hypothesis could be tested by primer extension analysis on transcripts from p28W8 transformants.

Table 5.5 summarises the results of this chapter.

TABLE 5.5

Class of transformant	ovary enhancement	pupae suppression	♂ suppression	♀carcass suppression	L3 activity
p186W8	++	++	++	++	-
p100W8	+	-	-	+	-
p40W8	-	-	-	-	-
p28W8	-	-	-	-	-
pΔ17W8	++	+	+/-	+	-

L3 = wandering third instar larvae

One intriguing observation is that CAT activity directed by all promoter fragments assayed is low in third instar larvae. There is as yet no evidence to suggest the reason for this, however one possible explaination is that third instar larvae might suppress *I* promoter activity or lack, or be deficient in a transcription factor(s) essential for *I* factor transcription. If this hypothesis is true, and for each promoter derivative, this function is mediated by the same protein, it would act through the sequences common to all promoters i.e. position +18 to +28. This sequence is CCGAAGAGATA. Like the other putative transcription factor binding motifs described in this thesis, this sequence does not resemble any known transcription factor binding sites (Wingender, 1988; Biggin and Tjian, 1989).

CHAPTER 6

6.1 INTRODUCTION:

The preceding two chapters demonstrated that the *I* factor promoter is involved in the ovary specificity of *I* transcription (and consequently transposition), and that the sequences conferring enhanced transcription in the ovaries lie between position +40 and +187. In addition to tissue specificity, *I* expression is regulated by cytotype, in that both transcription and transposition are highest in SF ovaries, lower in RSF ovaries, and barely detectible in inducer ovaries (Chaboissier *et al.*, 1990; Picard, 1976). This correlation of transcript levels and transposition frequency suggests that cytotype controls transposition at the level of transcription or RNA stability.

Cytotype regulation of *I* factor transcription and transposition is controlled directly or indirectly by an *I* factor encoded molecule. This was shown by P element transformation of a cloned *I* factor into a reactive strain. The resultant transformants evolved into strong inducer strains (Pritchard *et al.*, 1988). By switching cytotype from reactive to inducer, I-R hybrid dysgenesis offers a unique system in which to study autoregulation of transposition of a LINE-like element. This chapter examines the role that the *I* factor promoter plays in this.

6.2 RESULTS:

6.2.1 Construction of Isogenic Inducer Strains:

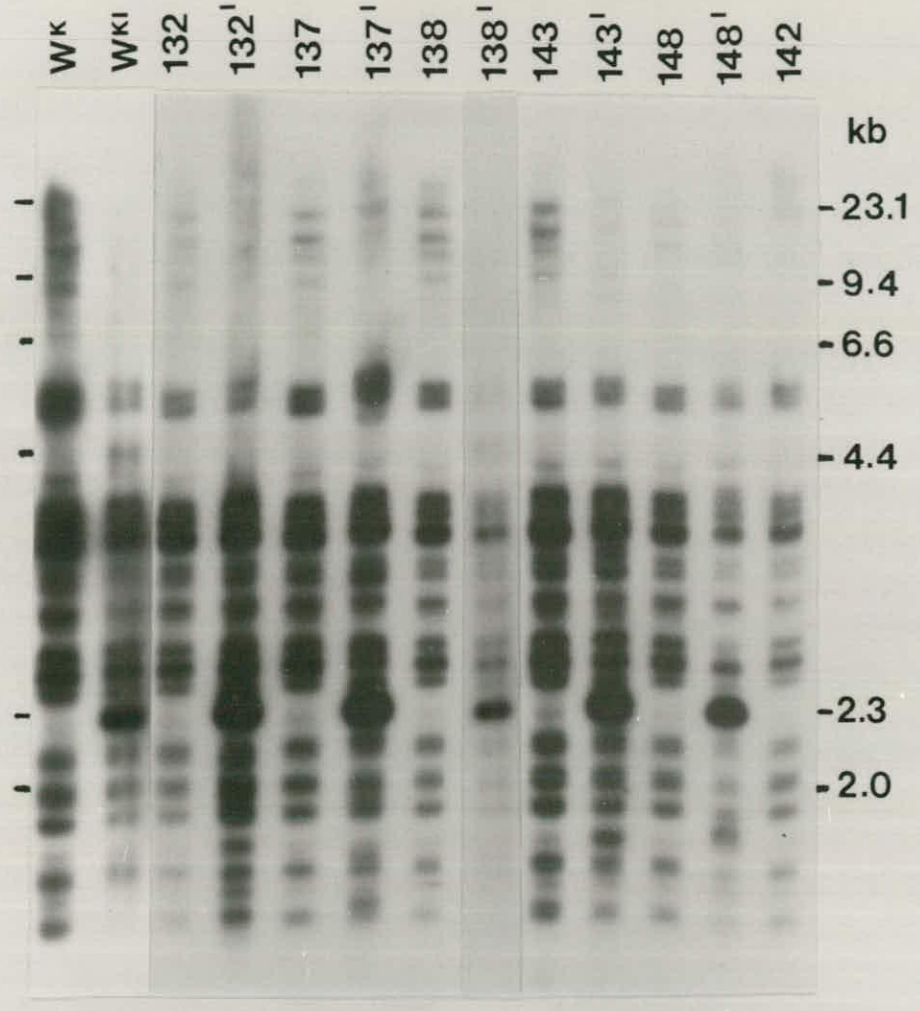
To examine the behaviour of the *I* factor promoter in an inducer cytotype, five of the p186W8 transformed lines, namely 186.132, 186.137, 186.138, 186.143, 186.148 and W^K were converted to inducer cytotype as described in materials and methods (section 2.3.4.8). The six strains have been designated 186.132^I, 186.137^I, 186.138^I, 186.143^I, 186.148^I and W^{KI}. The presence of complete *I* factors in the six lines was verified by Southern blot analysis (figure 6.1).

6.2.2 Comparison of the Developmental Expression from the *I* Promoter in Inducer and Reactive Lines:

Figures 6.2a-e illustrate graphically CAT activities for the five pairs of inducer and reactive lines corrected for the amount of protein used in the assay and figure 6.3

Figure 6.1 : **Southern blot analysis to determine the presence of *I* factors.** 3μg of DNA for each isogenic inducer and reactive line was digested with *Hind*III and *Pst*I, electrophoresed, Southern transferred, and hybridised with the 2.3kb *Hind*III/*Pst*I fragment of the *I* factor in plasmid pI407 (Bucheton *et al.*, 1984). Strong hybridisation to a 2.3kb fragment is diagnostic of inducer cytotype (Bucheton *et al.*, 1984) and is only seen in those lines that have been through a dysgenic cross. Numerous weaker bands of hybridisation are common to both reactive and inducer strains and are from defective pericentromeric *I* elements. Line 186.142 has been included in this Southern blot to verify its reactive state. The prefix '186' is omitted from each line.

Figure 6.1


W^K
W^KI
132
132'
137
137'
138
138'
143
143'
148
148'
142
kb
23.1
9.4
6.6
4.4
2.3
2.0

shows CAT assays of various developmental stages of males and females of the isogenic lines 186.143 and 186.143^{I}. In each pair of lines, the CAT activity of adult females does not increase after eclosion by the same magnitude as in reactive stocks. Female promoter activity is therefore switched down in an inducer cytotype. This effect is greatest for line 186.137^{I} in which adult female activity is less than adult males. In all other lines female CAT activity is between that of reactive females and males.

The lines differ only by the presence of *I* factors, therefore suppression of promoter activity must be causes by the *I* factors themselves or a molecule they encode. If the latter is true the degree of suppression may vary from line to line because of position effects on the *I* factors that modulate their ability to synthesise a regulatory protein.

Picard (1978c) demonstrated that lines in which the reactive genotype is reconstructed in the SF progeny (as in this case) take longer than 6 generations (but less than 100 generations) to become inducer (as measured by SF sterility). Lines 186.132^{I} and 186.142^{I}, were established for 8 generations when assayed, whereas the others were established for 18 generations. The inducer cytotype in some or all of these lines might not yet be fully established despite the presence of several copies of apparently intact *I* factors in all lines (figure 6.1). This could explain why CAT activity in adult females in most of the lines is not yet repressed to the level seen in males.

The alternative explanation for a reduction in CAT activity in the inducer cytotype is that the *I* factor promoter requires a transcription factor present in limiting amounts. The 10-15 functional *I* factors in inducer strains would therefore compete with the promoter on the transgene and reduce CAT activity.

Figure 6.2 : **Comparison of the developmental profile of CAT expression in inducer and reactive backgrounds in lines transformed with p186W8.** Graphs a-e illustrate the CAT activities of males and females of the inducer and reactive versions of lines 186.132, 186.137, 186.138. 186.143, and 186.148 respectively, normalised for the amount of protein. The y axis scale for each graph is different to emphasis the difference in CAT expression between reactive and inducer cytotypes. On the x axis: L3 - wandering third instar larvae; EP - early pupae; LP - late pupae; 0-1 day - 0-1 day after eclosion; 3-4 days - 3-4 days after eclosion.

