Transcriptional Regulation of the
$I$ factor, a $Drosophila$ melanogaster
transposable element

Ivan Benjamin Nile Clark

A thesis presented for the degree of PhD
University of Edinburgh
1999
I dedicate this thesis to my parents.
Declaration

This thesis has been composed by myself and the work presented herein is my own, except where stated.

Ivan Clark
August 1999
Acknowledgements

There are many people to whom I am indebted for their help and encouragement during the time I have been working on this thesis. First I would like to thank my supervisor David Finnegan for all his advice, support and ideas. I am also grateful to my second supervisor Bill Earnshaw for advice and many useful suggestions. I would also like to thank all of the past and present members of the Finnegan lab who have provided help, encouragement and friendship. They are Angela, Jill, Mary, Eve, Maki, Laura, Apinunt, Hong mei, Ruth, Petra, Lei and David. I would also like to thank Dannie, an excellent summer student, who carried out most of the mapping of transgenes to chromosomes. I am indebted to many other members of ICMB too numerous to mention for materials, technical advice, media and solutions.

I would like to thank Christophe Terzian who suggested the trinucleotide repeat experiment and professors Maniatis and Tolias who generously provided materials.

Some former colleagues at Newcastle university also deserve a mention, particularly Kathy, Adam and Paul Shore who supported me in my decision to come to Edinburgh during what was a difficult time. I would also like to mention certain of my friends, especially Azizah, Janette and Carola who have shown in very diverse ways that it is possible to remain human while working on a PhD.

Finally and most importantly I would like to thank my parents, without whose loving and generous support this thesis would have been impossible.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>adenosine</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine 5'-triphosphate</td>
</tr>
<tr>
<td>AAE</td>
<td>Adh adult enhancer</td>
</tr>
<tr>
<td>Adh</td>
<td>alcohol dehydrogenase</td>
</tr>
<tr>
<td>AEF-1</td>
<td>adult enhancer factor 1</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>B</td>
<td>C, G or T</td>
</tr>
<tr>
<td>bp</td>
<td>base pair or base pairs</td>
</tr>
<tr>
<td>c</td>
<td>centi (x10^-2)</td>
</tr>
<tr>
<td>C</td>
<td>cytosine</td>
</tr>
<tr>
<td>°C</td>
<td>degrees celsius</td>
</tr>
<tr>
<td>CAT</td>
<td>chloramphenicol acetyl transferase</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>C/EBP</td>
<td>CAAT/enhancer binding protein</td>
</tr>
<tr>
<td>Ci</td>
<td>curie</td>
</tr>
<tr>
<td>D</td>
<td>A, G or T</td>
</tr>
<tr>
<td>dATP</td>
<td>deoxyadenosine triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>deoxycytosine triphosphate</td>
</tr>
<tr>
<td>dGTP</td>
<td>deoxyguanosine triphosphate</td>
</tr>
<tr>
<td>dH2O</td>
<td>de-ionised water</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNase</td>
<td>deoxyribonuclease</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxynucleotide triphosphate</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>dTTP</td>
<td>deoxythymosine triphosphate</td>
</tr>
<tr>
<td>E. coli</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetra-acetic acid</td>
</tr>
<tr>
<td>ELISA</td>
<td>enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>F</td>
<td>faradays</td>
</tr>
<tr>
<td>g</td>
<td>grams</td>
</tr>
<tr>
<td>G</td>
<td>Guanosine or acceleration due to gravity (approximately 9.8m per second per second)</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>H</td>
<td>A, C or T</td>
</tr>
<tr>
<td>HEPES</td>
<td>N-[2-Hydroxyethyl]peperazine-N’-[2-ethane-sulphonic acid]</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>k</td>
<td>kilo (x10^3)</td>
</tr>
<tr>
<td>l</td>
<td>Litres</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Broth</td>
</tr>
<tr>
<td>LINE</td>
<td>long interspersed nucleotide element</td>
</tr>
<tr>
<td>LTR</td>
<td>long terminal repeat</td>
</tr>
<tr>
<td>λ</td>
<td>phage lambda</td>
</tr>
<tr>
<td>m</td>
<td>milli (x10^{-3}), or metres</td>
</tr>
<tr>
<td>M</td>
<td>molar (moles per litre)</td>
</tr>
<tr>
<td>min</td>
<td>minutes</td>
</tr>
<tr>
<td>μ</td>
<td>micro (x10^{-6})</td>
</tr>
<tr>
<td>n</td>
<td>nano (x10^{-9})</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>Ω</td>
<td>ohms</td>
</tr>
<tr>
<td>p</td>
<td>pico (x10^{-12})</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>pfu</td>
<td>plaque forming units</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethylsulphonyl fluoride</td>
</tr>
<tr>
<td>PO₄</td>
<td>Phosphate</td>
</tr>
<tr>
<td>Pu</td>
<td>purine</td>
</tr>
<tr>
<td>Py</td>
<td>pyrimidine</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SL₂</td>
<td>Schneider line 2 culture cells</td>
</tr>
<tr>
<td>SV40</td>
<td>simian virus 40</td>
</tr>
<tr>
<td>T</td>
<td>thymidine</td>
</tr>
<tr>
<td>TEMED</td>
<td>N, N, N’-tetramethylethylenediamine</td>
</tr>
<tr>
<td>Tris</td>
<td>tris (hydroxymethyl) aminomethane</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
</tbody>
</table>
UV    ultraviolet
V     A, C or G or volts
v/v   volume per volume
w/v   weight per volume
X-Gal 5-bromo-4-chloro-3-indol-β-D-galactopyranoside

Definitions are not given for the standard symbols for chemical elements and amino acids.
Transcriptional regulation of the \textit{I} factor, a \textit{Drosophila melanogaster} transposable element

Abstract

The \textit{I} factor is a transposable element of the LINE family found in \textit{Drosophila melanogaster}. High frequency \textit{I} factor transposition is associated with I-R hybrid dysgenesis, a syndrome of female sterility occurring in the progeny of males from an inducer strain, which carry active \textit{I} factors and females from a reactive strain which do not. Expression of the \textit{I} factor is restricted to the female germ line and is repressed in inducer strains. Transcription is regulated by sequences located internally in the 5' untranslated region of the element. A sequence-specific DNA binding protein present in ovaries recognises a 19bp site within this region, known as site 1, which a previous study suggested is important for germ line-specific transcription. The protein that binds to site 1 was identified as Adult Enhancer Factor 1, a transcriptional repressor that regulates the alcohol dehydrogenase and yolk protein genes. A series of mutations were made in site 1 and the effects on AEF-1 binding \textit{in vitro}, and the expression of a reporter gene controlled by \textit{I} factor regulatory sequences \textit{in vivo}, were investigated. It was discovered that deletion of site 1 and other mutations that reduce AEF-1 binding did not reduce expression, contrary to previous findings. The transgenes used in the previous study were characterized and found not to be as they were described. Constructs designed to express sense and antisense AEF-1 RNA in the female germ line were introduced into flies and the effect on ovarian expression from a reporter gene under the control of \textit{I} factor regulatory sequences was investigated. In each case expression was slightly lower than in a control experiment. The implications of these findings for models of \textit{I} factor regulation are discussed.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>i</td>
</tr>
<tr>
<td>Dedication</td>
<td>ii</td>
</tr>
<tr>
<td>Declaration</td>
<td>iii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iv</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>v</td>
</tr>
<tr>
<td>Abstract</td>
<td>viii</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>ix</td>
</tr>
</tbody>
</table>

## CHAPTER 1- INTRODUCTION

1.1 TRANSPOSABLE ELEMENTS

1.1.1 Classification of transposable elements

1.1.1.1 Retrovirus-like elements

1.1.1.2 LINE-like elements

1.1.1.4 Class II elements

1.1.2 Hybrid dysgenesis

1.1.2.1 P-M hybrid dysgenesis

1.1.2.2 I-R hybrid dysgenesis

1.1.2.3 Other hybrid dysgenesis systems

1.2 THE I FACTOR

1.2.1 Structure

1.2.2 RNA

1.2.3 Proteins

1.2.4 Defective pericentromeric I elements and the origins of the Drosophila melanogaster I factor

1.3 TRANSPOSABLE ELEMENT REGULATION

1.3.1 Regulation by the host

1.3.1.1 RIP and MIP in fungi

1.3.1.2 Methylation and the epigenetic regulation of plant transposable elements

1.3.1.3 Repeat associated gene silencing in plants

1.3.1.4 Silencing, methylation and transposable elements in animals

1.3.2 Autoregulation

1.3.3 Regulation of P element transposition

1.3.3.1 Introduction - tissue-specificity and the P cytotype

1.3.3.2 Tissue-specificity in P transposition is regulated at the level of pre-mRNA splicing

1.3.3.3 The role of internally-deleted P elements

1.3.3.4 Mechanisms of cytotype repression
1.3.3.4.1 The multimer poisoning model ................................................................. 19
1.3.3.4.2 Direct interference in transposition by repressor DNA binding .............. 20
1.3.3.4.3 Antisense RNA ...................................................................................... 20
1.3.3.4.4 Transcription and the P cytotype - trans-silencing .................................. 20
1.3.3.5 Maternal inheritance of the P cytotype ..................................................... 21

1.3.4 Regulation of I factor transposition ................................................................. 22
1.3.4.1 Levels of regulation ................................................................................... 22
1.3.4.1.1 Tissue-specificity ................................................................................... 22
1.3.4.1.2 Copy number control .......................................................................... 23
1.3.4.1.3 Reactivity ............................................................................................. 24
1.3.4.1.4 Influence of age, temperature and DNA damage on reactivity .............. 24
1.3.4.2 I factor transcription .................................................................................. 25
1.3.4.2.1 The I factor promoter .......................................................................... 26
1.3.4.2.2 The enhancer ....................................................................................... 26
1.3.4.3 Models of I factor regulation ................................................................. 27
1.3.4.3.1 I-encoded repressor model ............................................................... 27
1.3.4.3.2 Titration and homology-dependent gene silencing models .............. 28

1.3.5 Scope of the thesis ....................................................................................... 31

CHAPTER 2- MATERIALS AND METHODS .................................................................32

2.1 MATERIALS .................................................................................................. 33

2.1.1 Media ......................................................................................................... 33
2.1.1.1 Bacterial media ....................................................................................... 33
Luria Broth (LB) ........................................................................................................ 33
Luria Agar (L-agar) .................................................................................................. 33
SOC Buffer ............................................................................................................. 33
Top Agarose ........................................................................................................... 33
2 x TY Broth .......................................................................................................... 33

2.1.1.2 Drosophila media .................................................................................... 33
‘French’ Fly Food ................................................................................................... 33
Apple juice agar .................................................................................................... 34
Tissue culture medium .......................................................................................... 34

2.1.2 Materials .................................................................................................... 34
2.1.2.1 Chemicals ............................................................................................... 34

2.1.2.2 Solutions ................................................................................................. 34
TE .......................................................................................................................... 34
Nuclease mix .......................................................................................................... 34
Phage buffer .......................................................................................................... 34
PEG solution .......................................................................................................... 34
Kinase/ligase buffer ............................................................................................. 34
4 x Agarose gel loading buffer ............................................................................ 34
TAE ......................................................................................................................... 35
TBE ......................................................................................................................... 35
2.2 METHODS ..................................................................................................................... 41

2.2.1 Manipulation of bacteria ......................................................................................... 41
  2.2.1.1 Growth of E. coli cultures .................................................................................. 41
  2.2.1.2 Storage of E. coli cultures .................................................................................. 41
  2.2.1.3 Transformation of bacteria ................................................................................ 41

2.2.2 Manipulation of lambda phage .............................................................................. 42
  2.2.2.1 Phage growth ..................................................................................................... 42
  2.2.2.2 Preparation of phage lysates ............................................................................ 42
  2.2.2.3 Phage lambda DNA purification ...................................................................... 42
  2.2.2.4 Lysogeny of phage lambda .............................................................................. 43