Figure 6.2a 186.132

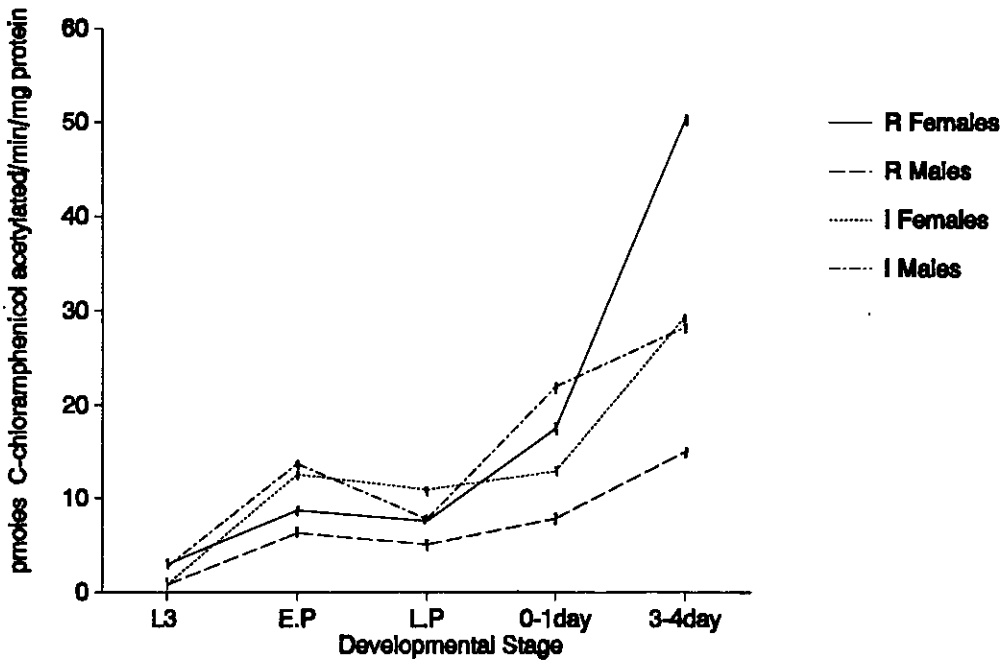


Figure 6.2b 186.137

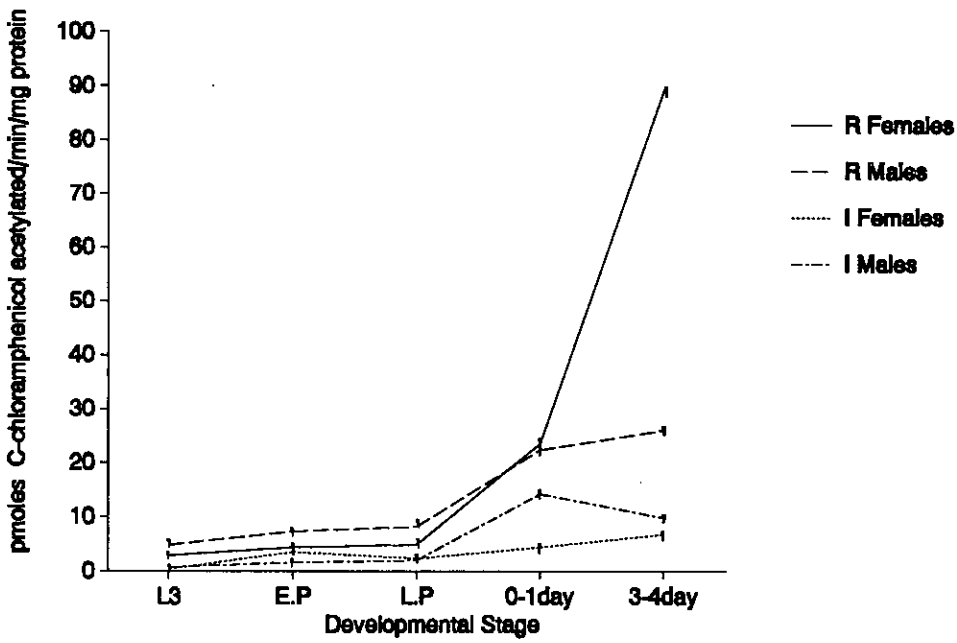


Figure 6.2c

186.138

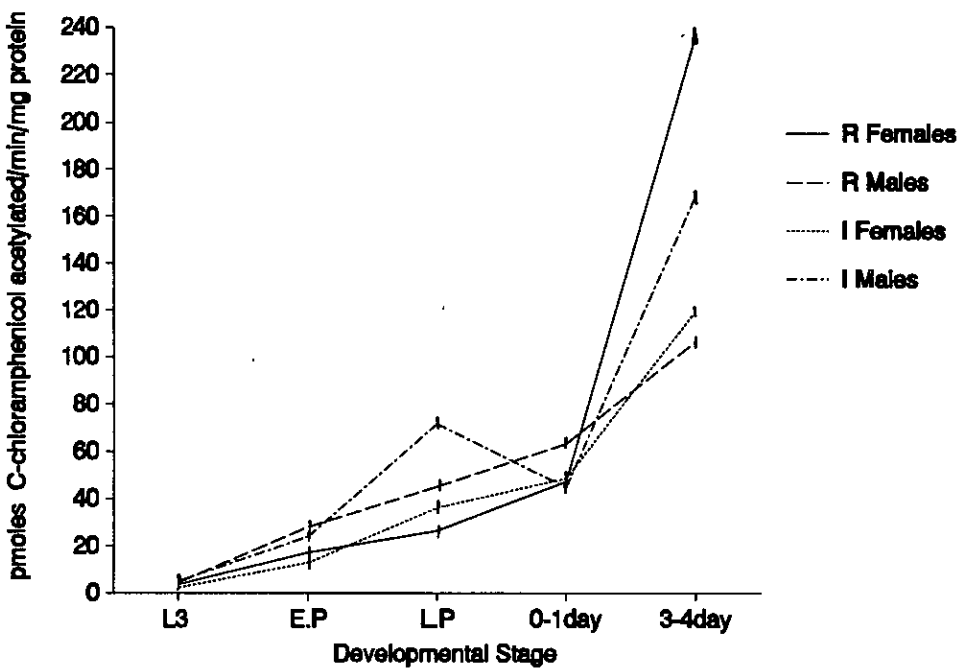


Figure 6.2d

186.143

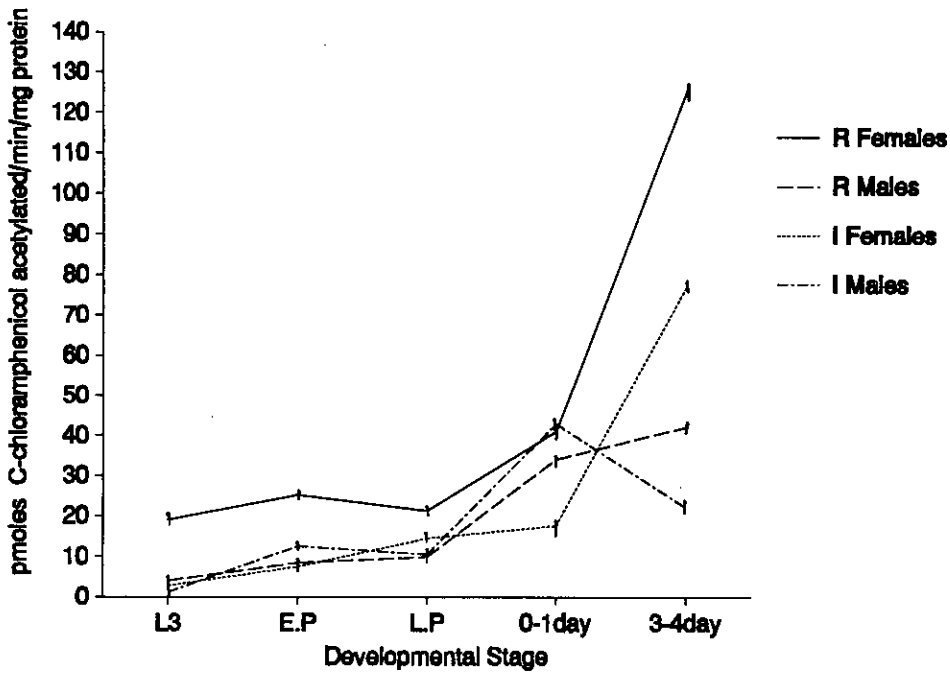


Figure 6.2e

186.148

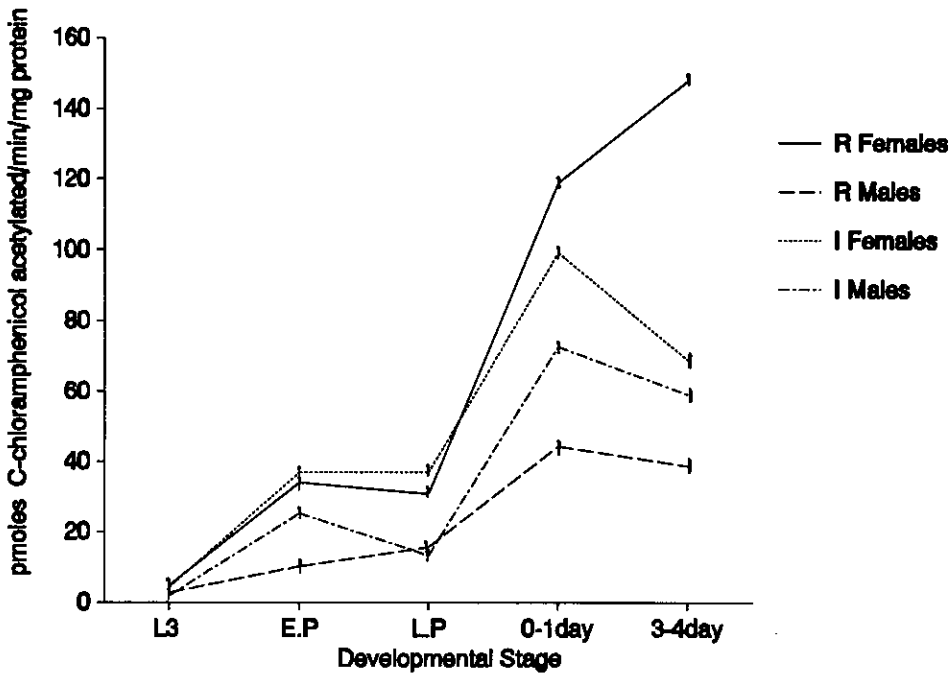


Figure 6.3 : **Comparison of CAT activity during development in lines 186.143 and 186,143^{I}.** The autoradiograph illustrates CAT assays of different developmental stages of males and females. L3 - wandering third instar larvae; EP - early pupae; LP - late pupae; 0-1 day - 0-1 day after eclosion; 3-4 days - 3-4 days after eclosion.

Figure 6.3

6.2.3 Comparison of CAT Activity in Ovary and Carcass Tissues from Inducer and Reactive Lines:

Ovaries were separated from carcass tissues in 3-4 day old inducer females and both samples were assayed for CAT activity. Figure 6.4 shows a CAT assays for lines 186.137 and 186.137^{I}, and 186.143 and 186.143^{I}. Table 6.1 lists the normalised values for all five lines.

CAT activity in the ovaries of inducer strain is less than in the isogenic reactive strain being reduced to a level similar to female carcass CAT activity which is similar in each cytotype. These observations are true for all lines except 186.132^{I} that has higher CAT activity in inducer ovaries than in carcass of either reactive or inducer lines. As argued above, inducer cytotype may not be fully established in this line, although some regulation has begun as CAT activity in 186.132^{I} ovaries is intermediate between that in reactive ovaries and female carcass of both cytotypes.

From the other four pairs of lines it can be concluded that absence of an increase in CAT activity after eclosion of inducer females is due to suppression of CAT expression in their ovaries. The inducer cytotype therefore specifically suppresses *I* promoter activity in ovary tissues to a level similar to that of males and female carcass.

TABLE 6.1

LINE	ovaries	carcass	ov:car
186.132	160*	15*	11:1*
186.132^I	92#	15#	6:1#
186.137	445 536*	19 1*	23:1 536:1*
186.137^I	16	4	4:1
186.138	229	50	4.6:1
186.138^I	84#	78#	1.1:1#
186.143	707 928*	61 55*	12:1 17:1*
186.143^I	55	34	1.6:1
186.148	684 1259*	109 72*	6:1 17:1
186.148^I	37	52	0.7:1

* = assays done approximately 36 generations post establishment, all other reactive assays done 20 generations post establishment.

= assays done 8 generations post establishment, all other inducer assays done 18 generations post establishment.

Values are expressed in pmoles ^{14}C-chloramphenicol acetylated/min/mg protein.

Figure 6.4 : **Comparison of CAT activity in ovary and carcass tissues in inducer and reactive cytotypes.** Assays of lines 186.137 and 186.143 are compared with those of lines 186.137^{I} and 186.143^{I} respectively. ♀ - female 3-4 days after eclosion; ♂ - male 3-4 days after eclosion; car - carcass from 3-4 day old females; ov - ovaries from 3-4 day old females.

Figure 6.5 : **CAT activity in SF and RSF flies.** The autoradiograph illustrates CAT assays of line 186.137 SF and RSF flies. Abbreviations are as above.

Figure 6.4

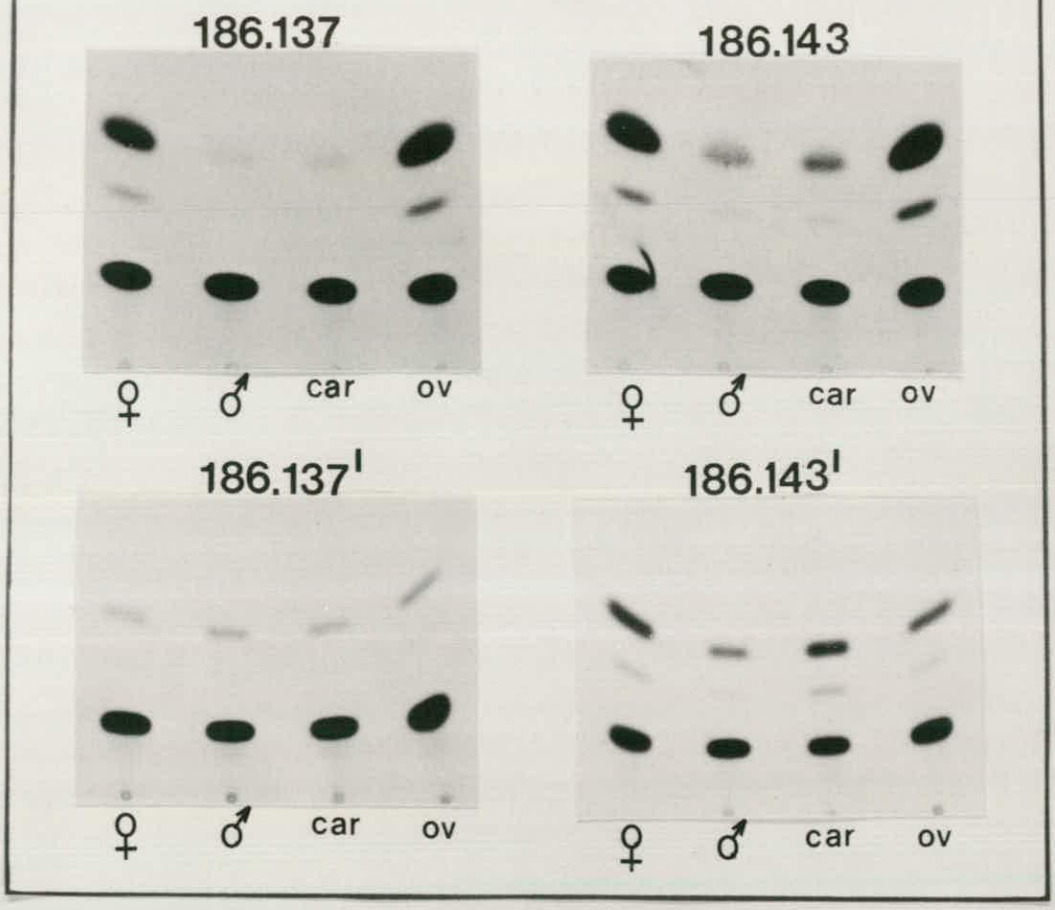
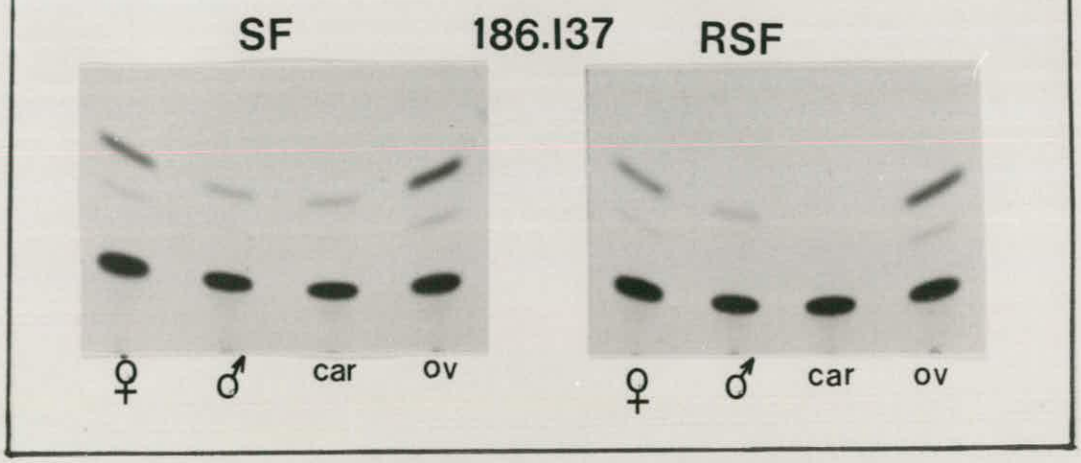


Figure 6.5



6.2.4 Behaviour of the *I* Factor Promoter in SF and RSF Flies:

Isogenic inducer and reactive flies of lines 186.137 and 186.148 were crossed in both directions to produce SF and RSF progeny that were assayed for CAT activity. The level of dysgenesis in each cross was determined by measuring the hatchability of eggs laid by SF and RSF females (see table 6.2).

TABLE 6.2

LINE	SF HATCHABILITY	RSF HATCHABILITY
186.137	4%	96%
186.148	1%	95%

Figure 6.5 shows the CAT activity of SF and RSF flies of line 186.137, CAT assays of the reactive and inducer parents are in figure 6.4. Tables 6.3a and 6.3b list the normalised CAT conversion results for lines 186.137 and 186.148 respectively.

TABLE 6.3a - 186.137

	REACTIVE	SF	RSF	INDUCER
♀ 3-4day	55	35	18	7
♂ 3-4day	7	20	8	10
carcass	1	8	1	4
ovaries	536	132	79	16

TABLE 6.3b - 186.148

	REACTIVE	SF	RSF	INDUCER
♀ 3-4day	129	192	127	69
♂ 3-4day	29	53	55	59
carcass	72	68	47	52
ovary	1260	327	202	37

Values are expressed in pmoles ^{14}C-chloramphenicol acetylated/min/mg protein.

A similar pattern of CAT expression is apparent in ovary and carcass tissues of both lines, with CAT levels being highest in reactive ovaries, four fold lower in SF ovaries, lower still in RSF ovaries, and similar to female carcass levels in inducer ovaries. In contrast, CAT activity in female carcass and males is similar in all four cytotypes. The difference in CAT levels between SF and RSF ovaries can only be due to an *I* factor autoregulatory protein that is maternally inherited, and not depletion of a limiting ovary specific transcription factor as SF and RSF individuals have an equivalent number of *I* factors. The first 186 nucleotides of *I* therefore contain sequences that are the target for autoregulation of *I* expression in inducer lines, and on which the maternally inherited regulator in RSF ovaries acts. Higher CAT levels in reactive compared with SF ovaries suggests regulation of *I* transcription is already underway in 3-4 day old SF females. The difference in CAT activity between SF and RSF ovaries is only twofold whereas transposition in RSF females is five times less than in SF females. The reason for this discrepancy shall be discussed in the following section.

6.2.5 Behaviour of the *I* Factor Promoter when Inherited Maternally or Paternally:

The above experiment demonstrates that the *I* promoter is influenced by the maternal effect of I-R dysgenesis but it does not distinguish between the activity of the paternally and maternally inherited chromosomes in SF and RSF flies. To determine the behaviour of the *I* factor promoter when inherited from either parent, strains 186.137 and 186.137^{I} were crossed with W^{K} and W^{KI} as follows, and the CAT activity of male and female progeny, and ovary and carcass tissues measured:

CROSS		CYTOTYPE of PROGENY
1: W^{K}♂ x 186.137♀	-	reactive
2: 186.137♂ x W^{K}♀	-	reactive
3: W^{KI}♂ x 186.137♀	-	SF
4: 186.137^{I}♂ x W^{K}♀	-	SF
5: W^{K}♂ x 186.137^{I}♀	-	RSF
6: 186.137♂ x W^{KI}♀	-	RSF
7: W^{KI}♂ x 186.137^{I}♀	-	inducer
8: 186.137^{I}♂ x W^{KI}♀	-	inducer

Figure 6.6 shows CAT assays from the eight crosses and table 6.4 catalogues the values for CAT activity corrected for protein concentrations. In each sample assayed there is only one copy of the transgene instead of two as in the previous assays.

One would predict that in a reactive background CAT levels directed by one transgene relative to two should drop by a maximum of 50%. Surprisingly, in crosses 1 and 2 a five fold reduction in ovary CAT activity is observed relative to homozygous line 186.137 (see table 6.1). The reason for this drop is unknown as both lines have the same W^{K} reactive background. Nevertheless the pattern of expression is similar, with ovary expression 60-83 times higher than female carcass activity. Cross 2 demonstrates that the *I* factor promoter is activated when inherited paternally into a reactive female cytotype, consistent with the idea that the *I* factor promoter is activated when introduced paternally into an SF female thus facilitating transposition and triggering I-R hybrid dysgenesis. Crosses 3 and 4 that produce SF progeny both

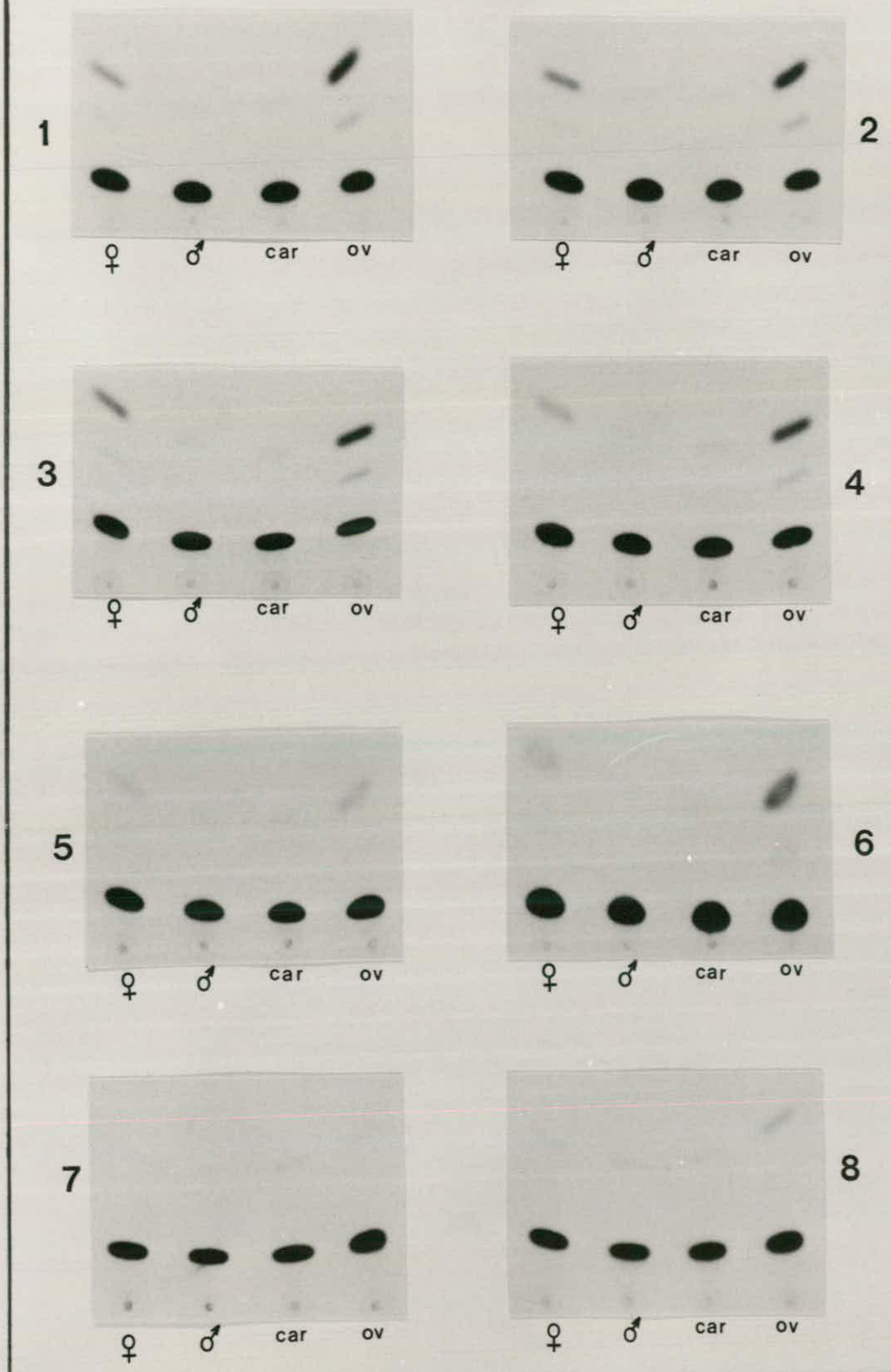
TABLE 6.4

CROSS	♀3-4day	♂3-4day	carcass	ovary	ov:car
1	10	1	1.5	92	60:1
2	13	1.5	1	83	83:1
3	22	1	1	70	70:1
4	17	3	1	66	66:1
5	6	1	1	12	12:1
6	10	1	1	31	31:1
7	1	1	1	2	2:1
8	2	1.5	1	10	10:1

Values are expressed in pmoles ^{14}C-chloramphenicol acetylated/min/mg protein.

Figure 6.6 : **CAT assays of flies that have inherited the p186W8 transgene paternally or maternally.** The number of each cross is given at the side of each autoradiograph. ♀ - female 3-4 days after eclosion; ♂ - male 3-4 days after eclosion; car - carcass of 3-4 day old females; ov - ovaries from 3-4 day old females.

Figure 6.6



give CAT activity in the ovaries at a level slightly lower than in crosses 1 and 2 indicating that regulation is underway. When inherited paternally (crosses 2 and 4) the *I* factor promoter directs slightly lower ovary CAT activity than in the corresponding cross when the transgene is maternally inherited (crosses 1 and 3). The reason for this is unknown, however it was observed in chapter 4 that the ovary CAT activity of p186W8 lines increases slightly with the length of time the line has been established. A promoter that is weakly expressed in males might therefore not be transcribed at its full potential when inherited by a daughter. This implies that the promoter activity detected from the first 186bp of *I* in long established reactive transformants is higher than that ever directed in SF females.

Crosses 5 and 6 produce RSF progeny. RSF ovary CAT activity is suppressed two fold when the transgene is inherited paternally (cross 6) relative to the corresponding SF ovary activity in cross 4, and six fold less when inherited maternally (cross 5) relative to the maternally inherited transgene in SF ovaries (cross 3). Two complementary hypotheses can explain this maternal effect:

1. The inducer cytotype in W^{KI} is weaker than in 186.137^{I} and is less able to regulate *I* transcription in RSF ovaries.