2.2.3 In vitro manipulation of DNA ............................................................................... 43
  2.2.3.1 Small scale preparation of plasmid DNA .......................................................... 43
  2.2.3.2 Large scale preparation of plasmid DNA .......................................................... 43
  2.2.3.3 Removal of protein from DNA by Phenol/chloroform extraction ....................... 44
  2.2.3.4 Precipitation of DNA using ethanol .................................................................. 44
  2.2.3.5 Quantification of DNA ...................................................................................... 44
  2.2.3.6 Cleavage of DNA by restriction endonucleases ................................................. 45
  2.2.3.7 Preparation of concatenated double-stranded oligonucleotides ....................... 45
  2.2.3.8 Agarose gel electrophoresis ............................................................................. 45
  2.2.3.9 Purification of DNA fragments from agarose ................................................. 46
2.2.3.10 Creation of blunt termini from 3’ recessed ends ................................................. 46
2.2.3.11 Ligation of DNA fragments .................................................................................. 46
2.2.3.12 Colony and plaque lifts ........................................................................................ 46
2.2.3.13 Southern blotting .................................................................................................. 47
2.2.3.14 UV cross-linking .................................................................................................. 47
2.2.3.15 Random primed labelling .................................................................................... 47
2.2.3.16 Nick translation ..................................................................................................... 47
2.2.3.17 End-labelling of DNA fragments using Klenow polymerase .................................. 48
2.2.3.18 Removal of unincorporated nucleotide following labelling ................................. 48
2.2.3.19 Hybridisation ....................................................................................................... 48
2.2.3.20 Autoradiography .................................................................................................. 48
2.2.3.21 Sequencing of double-stranded plasmid DNA ..................................................... 48
2.2.3.22 Polymerase chain reaction ................................................................................... 49
2.2.3.23 Cloning of PCR products .................................................................................... 50
2.2.5 Manipulation of Drosophila melanogaster flies, cells and tissues ......................... 50
2.2.5.1 Cell culture ........................................................................................................... 50
2.2.5.2 Maintenance of Drosophila stocks ....................................................................... 50
2.2.5.3 Preparation of Drosophila genomic DNA ............................................................. 50
2.2.5.4 P element-mediated germine transformation (Rubin and Spradling, 1982) .......... 51
2.2.5.5 Chromosome mapping of P element insertions in Drosophila lines ...................... 51
2.2.5.5 Ovary dissection .................................................................................................. 52
2.2.6 Protein extraction, purification and analysis ............................................................. 53
2.2.6.1 Preparation of protein extracts from E. coli lysogens ........................................... 53
2.2.6.2 Preparation of CAT extracts from Drosophila tissues ........................................ 53
2.2.6.3 Preparation of nuclear extracts ............................................................................ 53
2.2.6.4 Estimation of protein concentration by Bradford assay ....................................... 54
2.2.6.5 Measurement of CAT concentration in tissue extracts .......................................... 54
2.2.7 Protein-DNA interactions ........................................................................................ 55
2.2.7.1 cDNA library screen ............................................................................................. 55
2.2.7.2 Gel retardation assays .......................................................................................... 56

RESULTS

CHAPTER 3 – A SCREEN FOR SITE 1 BINDING PROTEINS ........................................... 57

3.1 INTRODUCTION ............................................................................................................ 58
3.2 RESULTS ....................................................................................................................... 58
3.2.1 Isolation and DNA binding specificity of cDNA clones encoding site 1 binding proteins ......................................................................................................................... 58
3.2.2 Gtoll and Gtol2 contain partial cDNAs of known Drosophila melanogaster genes .... 62
3.2.3 The Gtoll insert is a partial Adult Enhancer Factor-1 cDNA ...................................... 63
3.2.6 The Gtol2 insert is a partial cDNA of IsRFC/Gnf1 .................................................... 64
3.3 DISCUSSION ................................................................................................................ 66
6.2.2 Measurement of CAT expression ................................................................. 115
6.2.3 Statistical analysis ................................................................................... 116
6.3 DISCUSSION ................................................................................................ 118

CHAPTER 7 - REACTIVITY AND THE OPA MOTIF OF THE AEF-1 GENE .......... 121

7.1 INTRODUCTION .......................................................................................... 122
7.2 RESULTS ...................................................................................................... 123
7.3 DISCUSSION ............................................................................................... 125

CHAPTER 8 – GENERAL DISCUSSION .............................................................. 126

8.1 CONCLUSIONS ........................................................................................... 127
8.2 IMPLICATIONS FOR THEORIES OF FACTOR REGULATION ...................... 127
8.3 FUTURE WORK ............................................................................................ 130

APPENDIX - CHARACTERISATION OF TRANSGENIC DROSOPHILA LINES .. 131

REFERENCES .................................................................................................... 145
Chapter 1- Introduction
Introduction

The I factor is a Drosophila melanogaster transposable element which causes female sterility, as part of a syndrome known as I-R hybrid dysgenesis. This phenomenon represents a good model system for the study of transposable element regulation. Following a brief survey of eukaryotic transposable elements and a description of the features of the I factor, this chapter will discuss the mechanisms by which transposition is regulated with particular reference to hybrid dysgenesis.

1.1 Transposable elements

Transposable elements are segments of the genome that encode functions capable of copying or moving their own DNA to new chromosomal locations. They are generally regarded as 'selfish', conferring no direct selective advantage to the organism although this may not be the case for all transposable elements (Biessmann et al., 1992a, b). Transposition presents a threat to genome stability as it can result in mutagenesis by insertion and by stimulating ectopic recombination, either directly or through increasing element copy number providing regions of homology that can be recognised by the cellular recombination machinery (Engels and Preston, 1984, Busseau et al., 1989, Lim, 1989, Montgomery et al., 1991). It is accepted that selection against such events acts to eliminate elements from populations, although there is controversy as to the relative importance of direct insertion and ectopic recombination in this process (see, for example, Charlesworth et al., 1997, Biémont et al., 1997). This is opposed by transposition, which increases the copy number, maintaining elements in a population despite these selective forces.

1.1.1 Classification of transposable elements

Eukaryotic transposable elements are classified into two general groups on the basis of sequence and their method of transposition (Finnegan, 1989). Class I elements transpose by replicative mechanisms involving reverse transcription of an RNA intermediate while class II elements utilise conservative DNA-mediated mechanisms. Class I elements are further divided into two groups, the retrovirus-like and the LINE-like elements.
1.1.1.1 Retrovirus-like elements

The structure of retrovirus-like elements resembles that of the integrated proviral form of retroviruses (see Figure 1A). They have long terminal direct repeats (LTRs) at either end and open reading frames encoding polypeptides homologous to the retroviral \textit{gag}, \textit{pol} and in some elements \textit{env} proteins. The mechanism of transposition involves reverse transcription (Boeke \textit{et al.}, 1985) and is believed to be similar to the retrovirus life cycle. Transcripts are copied by reverse transcriptase eventually producing a double-stranded DNA intermediate which is integrated into the genome either by an integrase activity of the \textit{pol}-like gene product (Eichinger and Boeke, 1988) or by homologous recombination with a genomic copy of the same element (Melamed \textit{et al.}, 1992, Sharon \textit{et al.}, 1994, Ke and Voytas, 1997). Virus-like particles (VLPs) containing element-encoded proteins, RNA and reverse transcriptase activity are produced during transposition and are believed to be intermediates in the process (Shiba and Saigo, 1983, Garfinkel \textit{et al.}, 1985, Mellor \textit{et al.}, 1985, Syomin \textit{et al.}, 1993, Haoudi, \textit{et al.}, 1995, Atwood \textit{et al.}, 1996). The VLPs
of Ty1 contain full length double-stranded Ty1 DNA and can catalyse the transposition of Ty1 RNA into a DNA target in vitro (Eichinger and Boeke, 1988).

Examples of this family in Drosophila melanogaster include gypsy (Modolell et al., 1983), copia (Mount and Rubin, 1985) and 17·6 (Saigo et al., 1984).

1.1.1.2 LINE-like elements

LINE-like elements lack LTRs and are characterised by an A-rich sequence at the 3' end of the coding strand (see Figure 1B). The Drosophila melanogaster I factor falls into this class. Their name derives from the mammalian long interspersed nuclear elements (LINEs, Singer et al., 1993). Although some LINE-like elements contain only a single open reading frame in most cases there are two, designated ORF1 and ORF2. ORF1 of several elements, including the I factor but not mammalian LINEs, contains a CX$_2$CX$_4$HX$_4$C zinc finger nucleic acid binding motif also found in retroviral gag genes (Fawcett et al., 1986, Jakubczak et al., 1990, Udomkit et al., 1995, Priimagi et al., 1988). ORF2 in all cases is homologous to reverse transcriptases. Accepted models for LINE transposition are based on in vitro studies of the R2Bm element of Bombyx mori (Luan et al., 1993). This belongs to a class of ribosomal insertion sequences that occur only at specific sites in the ribosomal DNA repeats (Burke et al., 1987). The transposition mechanism is known as target-primed reverse transcription (TPRT) and is detailed in Figure 2.
The transposition cycle begins with the transcription of the element RNA. R2Bm is probably transcribed as part of the 28S ribosomal RNA (George and Eickbush, 1999), while other LINE-like elements, which are found at dispersed positions in the genome, contain internal promoters (Mizhroki et al., 1988, Swergold, 1990, McLean et al., 1993). R2Bm encodes a single protein which has reverse transcriptase and endonuclease activities (Luan et al., 1993, Xiong and Eickbush, 1988). This initiates transposition by cleaving one strand of the chromosomal DNA at the target site (Xiong and Eickbush, 1988, Yang and Eickbush, 1998). Reverse transcription and integration occur simultaneously as the free 3' end at the nick is used as a primer for DNA synthesis using R2Bm RNA as a template. The reverse transcriptases of several LINEs have motifs similar to *E. coli* RNaseH (Fawcett et al., 1986, Blesa and Martínez-Sebastián, 1997) which may function to degrade the RNA template following synthesis of the first element DNA strand. The remaining strand could be synthesised by the reverse transcriptase or by cellular DNA repair enzymes (George et al., 1996).
The sequence specificity of R2 insertion arises from sequence-specific cleavage of the target (Luan et al., 1993). In contrast human LINE 1 inserts at a wide variety of sites and encodes a nuclease activity with a corresponding lack of specificity (Feng et al., 1996). The ORF2 proteins of several elements contain endonuclease domains homologous to the AP family of nucleases involved in DNA base excision repair (Martín et al., 1995, Feng et al., 1996, Feng et al., 1998, Blesa and Martínez-Sebastián, 1997). The endonuclease domain of human LINE 1 is required for transposition in tissue culture cells (Feng et al., 1996).

Ribonucleoprotein particles associated with reverse transcriptase activity and the ORF1 protein have been described for LINE elements (Martin, 1991, Hohjoh and Singer, 1996). These are believed to be intermediates in transposition. The ORF1 proteins of the I factor and mouse LINE 1 show self-association and RNA binding activity, consistent with a structural role in such complexes (Dawson et al., 1997, Kolosha et al., 1997).

In addition to mammalian LINEs and insect ribosomal insertion sequences, representatives of this class include the *Drosophila melanogaster* I factor (Fawcett et al., 1986), F (Di Nocera et al., 1983), G (Di Nocera and Dawid, 1983), doc (Bender et al., 1983), D (Pittler and Davis, 1987), BS (Udomkit et al., 1995) and jockey (Priimagi et al., 1988) and the telomeric Het-A and TART elements (Biesman et al., 1992b, Levis et al., 1993).

### 1.1.1.4 Class II elements

Class II elements transpose by a DNA-mediated cut-and-paste mechanism (Kaufman and Rio, 1992, Vos et al., 1996, Van Leunen et al., 1994). They encode a transposase protein which catalyses a site-specific recombination event excising DNA from one site and inserting it elsewhere in the genome. The termini of class II elements have inverted repeat sequences which provide a potential means by which the ends of the element can be recognised by protein factors. They are usually flanked by target site duplications which suggests that integration occurs at staggered double-strand breaks with some replication to fill in the resulting gaps.

One of the best characterised type II elements is the *Drosophila melanogaster* P element which has been exploited in the development of vectors for genetic
transformation (Rubin and Spradling, 1982). It is responsible for the P-M system of hybrid dysgenesis which is discussed below.

1.1.2 Hybrid dysgenesis

Transposable elements are responsible for the phenomenon of hybrid dysgenesis in *Drosophila*. Crosses between certain strains give rise to offspring with genetic defects including sterility and increases in chromosome abnormalities and mutations. This is associated with high frequency transposition of certain elements when introduced into a new cellular environment. The two systems of hybrid dysgenesis that have been extensively investigated, the P-M and I-R systems, are associated with different types of transposable elements the P element and the I factor respectively. The systems share certain biological features but are functionally independent and are regulated in different ways.

1.1.2.1 P-M hybrid dysgenesis

The Paternal-Maternal (P-M) system gives rise to dysgenic progeny when males from a P strain, containing autonomous P elements, are crossed to females from an M strain (lacking P elements). Progeny of both sexes show genetic defects associated with P element transposition in the germ line, including high mutability, chromosome rearrangements, male recombination, and sterility (Engels, 1989). The progeny of the reciprocal cross between males from an M strain and females from a P strain are phenotypically normal. This lack of reciprocity in crosses between the sexes is also observed in the Inducer-Reactive (I-R) system.

1.1.2.2 I-R hybrid dysgenesis

In the I-R system all *Drosophila melanogaster* strains can be classified as either inducer (I) or reactive (R). When inducer males are mated to reactive females the F1 female progeny show reduced fertility, a proportion of their eggs failing to hatch. This is independent of the strain of males to which these F1 females are mated. Such females are designated stérilité femelle (SF) and also show other dysgenic traits including an increase in X chromosome non-disjunction and a high frequency of mutations (Picard *et al* 1978, Proust and Prudhommeau, 1982). They also show an increased level of meiotic recombination (Chaboissier, Lemeunier and
Bucheton, 1995). The female progeny (RSF) of the reciprocal cross, between reactive males and inducer females, show normal fertility as do males from either type of cross.

The I factor is present in 10-15 copies per haploid genome on the arms of inducer strain chromosomes (Bucheton et al, 1984) while active I factors are absent in reactive strains. The presence of such elements is the sole determinant of the inducer state as the introduction of a cloned I factor into a reactive strain results in an inducer phenotype capable of causing hybrid dysgenesis (Pritchard et al 1988).

Mutagenesis in SF females appears to be the result of I factor transposition. Several mutations in the white gene were found to have full length I factor insertions (Bucheton et al, 1984) while chromosome rearrangements are also generated at high frequency during dysgenesis (Picard et al, 1978, Sang et al, 1984, Busseau et al, 1989, Prudhommeau and Proust, 1990, Proust et al., 1992). Deletions and inversions are found which have I element sequences, either complete or truncated at the 5' end, at the junction points (Busseau et al, 1989, Proust et al., 1992). These are not sites at which the parental chromosomes contained I elements implying that the rearrangements occur either during or shortly after integration. They could result from homologous recombination between newly inserted copies.

1.1.2.3 Other hybrid dysgenesis systems

In addition to the P-M and I-R systems, two other systems of hybrid dysgenesis have been described in Drosophila species. One is associated with the type II hobo element (Blackman et al 1987). Flies containing active hobo give dysgenic progeny when mated with flies from strains lacking the element. The progeny show mutations associated with novel hobo insertions and chromosomal rearrangements generated by hobo mobilisation.

In Drosophila virilis a cross between males of the strain L160 and females of the strain B9 gives progeny with dysgenic symptoms including gonadal dystrophy, poor fertility and a high mutation rate (Lozovskaya et al., 1990). Analysis of mutants identified novel insertions of five different transposable elements, named Paris, Helena, Penelope, Ulysses and Telemac, all of which were apparently mobilised in this cross (Petrov et al., 1995, Vieira et al., 1998). It is not clear if this represents the simultaneous activation of different elements that are regulated.
independently. Alternatively there may be a common mechanism underlying their mobilisation, such as the removal or reduction of a host system that represses transposable element activity.

1.2 The I factor

1.2.1 Structure

Autonomous I factors span 5-4kb and show several features seen in other LINE-like elements (see Figure 3, Fawcett et al 1986). The sequence is usually flanked by 10-14bp duplications of the genomic DNA at the insertion site which is consistent with integration occurring at staggered nicks. There are two large open reading frames on the same strand, designated ORF1 and ORF2, separated by 54bp (Fawcett et al, 1986, Abad et al., 1989). At the 5’ end of the coding strand there is a 186bp untranslated region (5’UTR) containing an internal promoter and sequences involved in the regulation of transcription (Mclean et al 1993, Udomkit et al 1996). At the 3’end a 179bp untranslated region terminates in a variable number of TAA repeats.

1.2.2 RNA

A full length 5-4kb RNA is transcribed from the I factor during transposition (Chaboissier et al, 1990). RNA is known to be an intermediate in the process as introns within cloned I factors are found in the transposed copies to have been spliced precisely (Péisson et al, 1991, Jensen and Heidmann, 1991, Chaboissier et al, 1995, Busseau et al., 1998). The full length transcript is believed to function both as the substrate for reverse transcription and as a bicistronic messenger for the translation of the ORF1 and ORF2 proteins. Translation can initiate in vivo at low frequency from the ATG at the start of ORF2 (Bouhidel et al 1994). A number of mechanisms have been proposed for this, all of which are likely to be inefficient (Bouhidel et al 1994, Chaboissier et al 1990), which is consistent with the proposed transposition mechanism in which the product of ORF1, as a structural component of ribonucleoprotein particles, would be required in greater quantity than the catalytic ORF2.
In addition to the full length product of active $I$ factors, transcripts are produced from inactive $I$ elements located in the pericentromeric heterochromatin (see Section 1.2.4, Chaboissier et al, 1990). It has been suggested that they affect the regulation of $I$ factor transposition (Bucheton, 1990, Jensen et al., 1995).

**Figure 1.3** $I$ factor primary structure showing open reading frames, conserved motifs and untranslated regions. The 5' untranslated region is magnified to show the known regulatory elements.

### 1.2.3 Proteins

ORF1 contains one complete and one partial copy of the zinc finger motif $\text{CX}_{2}\text{CX}_{3}\text{H}_{2}\text{C}$ which is highly conserved in the products of retroviral $\text{gag}$ genes and is thought to be involved in RNA binding (Fawcett et al, 1986). ORF1 protein expressed in *E. coli* has a strong nucleic acid binding activity and is capable of multimerisation to form a high molecular weight nucleic acid-protein complex (Dawson et al., 1997). These properties, and analogy with other LINE-like elements, suggest a structural role in the formation of nucleoprotein particles. The protein promotes the annealing of complementary DNA strands, an activity also shown by the $\text{gag}$ gene product of human immunodeficiency virus 1 (Dawson et al., 1997,
Tsuchihashi and Brown, 1994). During transposition in SF females ORF1 accumulates in the cytoplasm of the oocyte which has led to the suggestion that it may be involved in transportation of I factor ribonucleoprotein particles from the nurse cells towards the oocyte nucleus (Seleme et al., 1999).

The putative product of ORF2 has short regions which are conserved in reverse transcriptases from retroviruses and transposable elements. Sequence comparisons show it is more similar to the reverse transcriptases of other LINEs and type II introns than to those of retroviruses or retrovirus-like transposable elements (Fawcett et al, 1986). There is a conserved motif characteristic of RNase H activity which is also found in other reverse transcriptases. An N-terminal region of ORF2 has similarities to AP endonucleases and exhibits in vitro endonuclease activity (Danielle Teninges and Alain Bucheton, unpublished results). Marked I elements lacking a functional ORF2 can be mobilised in trans by actively transposing I factors in the female germ line (Pélisson et al., 1991, Chaboissier et al., 1995) or in tissue culture cells by the expression of ORF2 protein from a trans gene (Jensen et al., 1994, Busseau et al., 1998).

1.2.4 Defective pericentromeric I elements and the origins of the Drosophila melanogaster I factor

In addition to functional I factors, which are present in inducer strains, all Drosophila melanogaster strains have defective I sequences located in the pericentromeric β heterochromatin (Bucheton et al, 1984, Vaury et al, 1989). This region divides the compact α heterochromatin, which consists of simple sequence repeats, from the largely euchromatic chromosome arms and contains many sequences related to transposable element families (Vaury et al., 1989). The location of defective I elements is similar in different reactive and inducer strains suggesting they are old and immobile (Bucheton et al., 1984). The pericentromeric elements differ from complete I factors in a number of ways such that none that have been sequenced could transpose autonomously (Crozatier et al., 1988, Vaury et al., 1990, Sezutsu et al., 1995). They appear to be derived from a common transposable ancestor not identical to the active I factors found in modern inducer strains. Studies of the distribution of I element sequences in related Drosophila species suggest that
this ancestor was active in flies before the speciation of Drosophila melanogaster and has subsequently been lost (Bucheton et al., 1986, Sezutsu et al., 1995). Modern active I factors are thought to have invaded Drosophila melanogaster some time in the first half of this century from one of its sibling species, perhaps Drosophila simulans which contains sequences more than 99% identical to active melanogaster I factors (Sezutsu et al., 1995). A more distantly related I factor cloned from Drosophila teissieri has been shown to transpose and cause hybrid dysgenesis when introduced into Drosophila melanogaster (Abad et al., 1989, Vaury et al., 1993) which supports the view that such an invasion is possible. All wild caught Drosophila melanogaster now contain active I factors while laboratory stocks established from flies caught before 1930 do not (Kidwell, 1993). Probably the isolation of these stocks from wild populations has prevented their invasion.

1.3 Transposable element regulation

Due to its mutagenic effects, transposition has serious consequences for the host organism and therefore for the survival of the transposable element itself. This is dramatically illustrated in hybrid dysgenesis, where mutations and sterility are thought to result from high frequency transposition. Consequently a variety of mechanisms have evolved that limit transposition including autoregulation by element-encoded molecules as well as host defence systems that protect the integrity of the genome. In multicellular organisms there is also frequently tissue-specificity and developmental regulation of transposition. In somatic cells transposition can reduce the fitness of the host while not contributing to the element’s survival by increasing the copy number in the next generation. Many elements have evolved ways of exploiting host factors to restrict transposition to the germ line.

1.3.1 Regulation by the host

Several host genes are known that affect the expression or the transposition rate of certain elements. Transposition of Tyl in Saccharomyces cerevisiae is affected by a number of factors including chromatin proteins and Rad6p, a ubiquitin conjugating enzyme (Picologlou et al., 1990, Qian et al., 1998 and references therein). In Drosophila several mutations affect levels of copia RNA (Rabinow et al., 1993, Csink et al., 1994a,b, Birchler et al, 1994) and the transposition of gypsy is

In many organisms there are silencing mechanisms that inactivate DNA present in multiple copies or otherwise recognised as foreign. Frequently inactivation is accompanied by methylation of cytosines, although silencing has been observed in *Drosophila melanogaster* and *Caenorhabditis elegans*, organisms that lack methylated DNA. Methylation is associated with the formation or maintenance of a repressive chromatin structure that can inhibit transcription (Antequera *et al.*, 1990, Siegfried *et al.*, 1999) and recombination (Maloisel and Rossignol, 1998). It may also lead eventually to the degradation of functional sequences in the silenced DNA through mutagenesis by spontaneous deamination of 5-methyl cytosine to thymine followed by mismatch repair. Silencing mechanisms have generally been discovered as a result of transgenic experiments, in which repeated DNA has been introduced artificially into an organism, but it has been speculated that their normal function is to provide a defence for the genome against viral infection and the proliferation of transposable elements.