2. The *I* factor regulatory molecule does not function efficiently in *trans*.

The results of crosses within an inducer cytotype show that the *I* promoter is activated slightly in the ovaries of inducer females when W^{KI} donates the maternal cytotype (cross 8), but not when 186.137^{I} donates the maternal cytotype (cross 7). At the time of the experiment W^{KI} was established for 10 generations and 186.137^{I} for 20. The inducer cytotype in W^{KI} might not be established and therefore unable to regulate *I* promoter activity fully. If so, slight activation of promoter activity would be expected in crosses 6 and 8.

To confirm this hypothesis one could use long established isogenic strong inducer and reactive lines instead of W^{KI} and W^{K} in crosses 1 to 8.

The delay in completion of inducer cytotype, as measured by *I* factor promoter suppression, may be because the efficiency of the repressor molecule in *trans* is not 100% (interpretation 2). However the repressor does function to some degree in *trans*

as the level of activation in cross 8 is much less than in when the *I* factor promoter is inherited paternally into a reactive cytotype (cross 4). A requirement for regulation of paternally inherited *I* factors in *trans* implies that these elements do not already carry a regulatory molecule. This is possible as the *I* factor promoter is weak and transcripts undetectable in males. Instead, transposition in males could be kept in check because they lack the ovary specific transcription factor(s) that facilitates high levels of synthesis of the transposition intermediate. Introduction of an *I* factor promoter paternally into a zygote would therefore rely on maternally inherited regulatory molecules in the oocyte to switch off/down the incoming *I* factor promoter. Inefficient *trans* regulation would result in activation of paternally inherited promoters in the germline (as in crosses 6 and 8). Complete regulation would require a critical concentration of the regulator which could take many generations to reach and thus explain perpetuation *I* factor promoter activity and SF sterility in the descendants of SF flies.

This model as it stands predicts that *I* factor transposition would occur in the germline of inducer and RSF females at the same rate because their oocyte cytoplasm is diluted to the same degree. However inducer flies have twice as many *I* factors as RSFs, which may make sufficient regulator in the germline of the inducer female to reduce transcription below the threshold prerequisite for detectable transposition.

Higher *I* factor promoter activity in the ovaries of RSF females when the transgene is inherited paternally than maternally explains the unexpectedly high level of CAT activity detected in the ovaries of RSF females in the previous section (6.2.4). In these RSF females a transgene was inherited from both parents, but only the maternally inherited copy would be repressed to a level representative of the RSF cytotype (six fold). The paternally inherited transgene would be repressed around two fold. The CAT activity in RSF ovaries in section 6.2.4 is therefore a summation of these two values and is not representative of *I* factor expression in RSF females which never inherit *I* factors from there fathers.

6.3 DISCUSSION:

The results presented in this chapter demonstrate that the inducer cytotype reduces the level of *I* factor transcript, and hence transposition, by suppressing promoter activity. This regulation acts on the first 186bp of *I* and functions only in the ovaries, reducing promoter strength to a level similar to that in males and female carcass tissues.

The isogenic nature of the lines means suppression of *I* promoter activity is mediated by an *I* factor encoded protein or RNA. A candidate for the regulatory molecule is the product of ORF1. This protein has both DNA and RNA binding properties, however experimental evidence to date suggests binding is not specific for *I* sequences (T. Paterson, personal communication). ORF1 could theoretically possess a high nonspecific affinity for nucleic acids but bind *I* factor sequences with slightly greater affinity. Such properties have been reported for the *P* element transposase that binds its target sequence at the *P* element ends with only 10 times more greater affinity than other sequences (Kaufman *et al.*, 1989). Poor target sequence specificity could mean that ORF1 operates by binding nucleic acid in close proximity as it is translated, that is *I* factor RNA. Upon entry into the nucleus and integration, ORF1 protein may bind the new *I* factor DNA copy and suppress further transcription. This mechanism complies with the hypothesis that the molecules conferring *I* cytotype do not function well in *trans*.

Arkhipova (personal communication) looked for proteins particular to inducer ovaries that bind the first 186 nucleotides of *I* on band shift assays but none were found. As discussed above, binding of the *I* factor autoregulator may be preferentially in *cis*, therefore it is not surprising that no inducer specific bands were seen in gel shift assays.

Antibodies raised against oligo peptides of the carboxyl terminus of ORF1 detected several proteins in extracts from SF and RSF developmental stages in Western blot analysis, but none were more abundant in RSF females (Lynch, 1989). The reason for this is unclear. However analysis of the results was complicated by interference from antibodies present in preimmune serum that cross hybridise with

Drosophila proteins. Work is in progress to affinity purify these antibodies further, which should make detection of the ORF1 protein more sensitive.

CHAPTER 7

7.1 GENERAL DISCUSSION:

The work presented in this thesis has identified and characterised a promoter in the 5'UTR of the *I* factor that functions in tissue culture cell and in *D. melanogaster* flies. This promoter has been shown to be internal to the transcription unit, and initiates transcription at position +1 of the *I* factor. These findings are consistent with the model for *I* factor transposition that predicts an internal promoter directs transcription of the *I* factor transposition intermediate. Within the promoter are regions that modulate expression in a tissue and developmental specific fashion that depends upon the inducer-reactive cytotype.

7.1.1 A Model for *I* Transcription in Reactive and Inducer Flies:

A model depicting the events that lead to transcription initiation from the *I* factor minimal promoter was presented in chapter 3 (section 3.5). This has been modified in figure 7.1 to include the interaction of *I* factor sequences with the putative transcription factors that mediate tissue specificity, developmental control, and cytotype regulation revealed by the analyses of transformed flies.

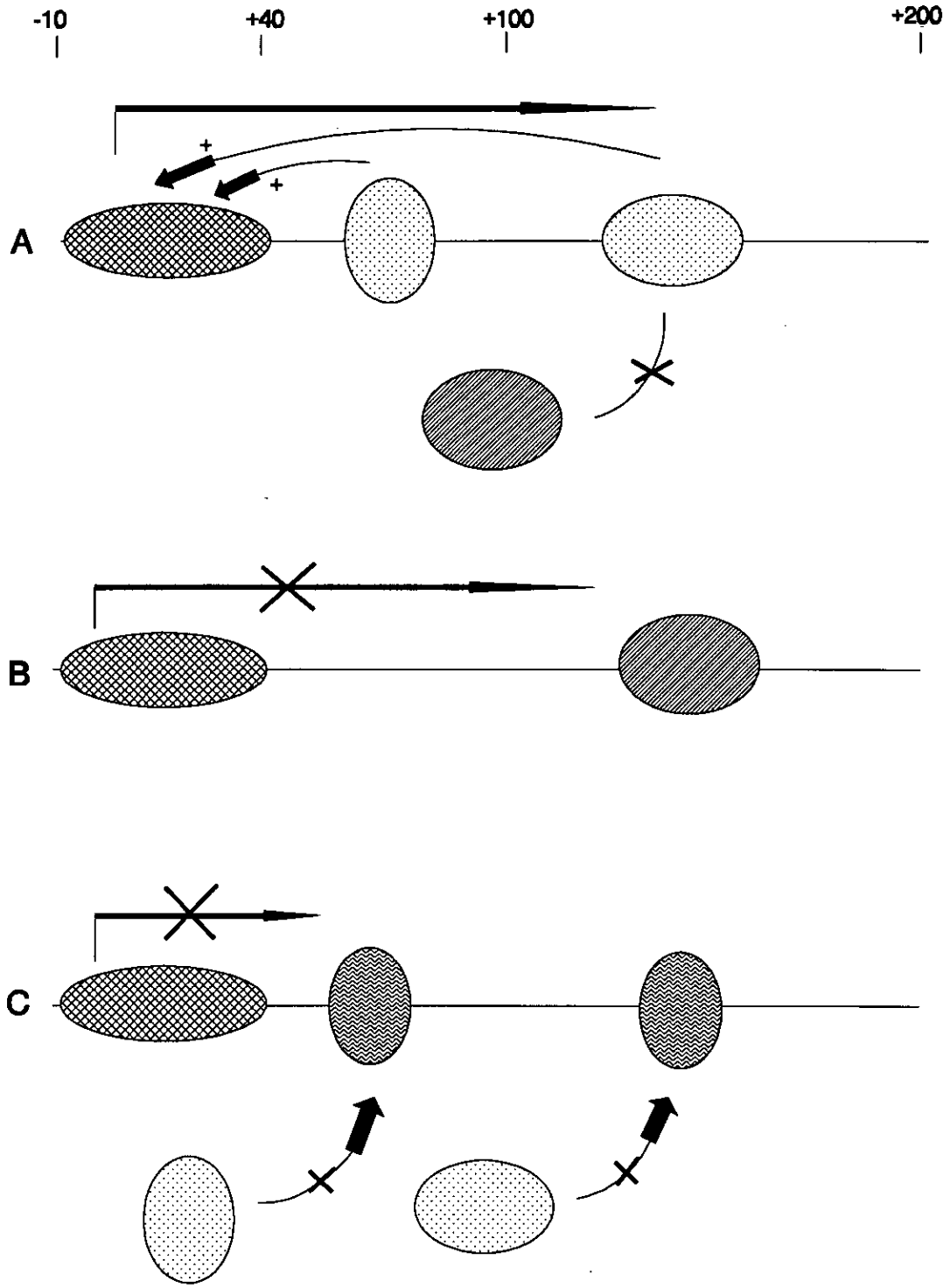
The first step in transcription is assembly of the RNA polymerase II initiation complex around sequences +1 to +37. In the context of nucleotides +1 to +186 in a reactive cytotype, high promoter activity is directed only in ovaries. This pattern of expression can be explained by two mechanisms that are not mutually exclusive:

1. Ovary specific transcription factor(s) activate expression.
2. In all tissues other than ovaries transcription is suppressed by a *trans* acting protein.

Deletion of nucleotides +101 to +186 resulted in an increase in expression in pupae, and adults of both sexes, and reduced expression in ovaries although some ovary enhancement is retained (section 5.2.3). This region of *I* therefore contains sequences that reduce promoter activity in males and females, and increase expression in ovaries. Position +135 to +152 of the *I* factor bind a protein found in pupae, adults and ovaries. If this binding activity represents the suppressor protein, its presence in ovaries suggests that the ovary specific transcription activator competes with the

Figure7.1 : **Possible mechanims of *I* factor promoter regulation in different tissues and cytotypes.** (A) depicts a possible mode of transcription regulation in reactive ovaries, (B) a model for all other reactive tissues, and (C) a model for regulation in inducer ovaries. The numbers indicate the distance relative to the transcription initiation site at +1. In (A), + signifies transcription activation. The horizontal arrows indicates transcription. X indicates blockage of that event.

Figure 7.1


-10
+40
+100
+200
A
B
C
Initiation complex
transcription suppressor
ovary specific trans acting transcription activator
I factor autoregulatory protein

repressor for the same binding site.

A second region that confers activation of expression in ovaries lies between position +40 and +101 as transformants with the first 40bp of *I* lack enhanced ovarian CAT expression.

Figure 7.1 depicts the interactions of transcription factors with the *I* factor promoter in reactive ovaries (figure 7.1a), and all other reactive tissues (figure 7.1b).

A model for *I* factor autoregulation in an inducer cytotype is presented in figure 7.1c. The target site of the repressor is unknown. However as discussed in chapter 6, the principal candidate is ORF1 which has high nonspecific nucleic acid binding properties but may have, as yet undefined, some sequence specificity. Figure 7.1c depicts ORF1 binding and blocking interaction of the ovary specific transcription activators. Alternatively ORF1 might prevent transcription by blocking RNA polymerase II progress by virtue of a strong interaction between it and its target DNA.

No RNA analysis has been performed on transformed lines. It is therefore possible that modulation of expression from the *I* factor 5'UTR (as measured by CAT activity) is through RNA stability or regulation of translation.

7.1.2 Implications for Regulation of I-R Hybrid Dysgenesis:

Regulation of I-R hybrid dysgenesis (as measured by sterility) takes several generations and is thought to be secured when a threshold number of *I* factors is reached (Finnegan, 1989). Evidence in support of this theory is weak as it is based on the slight difference in the number of euchromatic *I* elements in fertile and poorly fertile descendants of SF females (Pélisson and Bregliano, 1987) as measured by *in situ* hybridisation which cannot distinguish between functional and defective elements.