1.3.1.1 RIP and MIP in fungi

In the sexual cells of some filamentous fungi, repeated sequences are recognised and inactivated prior to meiosis by mechanisms involving DNA methylation. In *Neurospora crassa* the process of repeat induced point mutation (RIP) results in multiple G-C to A-T base pair transitions that affect both copies of a duplicated sequence (Selker *et al.* 1987, Selker and Garrett, 1988, Cambareri *et al.*, 1989). Where three copies of a sequence are present, RIP most frequently affects only two of them with all three being mutated in a minority of cases (Fincham *et al.*, 1989). Mutation of one out of two or three copies is never observed, supporting a model involving the physical pairing of identical sequences at the DNA level as a trigger for the mutagenesis of both copies. The products of RIP show extensive methylation of cytosines (Fincham *et al.*, 1989).

Another filamentous fungus, *Ascobolus immersus*, exhibits MIP (methylation induced premeiotically) in which repeated sequences are heavily methylated and transcriptionally repressed (Goyon and Faugeron, 1989). This shares many features with RIP including the pairwise inactivation of repeats (Faugeron *et al.* 1990).
Methylation induced by MIP affects transcriptional elongation resulting in gene silencing (Barry et al., 1993) and also represses recombination (Maloisel and Rossignol, 1998) so may play a role in preserving the integrity of the genome.

It has been suggested that RIP and MIP protect the genome from the effects of mobile elements (Selker and Garrett, 1988). Evidence to support this view comes from the discovery of sequences related to transposable elements that show features characteristic of the products of RIP (Kinsey et al., 1994, Margolin et al., 1998, Cambarei et al., 1998) and MIP (Goyon et al., 1996).

1.3.1.2 Methylation and the epigenetic regulation of plant transposable elements

The contribution of DNA methylation to the regulation of transposable elements in maize has been studied in some detail. At least two families of class II elements, Spm and Ac, are subject to reversible transcriptional inactivation, accompanied by the methylation of cytosines in their promoter regions. The Tam3 element of Antirrhinum majus can also be inactivated by methylation (Martin et al., 1989).

Spm elements exist in the maize genome in inactive as well as actively transposing forms. Inactive elements vary with respect to the stability of the inactive state. For some elements ("cryptic" Spm), the inactive state is highly stable and heritable, spontaneous reactivation occurring only rarely (Fedoroff, 1989). Other elements cycle frequently between the two states during plant growth, either apparently randomly, or according a heritable developmental programme (Fedoroff and Banks, 1988).

Both the active and the inactive forms of Spm are methylated throughout most of the element sequence (Banks et al., 1988). However active elements differ from inactive elements in that they are not methylated in the upstream control region, a 0.2kb sequence near the 5' end which contains the Spm promoter (Banks et al., 1988, Raina et al., 1993). Methylation in this region suppresses transcription of the genes required for transposition and also inhibits transposition of the element directly, even when the transposase proteins are provided in trans (Banks et al., 1988). Methylation in the downstream control region, a G-C-rich segment within the first exon, correlates with the stability of the inactive state, affecting the probability
that an element will reactivate spontaneously. Cryptic elements show greater methylation in this region than inactive elements that normally cycle between active and inactive forms (Banks et al., 1993). A protein expressed from active elements, TnpA, is able to reactivate inactive elements and to cause them to become demethylated (Schläppi et al., 1993).

The Ac element also shows reversible inactivation accompanied by changes in methylation, showing many similarities with the process observed for Spm elements. The inactive copies show methylation throughout the element, including in a region towards the 5' end which contains a promoter for the transposase gene as well as binding sites for the Ac transposase protein (Schwartz and Dennis, 1986, Chomet et al., 1987, Brutnell and Dellaporta, 1994). Elements methylated in this region show very little transcription (Kunze et al., 1988, Brutnell and Dellaporta, 1994), but can be mobilised by transposase provided by active Ac elements elsewhere in the genome (Schwartz and Dennis, 1986, Wang and Kunze, 1998). The inactive state is unstable, with spontaneous reactivation occurring at a low frequency (Schwartz and Dennis, 1986). Reactivation is accompanied by a partial removal of methylation at the 5' end and the restoration of transcription (Kunze et al., 1988, Brutnell and Dellaporta, 1994) and can be greatly accelerated by the presence of active Ac elements (Schwartz, 1989). It is believed that the Ac transposase promotes demethylation of the promoter allowing the reactivation of transcription.

In these systems it is not clear that methylation is triggered by the repetitive nature of transposable element sequences. When Spm elements are introduced into tobacco, which lacks endogenous copies of Spm, they can transpose and can also be inactivated and methylated (Schläppi et al., 1993). Reporter gene experiments demonstrated that the G-C-rich downstream control region must be present for methylation and inactivation to occur (Schläppi et al., 1994), suggesting that features of the transposable element sequences are necessary for recognition by the silencing system. Other elements can also be epigenetically inactivated in heterologous organisms (Scortecci, et al., 1996, Martin et al., 1989).

1.3.1.3 Repeat associated gene silencing in plants

In plants, transgenes introduced artificially are frequently subjected to silencing, particularly if they contain tandemly repeated sequences. A distinction is
made between transcriptional and post-transcriptional silencing mechanisms, which show different features, although it is not clear that the two are entirely independent. Transcriptional gene silencing (TGS) is accompanied by methylation of promoter regions (Amasino et al., 1984, Linn et al., 1990, Assaad et al., 1993, Matzke et al., 1993) and changes in the chromatin structure at the silenced loci (Ye and Signer, 1996). Copies of endogenous genes with sequences similar to a silenced transgene may also be methylated and silenced, a process known as co-suppression, which has led to the proposal that silencing can result from DNA-DNA interactions (Meyer et al., 1993, Park et al., 1996, Luff et al., 1999). There is evidence that DNA-RNA interactions can promote methylation and this has also been suggested as a possible mechanism for the transmission of silencing information throughout the genome (Wassenegger et al., 1994, Mette et al., 1999, Pelissier et al., 1999).

Post-transcriptional gene silencing (PTGS) involves the sequence-specific degradation of RNA derived from transgenes (Metzlaff et al., 1997, van Eldik et al., 1998). Endogenous genes homologous to the transgene may be co-suppressed (Napoli et al., 1990, van der Kroll et al., 1990, Smith et al., 1990, de Carvalho et al., 1992, Seymour et al., 1993, Boerjan et al., 1994). Genes subject to PTGS frequently show DNA methylation, in this case typically in the transcribed region of the gene rather than in the promoter (Ingelbrecht et al., 1994, van Houdt et al., 1997, Jones et al., 1998, Guo et al., 1999). The mechanism that triggers PTGS is unclear but probably involves aberrant RNA species which could be produced by a variety of mechanisms. It has been proposed that they represent a substrate for an RNA-dependent RNA polymerase which synthesises antisense copies complementary to the transgene mRNA. These may hybridise with the mRNA resulting in its degradation by a ribonuclease that recognises double-stranded RNA. Support for such a model has come from the discovery that a protein homologous to an RNA-dependent RNA polymerase is required for quelling, a post-transcriptional gene silencing mechanism operating in somatic cells in *Neurospora crassa* (Cogoni et al., 1999).

1.3.1.4 Silencing, methylation and transposable elements in animals

In mammalian cells transgenes integrated into the genome are sometimes found to be silenced at the transcriptional level (Pikaart et al., 1998, Garrick et al., 1998, Garrick et al., 1998).
1998), particularly if they contain viral sequences (Chalita and Kohn, 1994, Chen et al., 1997). Silencing is associated with changes in chromatin structure and may correlate with increased methylation (Chalita and Kohn, 1994, Pikaart et al., 1998, Garrick et al., 1998). In many cases silencing can be attributed to position effects but in one example the number of repeats inserted at a single location has been shown to affect the probability of silencing (Garrick et al., 1998) which may reflect a system analogous to repeat induced transcriptional gene silencing in plants. A proposal that methylation plays a role in the defence of mammalian genomes against transposable elements is the subject of much controversy (Yoder et al., 1997a, b, Bird, 1997, Simmen et al., 1999, Walsh and Bestor, 1999). Evidence for transcriptional gene silencing in *Drosophila melanogaster* will be discussed below in the context of I factor regulation.

Some evidence has emerged recently for the existence of post-transcriptional silencing mechanisms in animals. The injection of double-stranded RNA containing sequences present in endogenous genes results in a reduction in their expression in *Caenorhabditis elegans* and *Drosophila melanogaster* (Fire et al., 1998, Kennerdell and Carthew, 1998). In one case this has been shown to be due to a post-transcriptional mechanism (Montgomery et al., 1998). Double-stranded RNA is a particularly potent inducer of PTGS in plants (Waterhouse et al., 1998) and will also silence genes in *Trypanosoma brucei* (Ngo et al., 1998) which may imply that a highly conserved silencing mechanism is involved in all of these examples.

### 1.3.2 Autoregulation

Many transposable elements are believed to contribute to their own regulation. Overexpression of the transposase of the class II mariner element *mos1* results in a reduction rather than an increase in transposase activity, measured by the excision of a defective element inserted in the *white* gene (Lohe and Hartl, 1996). The mechanism of this effect, known as overproduction inhibition, is unknown. Other class II elements produce proteins that restrict transposase activity. These include the *Drosophila P* element, whose regulation is discussed in detail below and the maize *Spm* element which encodes the TnpA protein which inhibits transcription from the element promoter in its unmethylated state (Schläppi et al., 1994).
Many class II elements exist in genomes in two forms. A full length autonomous form is competent for transposition but there are also elements with internal deletions or other mutations that are unable to encode a functional transposase (O'Hare and Rubin, 1983, Kim and Kim 1999). It is proposed that these elements contribute to the regulation of transposition in two ways. They may encode mutant versions of transposase proteins which could inhibit transposition either by competing with the transposase for binding sites on the element DNA (Lee et al., 1998), by affecting transcription of the element (Kaufmann and Rio, 1991), or by interacting with functional transposase molecules by protein-protein interactions to form inactive multimers (Andrews and Gloor, 1995, Lohe et al., 1996). Alternatively nonautonomous elements may regulate transposition by titration if they contain transposase binding sites that would compete with those of functional elements for the transposase protein. (Simmons and Bucholz, 1985, Rasmusson et al, 1990).

1.3.3 Regulation of P element transposition

1.3.3.1 Introduction - tissue-specificity and the P cytotype

There are at least two levels of regulation acting on the P element, the causative agent of P-M hybrid dysgenesis. Transposition is restricted to germ line tissue (McElwain, 1986) and is also subject to an autoregulatory mechanism known as cytotype. In strains having P elements (P strains), transposition is inhibited, a condition known as the P cytotype. Conversely strains lacking P elements (M strains) have the M cytotype, a condition permissive for transposition. The P cytotype is inherited maternally, accounting for the absence of dysgenic symptoms in the progeny of crosses between P strain females and M strain males. In the dysgenic cross between M females and P males the progeny are not affected by a maternal P cytotype and experience high frequency transposition and the symptoms of dysgenesis.
1.3.3.2 Tissue-specificity in P transposition is regulated at the level of pre-mRNA splicing

Transposition is catalysed by the 87kD P transposase, which is encoded by autonomous P elements in four exons. These must be spliced together to produce the transposase-encoding mRNA. In somatic cells, splicing of the 3rd intron does not occur (Laski et al., 1986), due to the binding of inhibitory proteins to a regulatory element in the 5' exon RNA (Siebel et al., 1992). A complex is formed which includes the P element somatic inhibitor (PSI), a protein whose presence at high levels in somatic tissue but not in the germ line is sufficient to account for the observed tissue-specificity (Siebel et al., 1995, Adams et al., 1997). In the absence of 3rd intron splicing, P element RNA produces a shorter protein (66kD) which acts as a repressor of transposition (Misra and Rio, 1990, see below).

1.3.3.3 The role of internally-deleted P elements

Of the 40-50 P elements present in a typical P strain approximately one third are full length (2.9kb) and are capable of encoding the 87kD transposase. The rest have internal deletions and are non-autonomous but can transpose when the enzyme is provided in trans (O'Hare and Rubin, 1983). Some of these non-autonomous elements encode truncated derivatives of transposase which are capable of repressing transposition in vivo (Black et al., 1987, Simmons et al., 1990, Rasmusson et al., 1993) and may contribute to the repressive P cytotype.

1.3.3.4 Mechanisms of cytotype repression

Numerous mechanisms have been proposed to explain repression of transposition by the P cytotype. Most of these assume that repression depends on P-encoded proteins, the 66kD product of full length P elements as well as the products of internally-deleted elements.

1.3.3.4.1 The multimer poisoning model

One of the most common internally-deleted P elements is the KP element which has been the subject of considerable study. This encodes a protein of 207 amino acids which include the first 199 amino acids of the transposase. This region contains two protein-protein interaction domains, a leucine zipper and a C-terminal
basic region, which allow the repressor to dimerise (Lee et al. 1996). Mutagenesis of
the leucine zipper abolishes the ability of the protein to repress transposition in vivo
(Andrews and Gloor, 1995). This supports a “multimer poisoning” model in which
the truncated protein participates with the transposase in the formation of complexes
which are inactive in transposition. However mutagenesis of the leucine zipper does
not in itself abolish dimerization of the repressor in vitro (Lee et al., 1996). It has
been suggested that the mutations tested in vivo may have had a general effect on the
structural integrity of the protein that results in a loss of activity (Lee et al., 1998).

1.3.3.4.2 Direct interference in transposition by repressor DNA binding

The KP protein also contains a DNA binding domain which recognises the
transposase binding sites internal to each of the two P element ends (Lee et al.,
1996). It also binds to an 11bp transposition enhancer element and to parts of the
31bp terminal inverted repeats. This domain is required for the KP protein to repress
catalysis by transposase in vitro suggesting that it acts by competing for interactions
of transposase with P element DNA (Lee et al., 1998).

1.3.3.4.3 Antisense RNA

Insertion of a P element downstream of, and in an opposite orientation to a
cellular promoter could produce antisense P RNA. Dysgenesis can be repressed by
antisense P products (Simmons et al., 1996). This mechanism may explain the
repressive effects seen with a small P element (SP) having the potential to encode a
polypeptide of only 14 amino acids, considered unlikely to have an effect at the
protein level (Rasmusson et al., 1993). There is evidence that the effect of SP is
position-dependent which would be consistent with a model that relies on an external
promoter (unpublished results quoted in Simmons et al., 1996).

1.3.3.4.4 Transcription and the P cytotype - trans-silencing

The P cytotype inhibits transcription from the P element promoter in
germline and somatic tissue (Lemaitre and Coen, 1991, Lemaitre et al., 1993, Roche
et al., 1995). The 5’ transposase binding site, which is also recognised by truncated
P repressor proteins, overlaps the TATA box of the P promoter (Kaufman et al.,
1989) and full length transposase can repress transcription in vitro by interfering with
the binding of TFIID (Kaufman and Rio, 1991). In vivo the matter is more complicated as transcription from heterologous female germline promoters that do not contain the transposase binding site is also inhibited by the P cytotype when they are inserted into P element constructs (Roche et al., 1995). This suggests a more general mechanism of transcriptional repression, perhaps involving alterations in chromatin structure. Roche and Rio (1998) investigated the regulatory effects of two P elements inserted at cytological position 1A near one of the telomeres of the X chromosome. Flies containing only these two elements are able to repress transposition in the germ line (Ronserray et al., 1991). These elements were able to repress transcription from a non-autonomous reporter P element containing the hsp83 promoter. Repression was abolished in flies with mutations in two genes Enhancer of zeste E(z) and Su(var)205. E(z) encodes a member of the Polycomb group of proteins (PcG) which play a role in maintaining a repressed state of homeotic genes established early in development (Chan et al., 1994, Kassis, 1994, Gindhard and Kaufman, 1995). Su(var)205 encodes heterochromatin protein 1, a protein that colocalises with heterochromatin (James and Elgin, 1986). It is proposed that a repressive chromatin structure is established at the reporter P elements as a result of interactions with the telomeric regulatory elements, mediated by the recognition of sequence homology (Roche et al., 1998). This mechanism, known as ‘trans-silencing’, is analogous to models of RIP and MIP in fungi and repeat associated gene silencing in plants and is similar to an idea proposed for I factor regulation that is discussed below.

1.3.3.5 Maternal inheritance of the P cytotype

Repression by the P cytotype shows a maternal effect, acting on the progeny of P strain females. Repression in the germ line also shows maternal inheritance through several generations (Engels, 1979). Maternal inheritance requires a positive feedback mechanism, so that repressive products inherited cytoplasmically are maintained at high levels from one generation to the next. One hypothesis proposed to explain maternal inheritance is based on the observed inhibition of P element transcription in the P cytotype. A repressor of transcription, perhaps the 66kD protein, could be inherited cytoplasmically and inhibit zygotic transcription of P elements in the embryonic germ line. When P element mRNA levels are low, the
efficiency of splicing of the third intron is reduced, leading to the production of more of the 66kD repressor (Lemaitre et al., 1993, Roche et al., 1995). This could be due to the presence of a low level of the PSI splicing inhibitor in the germline which is titrated in the presence of high quantities of P RNA during dysgenesis. Low levels of transcription would therefore lead to the production of more repressor providing a positive feedback loop by which the repressive P cytotype could be maintained maternally through the generations.

1.3.4 Regulation of \textit{I} factor transposition

1.3.4.1 Levels of regulation

The features of I-R hybrid dysgenesis reflect several different levels of regulation of \textit{I} factor transposition. Transposition occurs in the germline of dysgenic females and can be detected in the phenomenon of chromosome contamination (Picard, 1976). In a dysgenic cross maternal chromosomes, which originally lack functional \textit{I} factors, become ‘contaminated’ at high frequency by \textit{I} factors transposing onto their arms. These chromosomes now segregate as determinants of the inducer phenotype and can cause hybrid dysgenesis when males carrying such chromosomes are mated to reactive females (Picard, 1976). Chromosome contamination has not been detected in the male germ line or in the female germ line of inducer stocks (Picard, 1976), indicating that transposition occurs at low frequency or not at all in these tissues. In \textit{RSF} females chromosome contamination does occur, but at a lower rate than in \textit{SF} females, reflecting a transposition frequency which is apparently not sufficient to cause a detectable reduction in fertility.

1.3.4.1.1 Tissue-specificity

Dysgenic females show no adverse symptoms other than reduced fertility suggesting that \textit{I} transposition is restricted to the female germ line. This conclusion is supported by two studies in which the \textit{I} element copy number in salivary gland chromosomes has been determined by \textit{in situ} hybridisation. In flies in which transposition was occurring in the germ line, the copy number was constant for
different salivary gland nuclei from the same individual, providing no evidence for somatic transposition (Pelisson and Bregliano, 1987, Pritchard et al., 1988).

1.3.4.1.2 Copy number control

The functional $I$ factors present in inducer strains are inactive. Females show normal fertility and chromosome contamination does not occur in the germ line suggesting that transposition is repressed. When $I$ factors are introduced into a reactive background, transposition occurs at high frequency for a limited period during which females show reduced fertility. After this period a stable inducer state is attained in which the transposition frequency is low (Picard, 1978a, Pritchard et al., 1988). These observations have led to the suggestion that transposition is repressed when the copy number reaches a certain threshold.

Pélisson and Bregliano (1987) looked at $I$ factor copy number through several generations of flies derived from a dysgenic cross. The reactive genotype was reconstituted by mating $SF$ females to males of the original reactive strain and discarding progeny that inherited markers present on chromosomes from the inducer strain. The stock was maintained by sibling matings selecting at each generation the progeny of females that showed low fertility. It was found that after three generations the $I$ factor copy number reached around 6 per haploid genome and this level remained stable for several subsequent generations. The authors therefore suggested that $SF$ sterility does not depend on transposition as there was no increase in copy number while sterility was maintained. As pointed out by McLean (1991), an alternative explanation is that germline cells in which transposition has occurred give rise either to eggs that fail to hatch or to flies with a higher $I$ factor copy number. In these flies transposition would be repressed in the female germ line so their fertility would be normal and they would be discarded from the experiment. The germline cells in which transposition has not occurred give rise to females with 6 $I$ factor copies, a number which allows transposition to proceed in the majority of germ cells so the female adults show reduced fertility and are selected to give rise to the next generation. The data are therefore compatible with the existence of a copy-number dependent mechanism regulating $I$ factor transposition.
1.3.4.1.3 Reactivity

The severity of the dysgenic phenotype in SF females varies considerably and depends on the maternal strain (Bucheton et al, 1976). This defines the property of "reactivity", which is measured by crossing a strain to a standard inducer line and counting the hatching percentage of the eggs laid by the female progeny.

Couples between strains that have different levels of reactivity show that most progeny inherit a level similar to that of their mothers (Bucheton and Picard, 1978). There is a clear difference in the progeny of reciprocal crosses demonstrating that inheritance of reactivity is maternal. However some of the progeny of such crosses show a level intermediate between those of the parental strains indicating that there is some paternal influence (Bucheton and Picard, 1978). This influence is tightly linked to the inheritance of the three major chromosomes and has no cytoplasmic element (Bucheton and Bregliano, 1982). The effects of chromosomes are additive, defining reactivity as a quantitative multigenic trait (Bucheton and Picard, 1978).

The actions of chromosomes on reactivity accumulate through several generations, eventually resulting in a reactivity level showing no dependence on the initial maternal state (Bucheton and Picard, 1978). This supports a model in which reactivity is influenced by an inherited cytoplasmic or epigenetic condition expressed in the oocyte and ultimately dependent on the genotype.

The frequency of I factor transposition, as measured by the efficiency of chromosome contamination, is dependent on the reactivity level (Picard, 1978b). There is therefore a correlation between SF sterility and I factor transposition suggesting a causal relationship.

1.3.4.1.4 Influence of age, temperature and DNA damage on reactivity

Reactivity is sensitive to environmental influences and to the age of the female in which the oocyte is developing. Keeping flies at 29°C, rather than 20°C, results in an increase in the fertility of SF females (Bucheton, 1978). This effect is also seen when the temperature treatment is applied to the reactive mothers of SF females, demonstrating that it is heritable (Bucheton, 1979a). The age of females also has an effect on reactivity. The fertility of SF females increases with age.
Correspondingly, flies hatching from eggs laid by older reactive females have a lower reactivity than the progeny of younger females (Bucheton, 1978).

The effects of heat treatment and age are cumulative over several generations, resulting in a gradual change in the reactivity of the stock. These effects are reversible, a change to the original conditions (temperature or time at which eggs are collected) resulting in a gradual return of the stock to its original level of reactivity (Bucheton, 1979b).

Recently it has been shown that reactivity is also affected by treatments causing DNA damage, including gamma irradiation and the application of drugs which inhibit nucleotide synthesis (Bregliano et al, 1995). Again the effects are heritable, cumulative and reversible. It is suggested that reactivity represents an inducible DNA repair or recombination system operating in the Drosophila female germ line analogous to the SOS response of E. coli (Bregliano et al, 1995). This model predicts that weakly reactive stocks will show a greater sensitivity to DNA damage than isogenic strongly reactive stocks and there are data to support this view (Laurençon and Bregliano, 1995, Laurençon et al., 1997).