Analysis of the *I* factor promoter in tissue culture and reactive flies demonstrated that the strength of *I* factor expression is influenced both by the identity of nucleotide +3 and its chromosomal location. Consequently the *I* factor's ability to establish transposition regulation, presumably by synthesise an autoregulatory

molecule, may vary from one inducer strain to another. Such variation was found in chapter 6 when comparing the degree promoter suppression between different inducer strains relative to their isogenic reactive strain. It is therefore possible that establishment of the inducer cytotype occurs when a threshold level of an autoregulatory molecule is reached. This will be influenced by insertion of *I* factors at suitable chromosomal sites, the identity of nucleotide +3, and the number of *I* factors.

Once the inducer cytotype is secured, *I* transcription rate (or RNA stability) will be reduced to the low level seen in inducer strains. Weak *I* factor transcription in inducer lines is probably required to maintain the repressor levels. Transposition might be prevented because translation of ORF2 (the putative transposase) is inefficient.

This hypothesis explains perpetuation of SF sterility in the descendants of SF females when either the inducer or reactive genotype is recovered in their daughters, because the threshold level of the *I* factor regulator may take several generations to be reached.

Variation in the strength of reactivity is thought to be influenced by the *I* elements located at the chromocentre as they are transcribed and contain large regions that hybridise to ORF1 sequences (Chaboissier *et al.*, 1990; Crozatier *et al.*, 1988). Sequence analysis of four *I* elements from a strong reactive strain however, showed the ORF1 region is incapable of encoding a full length protein (Vaury *et al.*, 1990). In two of the *I* elements sequenced nucleotides +1 to +37 are identical to the *I* factor and the region from position +38 to +186 has only 9 and 10 position changes, 7 of which are shared between the elements. Some *I* elements therefore have functional promoters from which transcription may take place at a high level in ovaries. If the promoter is involved in reactivity it must act through differences in strength, due to position effect, or sequence variation. One could imagine that *I* element promoters or their transcripts bind the regulatory protein - in a strong reactive cytotype there would be many *I* elements or transcripts that bind and sequester the regulator, thus preventing

early repression of transposition. Conversely, in a weak reactive cytotype, the promoters would be more divergent or under a suppressive position effect resulting in less homologous or rarer transcripts that do not interfere with the onset of transposition regulation. Although this theory can explain differences in reactivity strength from strain to strain and persistence of SF sterility in their descendants, it does not account for persistence of reactivity levels dictated by the maternal line several generations after the genotype has been replaced by reactive chromosomes of a different strength (Bucheton and Picard, 1978). An alternative explanation is that *I* elements from strong reactive strains cannot encode a functional regulator (the *I* elements sequenced by Vaury *et al.*, 1990 were from such a strain), and those from weak reactive strains can. Persistence of maternal reactivity strength may be due to positive autoregulation of the *I* factor transposition repressor.

7.1.3 Is the *I* factor Promoter and its Regulation Representative of Other LINEs?

The *I* factor is unique in that its transcription and transposition can be activated and suppressed by means of the I-R system of hybrid dysgenesis. The structural similarity between the *I* factor and other LINE-like elements suggests that they share aspects of regulation. However those elements that have been studied in detail questions whether one can extrapolate regulation mechanisms from the *I* factor to other LINEs. For example, apparent full length transcripts of the *D. melanogaster* element, *jockey*, have been found in all developmental stages tested and among different strains (Mizrokhi *et al.*, 1988). Yet these strains do not display higher than normal levels of somatic mutation. Therefore, unlike *I*, the presence of full length *jockey* transcripts does not correlate with transposition.

Expression of the human LINE *L1Hs* is similar to the *I* factor in that a tissue specific promoter confines high levels of full length transcripts to the teratocarcinoma cell line Ntera2D1 (Swergold, 1990). However correspondingly high transposition frequencies have not been reported in these cells despite association of reverse transcriptase activity with virus-like particles that contain *L1Hs* RNA (Deragon *et al.*, 1990). The *L1Hs* ORF1 protein has been detected in Ntera2D1 cells which, if it

functions as a repressor, could inhibit transposition, however the presence of full length *L1Hs* transcripts in these cells implies repression is post transcriptional.

7.1.4 Future Work:

Much of the work that remains to be done involves analysis of *I* factor expression at the RNA level. For example one could estimate the level of transcription from each promoter fragment by Northern blot analysis, primer extension, or S1 mapping to determine whether the CAT activity detected in tissue culture and flies is representative of transcription levels.

Primer extension analysis of transcripts synthesised *in vitro* demonstrated that a *Drosophila* nuclear extract directs transcription of the *I* factor promoter from position +1. However, to confirm the hypothesis that the CAT activity detected in ovaries is directed from the promoter that initiates transcription of the putative transposition intermediate, the 5' end of ovary transcripts should be mapped by primer extension or S1 protection analysis.

The role of the conserved promoter motif around position +30 has not been shown to have any significance as yet. It was proposed in chapter 3 that this motif might function in start site selection. This could be verified by mapping the 5' end of transcripts that initiate from the p28W8 transformants or p28T in tissue culture.

The 5' and 3' endpoints that define the *I* factor promoter also remain to be determined precisely, as at present the 5' limit is between position +1 and +5 and the 3' limit between +12 and +28. Construction of expression vectors that span these points could determine the promoter limits by the techniques used in this thesis or by RNA analysis.

The site of action of the inducer cytotype regulatory molecule has not yet been mapped. Conversion to the inducer cytotype of transformants carrying the p100W8 and the p40W8 expression vectors is underway and CAT assays shall be performed on various developmental stages once the lines are established.

From deletion analysis it appears that several positively and negatively acting transcription factors are involved in *I* factor expression. Identification of those involved in formation of the transcription initiation complex at the minimal promoter was discussed in chapter 3. Transcription factors that play a role in the ovary specificity of *I* expression could be identified by firstly defining more precisely the sequences to which they bind by synthesis of further promoter deletion derivatives, or determining which sequences specifically compete with promoter/ovary protein complex formation on band shift assays. An oligonucleotide of that sequence could then be used to screen an ovary specific expression library, or be used for purification of these factors by affinity chromatography. In both approaches the ovary transcription factor that interacts with the *I* factor promoter should bind the oligonucleotide and, in the first case lead directly to the gene and, in the second case lead indirectly to the gene via the protein. A similar approach could be used to isolate the protein(s) that suppress *I* expression in pupae.

APPENDICES

APPENDIX A:

Sequence of Nucleotides 1 to 186 of the *I* Factor:

(Taken from Fawcett *et al.*, 1986)

CATTACCACT TCAACCTCCG AAGAGATAAG TCGTGCCTCT CAGTCTAAAG
CCTCGCTTCG CGTAAGCCCA AAACTCTTAT CAGCAAAATC TTGATAAACA
AATATCAACC ACAAAGAGAA AATAAAAAAC TTAACAACAA AAACAACAAT
ACCGCTAATC CGGGCTCAAG CCCTTAACCA ACAATC

APPENDIX B:

Sequence of the Promoter Fragment in p40Tλ:

I factor sequences are shown in normal typeface; λcI857 sequences are shown in bold; junctions created by ligation of the *Bam*HI site immediately downstream of the *I* promoter to the *Sau*3A in λ are underlined.

λ sequences extend from position 33327 to 33206 on the minus strand (Daniels *et al.*, 1983).

CATTACCACT TCAACCTCCG AAGAGATAAG TCGTGCCTCT **GGATCAAACA**
CTTATGGCTA TCCAGACTAA ATTCACTATC GCCACTTTTA TTGGCGATGA
AAAGATGTTT CGTGAAGCCG TCGACGCTTA TAAAAAATGG ATATTAATAC
TGAAACTGAG ATCC

APPENDIX C:

Mapping Data for Transformed Lines:

LINE	CHROMOSOME TO WHICH TRANSGENE MAPS
CAT.187	3
CAT.188	X
CAT.192	2
CAT.197	3
CAT.198	3
186.132	3
186.137	3
186.138	2
186.142	X
186.143	X
186.148	X
100.178	3
100.179	2
100.182	2
100.183	X
100.184	3
40.177	2
40.185	3
28.165	3
28.166	3
28.167	3
Δ17.162	3
Δ17.163	3
Δ17.164	X

REFERENCES

Abad, P., Vaury, C., Pélisson, A., Chaboissier, M.-C., Busseau, I., and Bucheton, A., (1989) A long interspersed repetitive element - the *I* factor of *Drosophila teissieri* - is able to transpose in different *Drosophila* species. Proc. Natl. Acad. Sci. USA, **86**: 8887-8891

Ahearn, J.M., Jr., Bartolomei, M.S., West, M.L., Cisek, L.J., and Corden, J.L., (1987) Cloning and sequence analysis of the mouse genomic locus encoding the largest subunit of RNA polymerase II. J. Biol. Chem., **262**: 10695-10705

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

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

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., Mazo, A.M., Cherkasova, V.A., Gorelova, T.V., Schuppe, N.G., and Ilyin, Y.V., (1986) The steps of reverse transcription of *Drosophila* mobile dispersed genetic elements and U3-R-U5 structure of their LTRs. Cell, **44**: 555-563

Back, E., Meir, E.V., Muller, F., Schaller, D., Neuhaus, H., Aeby, P., and Tobler, H., (1984) Intervening sequences are of the ribosomal RNA genes of *Ascaris lumbricoides*: DNA sequences at junctions and genomic organisation. EMBO J., **3**: 2523-2529

Ballario, P., Filetici, P., Junakovic, N., and Pedone, F., (1983) Ty1 extrachromosomal circular copies in Saccharomyces cerevisiae. FEBS Lett., **155**: 225-229

Beaupain, D., Eléouët, J.F., and Roméo, 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. Nuc. Acids Res., **18**: 6509-6515

Belcourt, M.F., and Farabaugh, P.J., (1990) Ribosomal frameshifting in the yeast

retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. Cell, **62**: 339-352

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

Benoist, C., and Chambon, P., (1981) The SV40 early promoter region: the sequence requirements *in vivo.* Nature (London), **290**: 304-310

Berleth, T., Burri, M., Thoma, G., Bopp, P., Richstein, S., Frigerio, G., Noll, M., and Nüsslein-Volhard, C., (1988) The role of *bicoid* RNA in organizing the anterior pattern of the *Drosophila* embryo. EMBO J., **7**: 1749-1756

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

Biggin, M.D., and Tjian, R., (1989) Transcription factors and the control of *Drosophila* development. TIG, **5**: 377-383

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

Bingham, P.M., and Zachar, Z., (1989) Retrotransposons and the foldback transposon from *Drosophila melanogaster.* In 'Mobile DNA', Eds. Berg and Howe, **17**: 485-502

Bird, A.P., (1980) DNA methylation and the frequency of CpG in animal DNA. Nuc. Acids Res., **8**: 1499-1504

Bird, A.P., (1986) CpG-rich islands and the function of DNA methylation. Nature (London), **321**: 209-213

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

Blackman, R.K., and Gelbert, W.M., (1989) The transposable element *hobo*, in 'Mobile DNA', eds. Berg, D.E., and Howe, M.M., Ch.**20**: 523-529

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

Blackman, R.K., Macy, M., Koehler, D., Grimaila, R., and Gelbart, W., (1989) Identification of a fully functional *hobo* transposable element and its use for germ-line transformation of *Drosophila*. EMBO J., **8**: 211-217

Boeke, J.D., Garfinkel, D.J., Styles, C.A., and Fink, G.R., (1985) Ty elements transpose through an RNA intermediate. Cell, **40**: 491-500

Bolivar, 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

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

Bregliano, J.C., and Kidwell, M.G., (1983) Hybrid dysgenesis determinants. In 'Mobile Genetic Elements', J.A. Shapiro, ed., (London Acad. Press Inc.) p363-410

Brierley, C., and Flavell, A.J., (1990) The retrotransposon *copia* controls the relative levels of its gene products post-transcriptionally by differential expression from its two major mRNAs. Nuc. Acids Res., **18**: 2947-2951

Briggs, D., Jackson, D., Whitelaw, E., and Proudfoot, N.J., (1989) Direct demonstration of termination of signals for RNA polymerase II from the sea urchin H2A histone gene. Nucl. Acids Res., **17**: 8061-8071

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. Nuc. Acids Res., **12**: 9111-9122

Bucheton, A. (1973) Contribution à l'étude de la stérilité femelle non mendélienne chez *Drosophila melanogaster*. Transmission hereditaire des degres d'efficacite du facteur 'reacteur'. Comptes Rendus Ac. Sc. Paris D. **276**: 641-644

Bucheton, A., (1978) Non-mendelian female sterility in *Drosophila melanogaster*: influence of aging and thermic treatments. Heredity, **41**: 357-369

Bucheton, A., (1979a) Non-mendelian female sterility in *Drosophila melanogaster*: influence of aging and thermic treatments. III. Cumulative effects induced by these

factors. Genetics, **93**: 131-142

Bucheton, A., (1979b) 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, Ph., (1976) Non-mendelian female sterility in *Drosophila melanogaster*: quantitative variation in 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**: 207-223

Bucheton, A., Simonelig, M., Vaury, C., and Crozatier, M., (1986) Sequences similar to the *I* transposable element involved in I-R hybrid dysgenesis in *D. melanogaster* occur in other *Drosophila* species. Nature, **322**: 650-652

Buratowski, S., Hahn, S., Sharp, P.A., and Guarente, L., (1988) Function of a yeast TATA element binding protein in a mammalian transcription system. Nature (London), **334**: 37-42

Buratowski, S., Hahn, S., Guarente, L., and Sharp, P.A., (1989) Five intermediate complexes in transcription initiation by RNA polymerase II. Cell, **56**: 549-561

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

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., (1989a) *I* elements of *Drosophila*

melanogaster generate specific chromosomal rearrangements during transposition. Mol. Gen. Genet., **218**: 222-228

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

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

Carcamo, J., Buckbinder, L., and Reinberg, D., (1991) The initiator directs the assembly of a transcription factor IID-dependent transcription complex. Proc. Natl. Acad. Sci. USA, **88**: 8052-8056

Carcamo, J., Maldonado, E., Cortes, P., Ahn, M-H., Ha, -I., Kasai, Y., Flint, J., and Reinberg, D., (1990) A TATA-like sequence located downstream of the transcription initiation site is required for expression of an RNA polymerase II transcribed gene. Gen. Dev., **4**: 1611-1622

Carlou, M.-L., (1987) Biochemical phylogeny of the eight species in the *Drosophila melanogaster* subgroup, including *D. sechellia* and *D. orena*. Genet. Res. Camb., **50**: 181-185

Cavallini, B., Huet, J., Plassat, J., Sentenac, A., Egly, J., and Chambon, P., (1988) A yeast activity can substitute for the Hela cell TATA box factor. Nature (London) **334**: 77-80

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 melanogaster*. 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

Chapman, K.B., and Boeke, J.D., (1991) Isolation and characterisation of the gene encoding yeast debranching enzyme. Cell, **65**: 483-492

Chen, W., and Struhl, K., (1985) Yeast mRNA initiation sites are determined

primarily by specific sequences, not by the distance from the TATA element. EMBO J., 4: 3273-3280

Chretien, S., Dubart,A., Beaupain, D., Raich, N., Grandchamp, B., Rosa, J., Goossens, M., and Romeo, P.-H., (1988) Alternative transcription and splicing of the human porphobilinogen deaminase gene result either in a tissue-specific or in housekeeping expression. Proc. Natl. Acad. Sci., USA, 85: 6-10

Coffin, J.M., (1985) In 'RNA tumour viruses'. Eds. Weiss, R., Teich, N., Varmus, H.E., and Coffin, J. (Cold Spring Harbour Lab., CSH, NY), 2nd edn. p17-74

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

Cormack, B.P., Strubin, M., Ponticelli, A.S., and Struhl, K., (1991) Functional differences between yeast and human TFIID are locatized to the highly conserved region. Cell, 65: 341-348

Courey, A.J., and Tjian, R., (1988) Analysis of Sp1 *in vivo* reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. Cell, 55: 887-898

Crouse, J., and Amores, D., (1986) Double digestions of the multiple cloning site. Focus, 8: no.3, p9

Crouse, G.F., Leys, E.J., MeEwan, 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., Vaury, C., Busseau, I., Pélisson, A., and Bucheton, A., (1988) Structure and genomic organisation of *I* elements involved in I-R hybrid dysgenesis in *Drosophila melanogaster*. Nucl. Acids Res., 16: 9199-9123

Cully, D.F., Ip, H.S., and Cross, G.A.M., (1986) Coordinate transcription of variant

surface glycoprotein genes and an expression site associated gene family in *Trypanosoma brucei.* Cell **42**: 173-182

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

Daniels, D.L., Schroeder, J.L., Szybalski, W., Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F., Petersen, G.B., and Blattner, F.R., (1983) In 'Lambda II', Eds. Hendrix, R.W., Roberts, J.W., Stahl, F.W., and Weisberg, R.A., Appendix II, p519-661

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 analysis of long-established transformed lines. Genetics, **115**: 711-723

Das, G., Henning, D., Wright, D., and Reddy, R., (1988) Upstream regulatory elements are necessary and sufficient for transcription of a U6 RNA gene by RNA polymerase III. EMBO J., **7**: 503-512

David, J.R., and Merle, J., (1968) A reevaluation of the duration of egg chamber stages in oogenesis of *Drosophila melanogaster.* Dros. Inf. Ser., **431**: 122-123

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

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.* Nuc. Acids Res., **16**: 4041-4052

Di Nocera, P.P., and Casari, G., (1987) Related polypeptides are encoded by *Drosophila* F elements, I factors, and mammalian L1 sequences. Proc. Natl. Acad. Sci. USA, **84**: 5843-5847

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., and Dawid, I.B., (1983) Interdigitated arrangements of two oligo-A

terminated DNA sequences in *Drosophila*. Nuc. Acids Res., **11**: 5475-5482

Doolittle, R.F., Feng, D.-F., Johnstone, M.S., and McClure, M.A., (1989) Origins and evolutionary relationships of retroviruses. In 'The quarterly review of biology,' **64**: 1-30

Driever, W., and Nüsslein-Volhard, C., (1989) The *bicoid* protein is a positive regulator of *hunchback* in early *Drosophila* embryo. Nature, **337**: 138-143

Driver, A., Lacey, S.F., Cullingford, T.E., Mitchelson, A., and O'Hare, K., (1989) Structural analysis of *Doc* transposable elements associated with mutations at the *white* and *suppressor of forked* loci of *Drosophila melanogaster*. Mol. Gen. Genet., **220**: 49-52

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

Dynan, W.S., and Tjian, R., (1983a) Isolation of a transcription factor that discriminates between different promoters recognized by RNA polymerase II. Cell, **32**: 669-680

Dynan, W.S., and Tjian, R., (1983b) The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell **35**: 79-87

Economou-Pachnis, A., Lohse, M.A., Furano, A.V., and Tsichlis, P.N., (1985) Insertion of a long interspersed repeated element at the *Igh* (immunoglobulin heavy chain) and *Mlvi-2* (moloney leukemia virus integration-2) loci of rats, Proc. Natl. Acad. Sci. USA, **82**: 2857-2861

Eichinger, D.J., and Boeke, J.D., (1988) The DNA intermediate in yeast Ty1 element transposition copurifies with virus-like particles: cell-free Ty1 transposition. Cell, **54**: 955-966

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

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

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

Ephrussi, A., Dickinson, L.K., and Lehmann, R., (1991) *oskar* organises the germplasm and directs localization of the posterior determinant *nanos*. Cell, **66**: 37-50

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

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

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

Fasano, L., and Kerridge, S., (1988) Monitoring positional information during oogenesis in adult *Drosophila*. Development, **104**: 245-253

Fawcett, D.H., Lister, C.K., Kellet, E., and Finnegan, 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

Feinberg, A.P., and Vogelstein, B., (1984) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Addendum. Anal. Biochem., **137**: 266-267

Finnegan, D.J., (1988) *I* factor in *Drosophila melanogaster* and similar elements in other eukaryotes. In 'Transposition'. Eds. Kingsman, A.J., Kingsman, S.M., and Chater, K.F., Soc. Gen. Microbiol. Ltd., 271-285

Finnegan, D.J., (1989) The *I* factor and I-R hybrid dysgenesis in *Drosophila melanogaster*. In 'Mobile DNA', Eds. Berg, M.M., and Howe, M.M., **18**: 503-517

Finston, W.J., and Champou, J.J., (1984) RNA-primed initiation of Moloney murine leukemia virus plus strand by reverse transcription *in vitro*. J.Virol., **51**: 26-33

Fire, A., Samuels, M., and Sharp, P.A., (1984) Interactions between RNA polymerase II, factors, and template leading to accurate transcription. J. Biol. Chem., **259**: 2509-2516

Flanagan, P.M., Kelleher, R.J., Feaver, W.J., Lue, N.F., La Pointe, J.W., and

Kornberg, R.D., (1990) Resolution of factors required for the initiation of transcription by yeast RNA polymerase II. J. Biol. Chem., **265**: 11105-11107

Flavell, A.J., and Brierly, C., (1986) The termini of extrachromosomal linear *copia* elements. Nuc. Acids Res., **14**: 3659-3669

Flavell, A.J., and Ish-Horowicz, D., (1981) Extrachromosomal circular copies of the eukaryotic transposable element *copia* in cultured *Drosophila* cells. Nature (London), **292**: 591-594

Flores, O., Maldonado, E., Burto, Z., Greenblatt, J., and Reinberg, D., (1988) Factors involved in specific transcription by mammalian RNA polymerase II. RNA polymerase II associating protein 30 is an essential component of transcription factor IIF. J. Biol. Chem., **263**: 10812-10816

Flores, O., Lu, H., Killeen, M., Greenblatt, J., Burton, Z.F., and Reinberg., D., (1991) The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. Proc. Natl. Acad. Sci. USA, **88**: 9999-10003

Fouts, D., Ganguly, R., Gutierrez, A.G., Lucchesi, J.C., and Manning, J.E., (1988) Nucleotide sequence of the *Drosophila* glucose-6-phosphate dehydrogenase gene and comparison with the homologous human gene. Gene, **63**: 261-275

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

Fuscoe, J.C., Fenwick, R.G., Ledbetter, D.H., and Caskey, C.T., (1983) Detection and amplification of the HGPRT locus in Chinese hamster cells. Mol. Cell Biol. **3**: 1086-1096

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

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

Gibson, T.J., (1984) 'Studies on Epstien Barr virus genome.' PhD thesis, Cambridge

University, England.

Gill, G., and Tjian, R., (1991) A highly conserved domain of TFIID displays species specificity *in vivo*. Cell, **65**: 333-340

Gorman, C., Moffat, L., and Howard, B., (1982) Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol., **2**: 1044-1051

Gorman, C., (1985) in: DNA cloning - a practical approach, volume II. Ed. Glover, D.M. IRL Press, p143-165.