1.3.4.2 \( I \) factor transcription

There is good evidence that \( I \) factor transposition is regulated at the level of transcription. The presumed RNA transposition intermediate was detected by northern blotting in the ovaries of SF and, at a much lower level, RSF females but not in somatic tissues, males or inducer ovaries (Chaboissier et al, 1990). Its presence therefore correlates with transposition. This conclusion is disputed by de la Roche Saint André and Bregliano (1998) who detected \( I \) factor RNA in inducer ovaries at levels greater than in SF and RSF ovaries using a reverse transcriptase PCR assay and concluded that there are post-transcriptional controls on \( I \) factor transposition.

As detailed below, reporter genes under the control of the \( I \) factor promoter and regulatory region are expressed to the highest levels in the germ line of reactive females, the environment that is permissive for transposition. Reporter expression is sensitive to reactivity levels and is repressed in inducer flies (Lachaume et al, 1992, Lachaume and Pinon, 1993, McLean et al., 1993, Udomkit et al, 1996).
1.3.4.2.1 The \( I \) factor promoter

Like other LINE elements the \( I \) factor is transcribed from an internal RNA polymerase II promoter. The first 40bp are sufficient to direct transcription starting precisely at the 5' end (McLean et al, 1993). There is no TATA box as is found in typical upstream promoters. Instead, the promoter falls into a class of internal RNA polymerase II promoters which includes those of the mouse TdT, human porphobilinogen deaminase, *Drosophila Antennapaedia*, *engrailed*, *Ultrabithorax* and E74 genes (McLean et al, 1993 and references therein). The promoters of several other LINE-like elements also fall into this group. There are two conserved elements, an essential CA(G/T)T motif at position +1--i-4 and an AGACGTGTGPYPy motif at +28--i-41 which is not essential but stimulates \( I \) factor transcription two fold (McLean et al, 1993). This latter element is essential for jockey transcription (Mizrokhi et al., 1988).

1.3.4.2.2 The enhancer

In addition to the promoter, the 5' untranslated region contains other \( cis \) elements involved in the regulation of transcription. The region between nucleotides +41 and +186 confers enhanced expression of lacZ under the control of an hsp70 promoter in the female germ line (Udomkit et al, 1996). Staining of ovaries with X-gal reveals expression only in the germ-line nurse cells and oocyte. This occurs with the \( I \) factor sequence in either orientation. This experiment demonstrates that regulation of germ line expression occurs at the level of transcription, not RNA stability, as the predicted transcript contains no \( I \) factor sequences. Nucleotides +99 to +186 also confer germline specificity, although expression is reduced with respect to the longer construct. This identifies sequences that activate transcription within both the regions +41 to +98 and +99 to +186 (Udomkit et al, 1996).

The effect of sequences in the 5'UTR on expression from the natural \( I \) factor promoter has also been studied. A translational fusion consisting of part of ORF1 protein linked to \( \beta \) galactosidase downstream of the \( I \) factor 5' end is expressed only in the germ line cells of reactive and dysgenic females (Lachaine et al., 1992). The 186bp 5' UTR confers around 20-fold greater expression of the chloramphenicol acetyl transferase reporter gene (CAT) in ovaries than in non-ovarian tissue (McLean et al., 1992).
et al., 1993, Udomkit et al., 1996). The region +1 to +100 confers 4 fold activation in ovaries while the region +1 to +40, which still contains the promoter, gives two fold greater expression in non-ovarian tissue. This confirms that there are enhancers promoting ovary expression in the regions +40 to +100 and +100 to +186. It is reasonable to assume that they contribute to the tissue specificity of I factor transposition.

A sequence-specific DNA binding protein was detected in ovarian extracts that binds to base pairs 139-157, referred to as site 1, in gel retardation and DNase I footprinting assays (Udomkit et al., 1996). More precise DNase I footprinting identified the protected sequence as ACAAAAAACAACAT, representing base pairs 137 to 150 (Udomkit et al., 1996). A construct containing the 5'UTR-CAT reporter with site 1 deleted gave a drastically reduced level of expression and no difference in expression between ovary and carcass (Udomkit et al., 1996). The site 1 binding protein was therefore believed to be important in the regulation of the tissue-specificity of I factor transcription.

1.3.4.3 Models of I factor regulation

Several models have been proposed to account for the autoregulation of I factor transposition in inducer strains. The expression of reporter genes transcribed from the I factor promoter is around 30 fold lower in inducer compared with reactive ovaries (Lachaume et al., 1992, Udomkit et al., 1996). In those experiments the expected RNA transcripts contained the 5'UTR sequence so it is not clear whether regulation is at the level of transcription or if the stability of the RNA could be affected by the presence of I factor sequences.

1.3.4.3.1 I-encoded repressor model

By analogy with the P element it has been suggested that the I factor encodes a repressor protein which accumulates over several generations following the introduction of I factors into a reactive background, resulting in repression of transposition and a stable inducer stock (Jensen et al, 1995). It is proposed that this factor is inherited cytoplasmically in the oocyte, which would account for the difference in transposition frequency between genetically identical SF and RSF.
females. Flies whose mothers were inducer would inherit the repressor while those
with reactive mothers would not.

Jensen et al. (1995) found that a defective I element introduced into a reactive
strain reduces the reactivity over several generations eventually preventing I-R
hybrid dysgenesis. This element contains ORF1 and the N-terminal 322 codons of
ORF2 which potentially could encode regulatory proteins. One drawback of this
model is that the product of ORF1 does not have the properties expected of a specific
transcriptional repressor. The protein binds nucleic acid in a sequence non-specific
manner, with a higher affinity for RNA than DNA (Dawson et al, 1997). It is
therefore unlikely to interact specifically with I factor DNA to repress transcription.

1.3.4.3.2 Titration and homology-dependent gene silencing models

To test whether copy number regulation requires I factor proteins Chaboissier
et al. (1998) introduced different numbers of copies of the 5' untranslated region into
a reactive strain. The authors found a copy number-dependent reduction both in SF
sterility following dysgenic crosses and in the expression of a reporter gene
transcribed from the I factor 5'UTR. This demonstrates that the 5'UTR alone is
capable of regulating RNA levels and transposition in the absence of coding
sequences.

Two alternative interpretations could account for this result. There may be
transcription factors in the female germ line required for I factor expression that are
present at limiting concentration. When the copy number increases these factors
would be titrated so that a productive transcription initiation complex could not
assemble at any one I factor copy (Udomkit et al., 1996). The protein that binds to
site 1 is not a candidate for such a factor as copies of the 5'UTR with site 1 deleted
are almost as efficient at regulating transcription as the wild type sequence
(Chaboissier et al., 1998). This model does not account for the difference in
transposition frequency between SF and RSF females.

An alternative model proposes copy number dependent repression reflects the
action of an epigenetic homology-dependent silencing mechanism. It was argued
that the mechanism would not require transcription of the 5'UTR RNA as deleting
site 1 in the 5'UTR copies, which should reduce transcription drastically (Udomkit et
al., 1996), does not have a significant effect on repression (Chaboissier et al., 1998).
There is evidence for homology-dependent silencing mechanisms in *Drosophila melanogaster*. The discovery that double-stranded RNA can repress gene expression (Kennerdell and Carthew, 1998) suggests that a post-transcriptional silencing system may exist. Dörer and Henikoff (1994) found inactivation of a *white* transgene array in a process showing similarities with repeat associated transcriptional gene silencing in plants. Increasing the copy number increased the probability of silencing which was affected by modifiers of position effect variegation, suggesting that heterochromatin is formed. A *white* transgene on the homologous chromosome was also silenced in the presence of the repetitive array showing that inactivation can be communicated by a mechanism capable of recognising sequence homology (Dörer and Henikoff, 1997).

Expression of *Alcohol dehydrogenase* (*Adh*) from a transgene containing the promoter and regulatory sequences of the *white* gene is reduced as the number of transgene copies increases (Pal-Bhadra *et al.*, 1997). Multiple transgene copies also cause a reduction in the expression of the endogenous *Adh* gene which recalls the phenomenon of co-suppression in plant systems. Silencing is reduced by mutations in two Polycomb group (Pc-G) genes, *Polycomb* and *Polycomb-like*, and silenced transgene copies recruit *Polycomb* group proteins to their sites of insertion (Pal-Bhadra *et al.*, 1997). Pc-G genes function to maintain a transcriptionally repressed state at homeotic genes by interacting with cis regulatory elements known as Polycomb response elements (PREs, Chan *et al.*, 1994, Kassis, 1994, Gindhard and Kaufman, 1995). Two known features of the behaviour of PREs are of interest in this context. Firstly PREs communicate with each other so that silencing is enhanced when more than one copy of a PRE is present (Chan *et al.*, 1994, Kassis, 1994, Gindhard and Kaufman, 1995, Sigrist and Pirrotta, 1997). If Pc-G genes are involved in *I* factor regulation this could explain how the copy number in the genome could influence expression at each individual copy by a process similar to the trans-silencing mechanism observed for *P* element regulation (Roche *et al.*, 1998). Secondly the epigenetic state of a PRE, whether repressive or permissive for transcription, can be inherited through both mitotic and, as has recently been shown, meiotic cell divisions. A study of the *Fab-7* PRE of the bithorax complex found epigenetic inheritance through meiosis in females but not in males (Cavalli and Paro,
If a repressed state of \( I \) factor copies is subject to a similar mode of inheritance this could explain why transposition is greater in \( SF \) than in \( RSF \) females as the inducer mothers of \( RSF \) flies would be able to pass on a silenced state to their daughters.

Jensen et al. (1999) have carried out further studies on copy number-dependent regulation of the \( I \) factor. They found that introducing several copies of a transgene, containing a 969bp fragment from the 5' region of ORF2 downstream of the \( hsp70 \) promoter, resulted in repression of hybrid dysgenesis, even when the sequence was mutated to introduce stop codons in all three open reading frames. This supports the conclusion of Chaboissier et al., that \( I \) factor proteins are not required for repression and argues against the transcription factor titration hypothesis as this region is not required for germ line specific transcription (McLean et al., 1993). A transgene containing ORF1 was also capable of regulating hybrid dysgenesis when present in multiple copies, demonstrating that specific \( I \) factor sequences are not essential for repression and supporting the view that repression is mediated by the recognition of sequence homology. Experiments involving reciprocal crosses showed that the repressed state induced by a multicopy transgene is inherited maternally (Jensen et al., 1999). This could account for the difference in fertility and transposition rate between \( SF \) and \( RSF \) females.

A transgene containing both the sequences of ORF1 and the 969bp fragment of ORF2 was more effective than the other two transgenes at inducing silencing (Jensen et al., 1999). A version of this construct in which the \( hsp70 \) promoter was mutated had no silencing activity, implying that the production of RNA is necessary. This might suggest a post-transcriptional silencing mechanism, possibly with similarities to PTGS in plants. Alternatively it is possible that the RNA provides a trigger for the establishment of silencing at the transcriptional level, perhaps by a process related to RNA-mediated DNA methylation in higher plants (Wassenegger et al., 1994, Mette et al., 1999, Pelissier et al., 1999). In this respect there is an apparent disagreement with Chaboissier et al. (1998) who favoured a transcriptional silencing mechanism to explain the effect observed with multiple copies of the 5' UTR.
1.3.5 Scope of the thesis

A complete understanding of the mechanisms of I factor regulation will require characterisation of the protein factors that bind to the enhancer. This thesis describes the identification of a protein present in *Drosophila* ovaries that binds specifically to site 1, and attempts to characterise its role in I factor transcriptional regulation.
Chapter 2 – Materials and Methods
Materials and Methods

2.1 Materials

2.1.1 Media

2.1.1.1 Bacterial media

Luria Broth (LB)

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

Luria Agar (L-agar)

Luria broth with 15g/l Bacto agar (Difco). Ampicillin (Penbritin, Beecham Research) was added to LB and L-agar to a final concentration of 100μg/ml where indicated.