Grant, S.R., Gierl, A., and Saedler, H., (1990) En/Spm encoded tnpA protein requires a specific target sequence for suppression. EMBO J., **9**: 2029-2035

Greenblatt, J., (1991) Roles of TFIID in transcription initiation by RNA polymerase II. Cell, **66**: 1067-1070

Grosschedl, R., and Birnsteil, M.L., (1980) Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants *in vivo*. Proc. Natl. Acad. Sci. USA, **77**: 1432-1436

Grosveld, G.C., Shewmaker, C.K. Jat, P., and Flavell, R.A., (1981) Localization of DNA sequences necessary for transcription of the rabbit β-globin gene *in vitro*. Cell, **25**: 215-226

Grosveld, G.C., deBoer, E., Shewmaker, C.K., and Flavell, R.A., (1982) DNA sequences necessary for transcription of the rabbit β-globin gene *in vivo*. Nature (London), **295**: 120-126

Hahn, S., Hoar, E.T., and Guarente, L., (1985) Each of three 'TATA elements' specifies a subset of the transcription initiation sites at the *CYC-1* promoter of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA, **82**: 8562-8566

Hahn, S., Buratowski, S., Sharp, P.A., and Guarente, L., (1989a) Isolation of the gene encoding the yeast TATA binding protein TFIID: a gene identical to the *SPT15* suppressor of Ty element insertions. Cell, **58**: 1173-1181

Hahn, S., Buratowski, S., Sharp, P.A., and Guarente, L., (1989b) Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. Proc. Natl. Acad. Sci. USA, **86**: 5718-5722

Hasan, G., Turner, M.J., and Cordingley, J.S., (1984) Complete nucleotide sequence of an unusual mobile element from *Trypanosoma brucei.*, Cell. **37**: 333-341

Hay, B., Jan, Y.L., and Jan, Y.N., (1988) A protein component of *Drosophila* polar granules is encoded by *vasa* and has extensive sequence similarity to ATP-dependent helicases. Cell, **55**: 577-587

Heidmann, O., and Heidmann, T., (1991) Retrotransposition of a mouse IAP sequence tagged with an indicator gene. Cell, **64**: 159-170

Heiermann, R., and Pongs, O., (1985) *In vitro* transcription with extracts of nuclei of *Drosophila* embryos. Nucl. Acids Res., **13**: 2709-2730

Hemmati-Brivanlou, A., Frank, D.,Bolce, M.E., Brown, B.D., Sive, H.L., and Harland, R.M., (1990) Localization of specific mRNAs in *Xenopus* embryos by whole-mount *in situ* hybridization. Development, **110**: 325-330

Hen, R., Sassone-Corsi, P., Corden, J., Gaub, M.P., and Chambon, P., (1982) Sequences upstream from the T-A-T-A box are required *in vivo* and *in vitro* for efficient transcription from the adenovirus serotype 2 major late promoter. Proc. Natl. Acad. Sci. USA, **79**: 7132-7136

Hirsh, J., Morgan, B.A., and Scholnick, S.B., (1986) Delimiting regulatory sequences of the *Drosophila melanogaster Ddc* gene. Mol. Cell. Biol., **6**: 4548-4552

Hoey, T., Dynlach, B.D., Peterson, M.G., Pugh, B.F., and Tjian, R., (1990) Isolation and characterization of a *Drosophila* gene encoding the TATA box binding protein, TFIID. Cell, **61**: 1179-1186

Hoopes Jr., R.R., and Rohrmann, G.F., (1991) *In vitro* transcription of baculovirus immediate early genes: Accurate mRNA initiation by nuclear extracts from both insect and human cells. Proc. Natl. Acad. Sci. USA, **88**: 4513-4517

Horikoshi, M., Carey, M.F., Kakidani, H., Roeder, R.G., (1988) Mechanism of action of a yeast activator: direct effect of GAL4 derivatives on mammalian TFIID-promoter interactions. Cell **54**: 665-669

Horikoshi, M., Wang, C.K., Fujii, H., Cormlish, J.A., Weil, P.A., and Roeder, R.G., (1989) Purification of a yeast TATA box-binding protein that exhibits transcription factor IID activity. Proc. Natl. Acad. Sci., USA, **86**: 4843-4847

Horikoshi, M., Yamamoto, T., Ohkuma, Y., Wei, P.A., and Roeder, R.G., (1990) Analysis of structure-function relationship of yeast TATA box binding factor TFIID. Cell, **61**: 1171-1178

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

Hultmark, D., Klemenz, R., and Gehring, W.J., (1986) Translational and transcriptional control elements in the untranslated leader of the heat shock gene *hsp22*. Cell, **44**: 429-438

Ilyin Y.,V., Schuppe, N.G., Lyumbomirskya, N.V., Gorelova, T.V., and Arkhipova, I.R., (1984) Circular copies of mobile dispersed genetic elements in cultured *Drosophila melanogaster* cells. Nuc. Acids Res., **12**: 7517-7531

Ishii, S., Merlino, G.T., and Pastan, I., (1985) Promoter region of the human Harvey *ras* proto-oncogene: similarity to the EGF receptor proto-oncogene promoter. Science, **230**: 1378-1381

Jacks, T., Madhani, H.D., Masiarz, F.R., and Varmus, H.E., (1988) Signals for ribosomal frameshifting in the Rous sarcoma virus *gag-pol* region. Cell, **55**: 447-458

Jacks, T., and Varmus, H.E., (1985) Expression of Rous sarcoma virus *pol* gene by ribosomal frameshifting. Science, **230**: 1237-1242

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 retrotransposition in transgenic *Drosophila* demonstrates RNA-mediated transposition

of the LINE *I* elements. EMBO J., **10**: 1927-1937

Kadanoga, J.T., Jones, K.A., and Tjian, R., (1986) Promoter specific activation of RNA polymerase II transcription by Sp1. Trends Bioch. Sci., **11**: 20-23

Kamakaka, R.T., Tyree,C.M., and Kadanoga, J.T., (1991) Accurate and Efficient RNA polymerase II transcription with a soluable nuclear fraction derived from *Drosophila* embryos. Proc. Natl. Acad. Sci. USA, **88**: 1024-1028

Kao, C.C., Lieberman, P.M., Schmidt, M.C., Zhou, Q., Pei, R., and Berk, A.J., (1990) Cloning of the gene encoding the human TATA binding factor: expression and transcriptionally active TFIID protein. Science, **248**: 1646-1650

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

Kassis, J.A., Desplan, C., Wright, D.K., and O'Farrell, P.H., (1989) Evolutionary conservation of homeodomain-binding sites and other sequences upstream and within the major transcription unit of the *Drosophila* segmentation gene *engrailed*. Mol. Cell. Biol., **9**: 4304-4311

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

Kaufman, P.D., and Rio, D.C., (1991) *Drosophila* P-element transposase is a transcriptional repressor *in vitro*. Proc. Natl. Acad. Sci. USA, **88**: 2613-2617

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

Kempler, B., Jackson, P.D., and Felsenfeld, G., (1987) Protein binding sites within the 5' DNase-hypersensitive region of the chicken α^D-globin gene. Mol. Cell. Biol., **7**: 2059-2069

Kidwell, M.G., (1979) Hybrid dysgenesis in *Drosophila melanogaster*: the relationship between the P-M and I-R interaction systems. Genet. Res., **33**: 205-217

Kidwell, M.G., (1983) Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA, **80**: 1655-1659

Kidwell, M.G., Frydryk, T., and Novy, J.B., (1983) The hybrid dysgenesis potential of *Drosophila melanogaster* strains of diverse temporal and geographical natural origins. Dros. Inf. Ser., **59**: 63-69

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

Kikuchi,Y., Ando, Y., and Shiba, T., (1986) Unusual priming mechanism of RNA-directed DNA synthesis in *copia* retrovirus like particles in *Drosophila*. Nature (London). **323**: 824-826

Kikuchi, Y., Saski, N., and Ando-Yamagami, Y., (1990) Cleavage of tRNA within the mature tRNA sequence by the catalytic RNA of RNase P: implication for the formation of the primer tRNA fragment for reverse transcription in *copia* retrovirus-like particles. Proc. Natl. Acad. Sci. USA, **87**: 8105-8109

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

King, R.C., (1970) Ovarian development in *Drosophila melanogaster*. Academic Press: New York

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

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

Kurkulos, M., Weinberg, J.M., Pepling, M.E., and Mount, S.M., (1991) Polyadenylation in *copia* requires unusually distant upstream sequences, Proc. Natl.

Acad. Sci. USA, **88**: 3038-3042

Lakshmikumaran, M.S., D'Ambrosio, E., Laimans, L.A., Lin, D.T., and Furano, A.V., (1985) Long interspersed repeated DNA (LINE) causes polymorphism at the rat *insulin-1* locus. Mol.Cell Biol., **5**: 2197-2203

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

Lasko, P.F., and Ashburner, M., (1988) The product of the *Drosophila* gene *vasa* is very similar to eukaryotic initiation factor-4A. Nature, **335**: 611-617

Lathe, R., Kieny, M.P., Skory, S., and Lecocq, J.P., (1984), DNA **3**: 173-182

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

Lavige, J.M., and Lecher, P., (1982) Mitoses anomales dans les embryons à développement bloqué dans le système I-R de dysgénésie hybride chez *Drosophila melanogaster*., Biology of the Cell, **44**: 9-14

Leeton, P., (1991) PhD thesis entitled 'A highly repeated DNA sequence dispersed throught he genome of *Lilium speciosum*', Monash Uni., Melbourne, Australia.

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

Liao, X.-B., Clare, J.J., and Farabaugh, P.J., (1986) The UAS site of a Ty2 element of yeast is necessary but not sufficient to promote maximal transcription of the element. Proc. Natl. Acad. Sci. USA, **84**: 8520-8524

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

Long, E.O., and Dawid, I.B., (1979) Expression of ribosomal DNA insertions in

Drosophila melanogaster. Cell, **18**: 1185-1196

Lozovskaya, E.R., Scheinker, V.Sh., and Evgen'ev, M.B., (1990) A hybrid dysgenesis syndrome in *Drosophila virilis*. Genetics, **126**: 619-623

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

Lynch, M., (1989) Ph.D. thesis - A molecular study of the I factor of *Drosophila melanogaster*.

Mahowald, A.P., and Hardy, P.A., (1985) Genetics of *Drosophila* embryogenesis. Ann. Rev. Genet., **19**: 149-177

Mahowald, A.P., and Kambysellis, M.P., (1980) Oogenesis. In 'The genetics and biology of *Drosophila* (ed. M. Ashburner and T.R.F. Wright), **2d**:141-224 Acad. Press, London and New York.

Mahowald, A.P., and Tiefert, M., (1970) Wilhelm Roux's Arch. Devl. Biol., **165**:8-25

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

Maniatis, T., Fritsch, E.F., and Sambrook, J., (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbour Laboratories, Cold Spring Harbour, New York.

McKnight, S.L., and Tjian, R. (1986) Transcriptional selectivity of viral genes in mammalian cells. Cell, **46**: 795-805

Mellor, J., Fulton, A.M., Dobson, M.J., Wilson, W., Kingsman, S.M., and Kingsman, A.J., (1985a) A retrovirus-like strategy for the expression of fusion protein encoded by yeast transposon Ty1. Nature (London), **313**: 243-246

Mellor, J., Malim, M.H., Gull, K., Tuite, M.F., McCready, S., Dibbayawan, T., Kingsman, S.M., and Kingsman, A.J., (1985b) Reverse transcriptase activity and Ty RNA are associated with virus-like particles in yeast. Nature (London), **318**: 583-586

Melton, D.W., (1987) HPRT gene organization and expression. In Oxford surveys on eukaryotic genes, ed. MacLean, N., **4**: 34-76

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

Messing, J., (1983) New M13 vectors for M13mp19 cloning. Methods Enzymeol. **101**: 20-78

Minchiotti, G., and Di Nocera, P.P., (1991) Convergent transcription initiates from oppositely orientated 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 66kd protein is a repressor of transposase activity. Cell, **62**: 269-284

Miyake, T., Mae, N., Shiba, T., and Kondo, S., (1987) Production of virus-like particles by the transposable genetic element, *copia*, of *Drosophila melanogaster*. Mol. Gen. Genet., **207**: 29-37

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

Mizrokhi, L.J., and Mazo, A.M., (1990) Evidence for horizontal transmission of the mobile element *jockey* between distant *Drosophila* species. Proc. Natl. Acad. Sci. USA, **87**: 9216-9220

Morse, B., Rothberg, 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 human breast carcinoma. Nature (London), **333**: 87-90

Mossie, K.B., Young, M.W., and Varmus, H.E., (1985) Extrachromosomal DNA forms of *copia*-like transposable elements, F elements and middle repetitive DNA sequences in *Drosophila melanogaster*.. J. Mol. Biol., **182**: 31-42

Mullins, M.C., (1990) In Rubin Lab methods book, 2nd edn., p187

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

Murphy, N.B., Guyaux, M., Pays, E., and Steinert, M., (1986) In proceedings of UCLA symposium on molecular strategies of parasitic invasion. Agabian, N., Goodman, H., Noguiera, N., Eds. A.R. Liss Inc. N.Y.