SOC Buffer

LB with 3.6 g/l glucose, 0.1M MgSO4 and 0.1M MgCl2

Top Agarose

LB with 0.6% agarose and 1mM MgSO4

2x TY Broth

Bacto tryptone (Difco), 16g; Bacto yeast extract (Difco), 10g; NaCl, 10g; per litre adjusted to pH 7.4

2.1.1.2 Drosophila media

‘French’ Fly Food

Oxoid No.3 agar, 7.5g; polenta, 55; dried flake yeast, 550g; nipagen (150mg/ml made up in 95% ethanol), 10ml; dH2O, 100ml
Apple juice agar

Bacto agar (Difco), 2.5g per 100ml pure apple juice

2.1.1.3 Tissue culture medium

Schneider's Drosophila medium (Gibco) with 5% foetal calf serum (sera-lab), 100μg/ml ampicillin and 100μg/ml streptomycin (Gibco).

2.1.2 Materials

2.1.2.1 Chemicals

Chemicals were supplied by Sigma, Fisons, BDH and Boehringer Mannheim.

2.1.2.2 Solutions

TE

10mM Tris; 50mM EDTA; adjusted to pH 8

Nuclease mix

10mg/ml DNaseI with 10mg/ml RNAse A

Phage buffer

50mM Na₂HPO₄; 22mM KH₂PO₄; 86mM NaCl; 1mM MgSO₄; 0.1mM CaCl₂; 0.001% gelatin

PEG solution

Polyethylene Glycol, MW 8,000, 20% w/v; 2M NaCl in phage buffer

Kinase/ligase buffer

50mM Tris.Cl (pH7.6); 100mM MgCl₂

4x Agarose gel loading buffer

20% glycerol (v/v); 0.05% bromophenol blue in TE
**TAE**

40mM Tris-acetate; 1mM EDTA

**TBE**

45mM Tris-borate; 1mM EDTA

1 x DNA polymerase buffer

10mM Tris.Cl; 5mM MgCl₂; 7.5mM DTT; pH 7.5

1 x T4 ligase buffer

50mM Tris.Cl (pH 7.8); 10mM MgCl₂; 10mM dithiothreitol, 1mM ATP; 50μg/ml bovine serum albumin;

**Denaturation solution**

0.5M NaOH; 1.5M NaCl

**Neutralisation buffer**

3M NaCl; 1M Tris.Cl pH7.5

**20 x SSC**

3M NaCl; 300mM tri-Na citrate

**Prehybridisation solution**

500mM NaPO₄ (Molar ratio Na₂HPO₄: NaH₂PO₄ 18:7); 7% (w/v) SDS; 1mM EDTA; 100μg/ml denatured calf thymus DNA (Sigma)

**PCR buffers**

For vent DNA polymerase: 10mM KCl; 20mM Tris.Cl (pH8.8 at 25°C); 10mM (NH₄)₂SO₄; 2mM MgSO₄; 0.1% Triton X-100

For Taq DNA polymerase: 50mM KCl; 10mM Tris.Cl (pH 9 at 25°C); 0.1% Triton X-100

**Drosophila DNA extraction buffer**

100mM Tris.Cl, pH9; 100mM EDTA; 1% (w/v) SDS
Injection buffer

5mM KCl; 100μM NaPO₄ pH 6.8

PBS

137mM NaCl; 2.68mM KCl; 10mM Na₂HPO₄; 1.76mM KH₂PO₄ pH 7.4

Nuclear extract buffer A

10mM HEPES pH 7.9; 10mM KCl; 1.5mM MgCl₂; 0.5mM DTT; 0.5mM PMSF

Nuclear extract buffer B

10mM HEPES pH 7.9; 100mM NaCl; 1.5mM MgCl₂; 0.5mM DTT; 0.5mM PMSF; 5% glycerol

HEPES binding buffer

25mM NaCl; 5mM MgCl₂; 0.5mM DTT; 25mM HEPES pH 7.0

Block solution

5% Marvel non fat dried milk in HEPES binding buffer.

2.1.2.3 Enzymes

Restriction enzymes, vent polymerase, Klenow polymerase and T4 DNA ligase were purchased from New England Biolabs. Taq polymerase was supplied by Promega, RNase and DNaseI by Sigma and T4 polynucleotide kinase by Boehringer Mannheim.

2.1.2.4 Isotopes

α-³²P-dCTP (3000Ci/mM) and α-³⁵S-dATP (400Ci/mM) were supplied by Amersham.
### 2.1.2.5 Plasmids

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Reference/Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBluescript II</td>
<td>General cloning vector.</td>
<td>Stratagene</td>
</tr>
<tr>
<td>pGEM-T</td>
<td>TA cloning vector, for cloning PCR products.</td>
<td>Promega</td>
</tr>
<tr>
<td>pr25.7wc</td>
<td>Encodes P transposase. Helper plasmid for P element-mediated transformation</td>
<td>Karess and Rubin, 1984</td>
</tr>
<tr>
<td>pW8</td>
<td>Vector for P element-mediated transformation. Contains the white gene under</td>
<td>Klemenz et al., 1987</td>
</tr>
<tr>
<td></td>
<td>the control of the hsp70 promoter as a marker for selection.</td>
<td></td>
</tr>
<tr>
<td>pCAT.1</td>
<td>Contains the CAT gene linked to the SV40 small t intron and polyadenylation</td>
<td>McLean et al., 1993.</td>
</tr>
<tr>
<td></td>
<td>signals for expression in eukaryotes.</td>
<td></td>
</tr>
<tr>
<td>p186T.1</td>
<td>Contains the 5' UTR of the I factor inserted into pCAT1 upstream of the CAT</td>
<td>McLean et al., 1993</td>
</tr>
<tr>
<td></td>
<td>open reading frame</td>
<td></td>
</tr>
<tr>
<td>pW8.186.CAT</td>
<td>Contains the I factor 5'UTR upstream of the CAT reporter of pCAT.1 inserted</td>
<td>McLean et al., 1993</td>
</tr>
<tr>
<td></td>
<td>into pW8 for transformation of Drosophila</td>
<td></td>
</tr>
<tr>
<td>p186Δ.Ti</td>
<td>As p186T.1 but with site 1 deleted in the I factor 5'UTR</td>
<td>Udomkit et al., 1996</td>
</tr>
<tr>
<td>pBSφ35</td>
<td>Contains AEF-1 open reading frame downstream of codons 408-439 of the</td>
<td>Falb and Maniatis, 1992a</td>
</tr>
<tr>
<td></td>
<td>human c-myc gene.</td>
<td></td>
</tr>
<tr>
<td>pBBS</td>
<td>Contains two copies of the B germline regulatory element of the hsp26 gene</td>
<td>Frank et al., 1992</td>
</tr>
<tr>
<td></td>
<td>upstream of the SGS3 promoter</td>
<td></td>
</tr>
<tr>
<td>pSV2βglobin</td>
<td>Contains SV40 small t intron and polyadenylation sequence</td>
<td>Derived from pSV2, Gorman</td>
</tr>
<tr>
<td></td>
<td></td>
<td>et al., 1982</td>
</tr>
</tbody>
</table>
### 2.1.2.6 Bacterial strains

<table>
<thead>
<tr>
<th>Name</th>
<th>Genotype and use</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH5α</td>
<td><code>deoR, endA1, gyrA96, hsdR17 (rKmK⁺), supE44, thi-1, recA1, relA, Δ(lacZYA-argF)U169, deoR(p808lacZΔM15), F, λ</code>&lt;br&gt;Used for plating λZAP.</td>
<td>Hanahan, 1983</td>
</tr>
<tr>
<td>XL1-blue</td>
<td><code>Δ(mcrCB-hsdSMR-mrr), lac [F' proAB, lacIqZ ΔM15] Su-. Used for plating λZAP.</code>&lt;br&gt;Used for expression of recombinant proteins.</td>
<td>Bullock et al., 1987</td>
</tr>
<tr>
<td>BL21(DE3)</td>
<td><code>F'ompT [lon] hsdS6(rB-mB⁺) has DE3, a λ prophage carrying the T7 RNA polymerase gene. Used for expression of recombinant proteins.</code>&lt;br&gt;Used for lysogeny of λgt22a phage containing cDNA inserts</td>
<td>Studier et al., 1990</td>
</tr>
<tr>
<td>Y1089</td>
<td><code>F' Δ(lac)U169, lon-100, araD139, strA, hflA 150::Tn10, (pMC9; Tetr'Amp') Used for lysogeny of λgt22a phage containing cDNA inserts</code>&lt;br&gt;Used for plating λgt22a phage containing cDNA inserts</td>
<td>Huynh et al., 1985</td>
</tr>
<tr>
<td>Y1090</td>
<td><code>F' Δ(lac)U169, lon-100, araD139, rpsL(Str'), supF, mcrA, trpC22::Tn10, (pMC9; Tetr'Amp') Used for plating λgt22a phage containing cDNA inserts</code>&lt;br&gt;Used for plating λgt22a phage containing cDNA inserts</td>
<td>Huynh et al., 1985</td>
</tr>
</tbody>
</table>

### 2.1.2.7 Phage

Exassist™ (Stratagene) was used for coinfection of *E. coli* with λZAP containing cDNA inserts to excise plasmids by site-specific recombination (reference - Stratagene predigested λZAP II/EcoRI/CIAP cloning kit instruction manual)

### 2.1.2.8 Phage Libraries

1. Ovarian cDNA expression library in λgt22a (Stroumbakis et al., 1994)
2. Ovarian cDNA expression library in λZAP (Jongens et al., 1992)

### 2.1.2.9 Drosophila melanogaster strains

<table>
<thead>
<tr>
<th>Name (genotype)</th>
<th>Source/reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charolles (wild type)</td>
<td>Pélisson, 1981</td>
</tr>
<tr>
<td>JA (yellow, white)</td>
<td>Marie-Christine Chaboissier</td>
</tr>
<tr>
<td>Cy/Pm: Dcxf/H</td>
<td>Marie-Christine Chaboissier</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>W (white)</td>
<td>Lüning, 1981</td>
</tr>
<tr>
<td>137 (W with P186.CAT insertion)</td>
<td>McLean et al., 1993</td>
</tr>
<tr>
<td>230 (W with multiple insertions of P186.ΔCAT)</td>
<td>Udomkit et al., 1996, referred to as line 9</td>
</tr>
<tr>
<td>231 (W with P186.ΔCAT insertion)</td>
<td>Udomkit et al., 1996 (referred to as line 11)</td>
</tr>
<tr>
<td>232 (W with P186.ΔCAT insertion)</td>
<td>Udomkit et al., 1996, (referred to as line 13)</td>
</tr>
<tr>
<td>233 (W with P186.ΔCAT insertion)</td>
<td>Udomkit et al., 1996, (referred to as line 14)</td>
</tr>
<tr>
<td>234 (W with P186.ΔCAT insertion)</td>
<td>Udomkit et al., 1996, (referred to as line 17)</td>
</tr>
</tbody>
</table>

### 2.1.2.10 Cell culture line

Schneider line 2 embryonic culture cells (Schneider, 1971).