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

Nagawa, F., and Fink, G.R., (1985) The relationship between the 'TATA' sequence and transcription initiation sites at the *HIS4* gene of *Saccharomyces cerevesiae*. Proc. Natl. Acad. Sci. USA, **82**: 8557-8561

Neuhaus, H., Muller, F., Etter, A., and Tobler, H., (1987) Type 1-like intervening sequences are found in the rDNA of the nematode *Ascaris lumbricoides*. Nuc. Acid Res. **15**: 7689-7705

Norrander, J., Kempe, T., and Messing, J., (1985) Construction of improved M13 vectors using oligonucleotide directed mutagenesis. Gene **26**: 101

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

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

O'Hare, K., Murphy, C., Levis, R., and Rubin, G.M., (1984) DNA sequence of the *white* locus of *Drosophila melanogaster*. J. Mol. Biol., **180**: 437-445

Ohkuma, Y., Horikoshi, M., Roeder, R.G., and Desplan, C., (1990) Engrailed, a homeodomain protein, can repress *in vitro* transcription by competition with the TATA box-binding protein transcription factor IID. Proc. Natl. Acad. Sci. USA, **87**: 2289-2293

Padgett, R.W., Hutchison III, C.A., and Edgell, M.H., (1988) The F type 5' motif of mouse L1 element: a major class of L1 termini similar to the A-type in organization but unrelated in sequence. Nuc. Acid 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

Paskewitz, S.M., and Collons, F.H., (1989) Site specific ribosomal DNA insertion elements in *Anopheles gambiae* and *A. arabiensis*: nucleotide sequence of gene-element boundaries. Nuc. Acid Res., **17**: 8125-8133

Pélisson, A., (1978) Non-mendelian female sterility in *Drosophila melanogaster*: variations of chromosomal contamination when caused by chromosomes of various inducer efficiencies. Genet. Res., Camb., **32**: 113-122

Pélisson, A., (1979) The I-R system of hybrid dysgenesis in *Drosophila melanogaster*: influence of SF sterility of their inducer and reactive paternal chromosomes. Heredity, **43**: 423-428

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., 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

Pélisson, A., and Bregliano, J.C., (1981) The I-R system of hybrid dysgenesis in *Drosophila melanogaster*: construction and characterisation 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

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. Gen. Dev., **2**: 1615-1626.

Peterson, M.G., Tanese, N., Pugh, B.F., and Tjian, R., (1990) Functional domains and upstream activation properties of a cloned human TATA binding protein. Science, **248**: 1625-1630

Picard, G., (1976) Non mendelian female sterility in *Drosophila melanogaster*:

heredity and transmission of *I* factor. Genetics, **83**: 107-123

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

Picard, G., (1978b) Non mendelian female sterility in *Drosophila melanogaster*, further data on chromosome contamination. Mol. Gen. Genet., **164**: 235-247

Picard, G., (1978c) Non mendelian female sterility in *Drosophila melanogaster*: sterility in stocks derived from the genotypically inducer or reactive offspring of SF and RSF females. Biol. Cell., **31**: 245-254

Picard, G., (1979) *Drosophila melanogaster*: principal characteristics of chromosomes from inducer and reactive origin after chromosome contamination. Genetics, **91**: 455-471

Picard, G., Bucheton, A Lavige, J.M., and Fleuriet, A., (1972) Contribution à l'étude d'un phénomène de stérilité à déterminisme non mendélien chez *Drosophila melanogaster*. Comptes Rendus Ac. Sc. Paris D. **275**: 933-936

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

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

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

Pirrotta, V., Manet, E., Hardon, E., Bickel, S.E., and Benson, M., (1987) Structure and sequence of the *Drosophila zeste* gene. EMBO J., **6**: 791-799

Preston, C.R., and Engels, W.R., (1989) Spread of *P* transposable elements in inbredlines of *Drosophila melanogaster*. In Prog. Nucl. Acid Res. Mol. Biol., eds. S.W. Cohn, S.W., and Moldave, K., Acad. Press.

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., Pelisson, A., Bucheton, A., and Finnegan. 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

Pugh, B.F., and Tjian, R., (1990) Mechanism of transcription activation by Sp1: Evidence for coactivators. Cell, **61**: 1187-1197

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

Puro, J., and Nokkala, S., (1977) Meiotic segregation of chromosomes in *Drosophila melanogaster* oocytes. A cytological approach. Chromosoma (Berl.), **63**: 273-286

Rae, P.M.M., Kohorn, B.D., and Wade, R.P., (1980) The 10kb *Drosophila virilus* 28S rDNA intervening sequence is flanked by a direct repeat of 14 base pairs of coding sequence. Nuc. Acids Res., **8**: 3491-3504

Read, S.M., and Northcote, D.H., (1981) Minimisation of variation in the response to different proteins of the Coomassie blue G dye binding assay of protein. Anal. Biochem. **116**: 53-64

Reed, K.C., and Mann, D.A., (1985) Rapid transfer of DNA from agarose gels to nylon membranes. Nuc. Acid Res., **13**: 7207-7221

Regulski, M., McGinnis, N., Chadwick, R., and McGinnis, W., (1987) Developmental and molecular analysis of *Deformed*; a homeotic gene controlling *Drosophila* head development. EMBO J., **6**: 767-777

Resnick, R., Omer, C.A., and Faras, A.J., (1984) Involvement of retroviral reverse-transcriptase-associated RNase H in the initiation of strong stop(+) DNA synthesis and the generation of the long terminal repeat. J. Virol., **51**: 813-821

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 *Drosophila*

protein that binds to the terminal 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

Roiha, H., Muller, J.R., Woods, L.C., and Glover, D.M., (1981) Arrangements and rearrangements of sequences flanking the two types of rDNA insertion in *Drosophila melanogaster*. Nature, **290**: 749-753

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

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

Saigo, K., (1986) A potential primer for reverse transcription of *mdg-3*, a *copia*-like element is a leucine tRNA lacking its 3' terminal five bases. Nuc. Acids Res., **14**: 4370-4371

Saltzman, A.G., and Weinmann, R., (1989) Promoter specificity and modulation of RNA polymerase II transcription. FASEB J., **3**: 1723-1733

Sanfacon, H., and Hohn, T., (1990) Proximity to the promoter inhibits recognition of cauliflower mosaic virus polyadenylation signal. Nature (London), **346**: 81-84

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

Santoro, C., Mermod, N., Andrews, P.C., and Tjian, R., (1988) Cloning of cDNAs encoding human CCAAT-box binding proteins: a family of transcription and replication factors. Nature (London), **334**: 218-224

Scherf, S.J., Horn, G.T., and Erlich, H.A., (1986) Direct cloning and sequence analysis of enzymatically amplified genomic sequences. Science, **233**: 1076-1078

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

Schwarz-Sommer, Z., Leclercq, L., Ciobel, E., and Saedler, H., (1987) *Cin4*, an

insert altering the structure of the *A1* gene in *Zea mays*, exhibits properties of non-viral retrotransposons. EMBO J., **6**: 3873-3880

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

Shepard, B.M., and Finnegan, D,J., (1984) Structure of circular copies of *412* transposable elements present in *Drosophila melanogaster* tissue culture cells and isolation of free *412* long terminal repeat. J. Mol. Biol., **180**: 21-40

Shiba, T., and Saigo, , K., (1983) Retrovirus-like particles containing RNA homologous to the transposable element *copia* in *Drosophila melanogaster*. Nature (London), **302**:119-124

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, C., Pélisson, A., and Bucheton, A., (1988) Transposable and nontransposable elements similar to the *I* factor involved in inducer-reactive (IR) hybrid dysgenesis in *Drosophila melanogaster* coexist in various *Drosophila* species. Proc. Natl. Acad. Sci. USA, **85**: 1141-1145

Singer, V.L., Wobbe, C.R., and Struhl, K., (1990) A wide variety of DNA sequences can functionally replace yeast TATA element for transcriptional activation. Genes Dev., **4**: 636-645

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 cytoplacmic 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

Smith, J.K., Cywinski, A., and Taylor, J.M., (1984) Initiation of plus-strand DNA synthesis during reverse transcription of an avian retrovirus genome. J.Virol., **51**: 26-33

Soeller, W.C., Poole, S.J., and Kornberg, T., (1988) *In vitro* transcription of the *Drosophila engrailed* gene. Genes Dev., **2**: 68-81

Sopta, M., Carthew, R.W., and Greenblatt, J., (1985) Isolation of three proteins that bind to mammalian RNA polymerase II. J. Biol. Chem., **260**: 10353-10360

Southern, E.M., (1975) Detection of specific sequences among DNA fragments separated by agarose gel electrophoresis. J. Mol. Biol. **98**: 503-517

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

Sprenger, F., Stevens, L.M., and Nüsslein-Volhard, C., (1989) The *Drosophila* gene *torso* encodes a putative receptor tyrosine kinase. Nature, **338**: 478-483

St. Johnston, D., Beuchle, D., and Nüsslein-Volhard, C., (1991) *staufen*, a gene required to localize maternal RNAs in the *Drosophila* egg. Cell, **66**: 51-63

Steller, H., and Pirrotta, V., (1985) A transposable *P* vector that confers selectable G418 resistance to *Drosophila* larvae. EMBO J., **4**: 167-171

Sun, X-H., Lis, J.T., and Wu, R., (1988) The positive and negative transcriptional regulation of the *Drosophila Gapdh-2* gene. Gen. & Dev., **2**: 743-753

Sussman, R., and Jacob, F., (1962) Sur un système de répression thermosensible chez le bacteriophage λ *d'Escherichia coli.* CR Acad Sci (Paris) **254**: 1517-1519

Swadago, M., and Sentenac, A., (1990) RNA polymerase B 9II) and general transcription factors. Ann. Rev. Biochem., **59**: 711-754

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

Tautz, D., Lehmann, R., Schnürch, H., Schuh, R., Seifert, E., Kienlin, A., Jones, K., and Jäckle, H., (1987) Finger protein of novel structure encoded by *hunchback*, a second member of the gap class of *Drosophila* segmentation genes. Nature, **327**:

383-389

Tautz, D., and Pfeifle, C., (1989) A non-radioactive *in situ* hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene *hunchback*. Chromosoma (Berl.), **98**: 81-85

Thummel, C.S., (1989) The *Drosophila E74* promoter contains essential sequences downstrean from the start site of transcription. Genes Dev., **3**: 782-792

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

Tso, J.Y., Sun, X.H., and Wu, R., (1985) Structure of two unlinked *Drosophila melanogaster* glyceraldehyde-3-phosphate dehydrogenase genes. J. Biol. Chem., **260**: 8220-8228

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

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

Voliva, C.F., Jahn, C.L., Comer, M.B., Hutchison III, C.A., and Edgell, M., (1983) The L1Md long interspersed repeat family in the mouse: almost all examples are truncated at one end. Nuc. Acid Res. **11**: 8847-8859

Wasylyk, B., Derbyshire, R., Guy, A., Molko, D., Roget, A., Teoule, R., and Chambon, P., (1980) Specific *in vitro* transcription of conalbumin gene is drastically decreased by a single-point mutation in TATA box homology sequence. Proc. Natl. Acad. Sci. USA, **77**: 7024-7028

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

Wharton, R.P., and Struhl, G., (1989) Structure of the *Drosophila Bicaudal D* protein and its role in localizing the posterior determinant *nanos*. Cell, **59**: 881-892

Wincker, P., Jubier-Maurin, V., and Roizès, G., (1987) Unrelated sequences at the

5' end of mouse LINE-1 repeated elements define two distinct subfamilies. Nuc. Acids Res., **15**: 8593-8605

Wingender, E., (1988) Compilation of transcription regulating proteins. Nucl. Acids Res., **16**: 1879-1902

Xiao, H. and Lis, J.T., (1988) Germline transformation used to define key features of heat-shock response elements. Science, **239**: 1139-1142

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. Nuc. Acid Res., **16**: 10561-10573

Xiong, Y., and Eickbush, T.H., (1988a) Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. Mol. Biol. Evol., **5**: 675-690

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

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

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* bearing the male recombination factor 23.5 MRF. Cell, **49**: 487-495

Yoshioka, K., Honma, H., Zushi, M., Kondo, S., Togashi, S., Miyake, T, and Shiba, T., (1990) Virus-like particle formation of *Drosophila copia* through autocatalytic processing. EMBO J., **9**: 535-541

Yuki, S., Inouye, S., Ishimaru, S., and Saigo, K., (1986) Nucleotide sequence characterisation of a *Drosophila* retrotransposon, *412*. Eur. J. Biochem., **158**: 403-410

Zoller, M.J., and Smith, M., (1984) Oligonucleotide-directed mutagenesis: A simple method using two oligonucleotide primers and a single stranded DNA template. DNA, **3**: 479-488