### 2.1.2.11 Oligonucleotides

The table shows the sequences of oligonucleotides. Nucleotides shown in bold are from the I factor sequence. Underlined nucleotides are within recognition sites for restriction endonucleases.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1 5'-3'</td>
<td>GAT CCA AAA ACA ACA ATA CCG CTA</td>
<td>Preparation of site 1 probe for library screen and site 1 competitor for gel retardation assays</td>
</tr>
<tr>
<td>S1 3'-5'</td>
<td>GAT CCT AGC GGT ATT GTT GTT TTT G</td>
<td>Preparation of site 1 probe for library screen and site 1 competitor for gel retardation assays</td>
</tr>
<tr>
<td>S1mut 5'-3'</td>
<td>GAT CCA GAT GTT CAT CTA CAC CTA G</td>
<td>Preparation of negative control probe for filter binding assay</td>
</tr>
<tr>
<td>S1mut 3'-5</td>
<td>GAT CCT AGG TGT AGA TGA ACA TCT G</td>
<td>Preparation of negative control probe for filter binding assay</td>
</tr>
<tr>
<td>664C</td>
<td>GTC TAG ACA TTA CCA CTT CAA CCT CCG</td>
<td>+ strand primer annealing at the 5' end of the I factor</td>
</tr>
<tr>
<td>Y3411</td>
<td>CCG CTC GAG GAT TGT TGG TTA AGG GC</td>
<td>- strand primer annealing at the 3' end of the I factor 5'UTR</td>
</tr>
<tr>
<td>CAT-R</td>
<td>TGT TCT TTA CGA TGC CAT TGG</td>
<td>- strand primer annealing in the CAT open reading frame</td>
</tr>
<tr>
<td>R135.136</td>
<td>CTT AAD BAC AA AAC AAC AAT ACC</td>
<td>+ strand primer for mutagenesis of C135 and A136</td>
</tr>
<tr>
<td></td>
<td>GCT AAT CC</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>-----------</td>
<td>-----------------------------------------------------------------</td>
</tr>
<tr>
<td>L135.136</td>
<td>GGT ATT GTT GTT TTT GTV HTT AAG TTT TTT ATT TTC TC</td>
<td>- strand primer for mutagenesis of C135 and A136</td>
</tr>
<tr>
<td>R137.138</td>
<td>CTT AAC ABD AAA AAC AAC AAT ACC GC</td>
<td>+ strand primer for mutagenesis of A137 and C138</td>
</tr>
<tr>
<td>L137.138</td>
<td>GGT ATT GTT GTT TTT HVT GTT AAG TTT TTT ATT TTC TC</td>
<td>- strand primer for mutagenesis of A137 and C138</td>
</tr>
<tr>
<td>R143.144</td>
<td>CTT AAC AAC AAA ABD AAC AAT ACC GCT AAT CC</td>
<td>+ strand primer for mutagenesis of A143 and C144</td>
</tr>
<tr>
<td>L143.144</td>
<td>GGT ATT GTT HVT TTT GTT GTT AAG TTT TTT ATT TTC TC</td>
<td>- strand primer for mutagenesis of A143 and C144</td>
</tr>
<tr>
<td>R150.151</td>
<td>CTT AAC AAC AAA AAC AAC AAV BCC GCT AAT CC</td>
<td>+ strand primer for mutagenesis of T160 and A161</td>
</tr>
<tr>
<td>L150.151</td>
<td>GGV BTT GTT GTT TTT GTT GTT AAG TTT TTT ATT TTC TC</td>
<td>- strand primer for mutagenesis of T160 and A161</td>
</tr>
<tr>
<td>R137.138.143.144</td>
<td>CTT AAC ABD AAA ABD AAC AAT ACC GCT AAT CC</td>
<td>+ strand primer for mutagenesis of A137, C138, A143 and C144</td>
</tr>
<tr>
<td>L137.138.143.144</td>
<td>GGT ATT GTT HVT TTT HVT GTT AAG TTT TTT ATT TTC TC</td>
<td>- strand primer for mutagenesis of A137, C138, A143 and C144</td>
</tr>
<tr>
<td>5' opa</td>
<td>ACC CAG GCC ACG CAT CCG GCT CAC</td>
<td>+ strand primer for amplification across the AEF-1 opa repeat</td>
</tr>
<tr>
<td>3' opa</td>
<td>GGC GGC GGC ACA GAG GGT GTT CCG</td>
<td>- strand primer for amplification across the AEF-1 opa repeat</td>
</tr>
<tr>
<td>T7</td>
<td>GTA ATA CGA CTC ACT ATA GGG C</td>
<td>Primer for sequencing from pBluescript</td>
</tr>
<tr>
<td>T3</td>
<td>AAT TAA CCC TCA CTA AA GGG</td>
<td>Primer for sequencing from pBluescript</td>
</tr>
<tr>
<td>SGS3</td>
<td>TTA GAA TTG CAG TTT CAT TG</td>
<td>Primer for sequencing from SGS3 promoter in the direction of transcription</td>
</tr>
<tr>
<td>CATb</td>
<td>TCA TCA GCC GGG CAA GAA TGT G</td>
<td>- strand primer annealing in the CAT open reading frame</td>
</tr>
</tbody>
</table>

40
2.2 Methods

2.2.1 Manipulation of bacteria

2.2.1.1 Growth of E. coli cultures

*E. coli* cultures were grown by inoculation of bacteria from a single colony into LB or 2 x TY broth and incubation for 14-16 hours at 37°C with aeration by vigorous shaking. For strains carrying ampicillin-resistant plasmids and for Y1089 and Y1090, LB was supplemented with ampicillin.

2.2.1.2 Storage of E. coli cultures

For long term storage *E. coli* cultures in logarithmic phase growth were mixed with an equal volume of glycerol, placed in sterile tubes and kept at -70°C. To grow bacteria from frozen culture a small portion was removed using a sterile loop and streaked on an L-agar plate, with ampicillin if required.

For short term storage up to six weeks bacteria were streaked onto agar plates which were incubated at 37°C 14-16 hours for colony growth then kept at 4°C.

2.2.1.3 Transformation of bacteria

Transformation of *E. coli* by purified plasmid DNA or ligation products was carried out by electroporation according to Heery and Duncan (1989). Cultures were grown to early stationary phase in 2 x TY medium and cells harvested by centrifugation at 9000 x G for 10 minutes at 4°C. Cells were washed by resuspension in a volume of ice-cold dH2O equal to that of the original culture and again collected by centrifugation. This wash was repeated twice more and the cells resuspended in an equal volume of ice-cold dH2O. 40μl of cells were mixed with 1μl DNA solution and transferred to an electroporation cuvette (0.2cm, Invitrogen). A single pulse at 2.5kV, 15μF, 200Ω was applied. 1ml SOC buffer was added immediately and the mixture transferred to a culture tube. Cells were then incubated at 37°C with shaking for 20 minutes. Several dilutions in SOC buffer were made and plated on L-agar with ampicillin. For selection for inactivation of β galactosidase
plated on L-agar with ampicillin. For selection for inactivation of β galactosidase expression 100μl of 100mM IPTG and 20μl 50mg/ml X-gal were spread onto the plates which were then incubated for 30 minutes at 37°C for absorption prior to use.

2.2.2 Manipulation of lambda phage

2.2.2.1 Phage growth

_E. coli_ Y1090 cells were used for plating λgt22a phage while XL1-blue was used for λZAP. Cells were grown overnight from inoculation of a single colony at 37°C in 50ml LB with 0.2% maltose, 0.1mM MgSO₄ and, for Y1090 cultures only, 100μg/ml ampicillin. Cells were collected by centrifugation, resuspended in 20ml 0.1M MgSO₄ and incubated at 37°C for 60 minutes before storage at 4°C. λ phage were mixed with plating cells and incubated at 37°C for 20 minutes to allow adsorption. Molten top agarose was added before pouring onto L-agar plates (containing ampicillin where Y1090 plating cells were used). Plates were left at room temperature for 10 minutes to allow the agarose to set before incubation at 37°C.

2.2.2.2 Preparation of phage lysates

Lambda phage was amplified for storage and DNA purification using the plate lysate method. Phage were plated at a density of 1 x 10⁵ plaque-forming units per plate on 75mm diameter plates and incubated until phage growth was confluent. 3ml of phage buffer was poured onto the plates and the top agarose scraped off into centrifuge tubes. After incubation for 30 minutes at room temperature with occasional shaking the agarose and cell debris was pelleted by centrifugation at 8,000 x G for 10 minutes at 4°C. The supernatant was transferred to a new tube and chloroform added to a final concentration of 0.3% v/v. Phage lysates were stored at 4°C.

2.2.2.3 Phage lambda DNA purification

To purify DNA from phage lysates, 5μl nuclease mix was added to 5ml lysate (see Section 2.2.2.2) and the mixture incubated 30 minutes at 37°C. 4ml PEG
solution was added and the mixture stored for 16 hours at 4°C for phage precipitation. Phage were collected by centrifugation at 10,000 x G for 20 minutes at 4°C. Following removal of excess liquid the pellet was resuspended in 500μl phage buffer and centrifuged for 1 minute in a microcentrifuge to remove any insoluble material. The supernatant was extracted successively with 500μl chloroform, 500μl phenol, 500μl of a 50% phenol/50% chloroform mixture and 500μl chloroform. DNA was then precipitated by the addition of 800μl absolute ethanol and incubation for 10 minutes on ice. DNA was recovered by centrifugation for 10 minutes at full speed in a microcentrifuge, washed with 70% ethanol and resuspended in 450μl TE buffer. The DNA was again precipitated by ethanol according to Section 2.2.3.4 and finally resuspended in 100μl TE.

2.2.2.4 Lysogeny of phage lambda

Lysogens were obtained from λgt22a clones by spotting on a lawn of E. coli Y1089 and growth at 30°C. Cells from the interior of turbid plaques were streaked to single colonies then streaked on duplicate plates. To confirm lysogeny one plate was incubated at 30°C and the other at 42°C. Clones showing growth on the former plate but not the latter were taken to be lysogens.

2.2.3 In vitro manipulation of DNA

2.2.3.1 Small scale preparation of plasmid DNA

Small scale preparation of plasmid DNA from E. coli cultures was carried out using the Wizard® Plus SV minipreps DNA purification system (Promega) according to the manufacturer's instructions. This method involves alkaline lysis of bacteria followed by a brief treatment with alkaline protease to inactivate endonucleases released on cell lysis. Plasmid DNA is then purified by binding to a column, washing in a 60% ethanol solution to remove impurities and finally elution in dH2O.

2.2.3.2 Large scale preparation of plasmid DNA

Preparation of up to 100μg of plasmid DNA from E. coli cultures was carried out using the Qiagen plasmid midi kit (Qiagen GmbH and Qiagen Inc) according to
the manufacturer’s directions. This method is similar to the miniprep method described above (Section 2.2.3.1). Alkaline lysis of *E. coli* is followed by binding of plasmid DNA to an anion exchange resin under low salt and pH conditions. The resin is washed in a medium salt buffer, and the DNA eluted by high salt. Finally the DNA is concentrated by isopropanol precipitation.

### 2.2.3.3 Removal of protein from DNA by Phenol/chloroform extraction

Water-saturated distilled phenol (Rathburn chemicals), containing 0.1% hydroxyquinolone, was mixed with an equal volume of 0.5M Tris.Cl pH8 and the phases allowed to separate. This step was repeated until the pH of the organic phase was 8. Equilibrated phenol was stored in the presence of 100mM Tris.Cl pH8, containing 0.2% β mercaptoethanol. Prior to use, equilibrated phenol was mixed with an equal volume of chloroform. DNA to be extracted was added to an equal volume of this phenol/chloroform mixture and mixed thoroughly. The phases were separated by centrifugation in a microcentrifuge for 5 minutes. The aqueous phase was removed and extracted with an equal volume of chloroform to remove any residual phenol. After separation of the phases by centrifugation the DNA solution was removed to a new tube.

### 2.2.3.4 Precipitation of DNA using ethanol

DNA in solution was precipitated by the addition of 1/9 volume 3M sodium acetate pH 5.2 followed by 3 volumes absolute ethanol. After mixing, the solution was incubated 20 minutes on ice and DNA recovered by centrifugation at 13,000 rpm for 10 minutes in a Biofuge 13 microcentrifuge (Heraeus). Following removal of the supernatant the pellet was washed with 70% ethanol and dried for 10 minutes at room temperature. DNA was dissolved in dH₂O or TE.

### 2.2.3.5 Quantification of DNA

DNA concentrations were estimated by measurement of absorption at 260nm using a lambda 15 UV/VIS spectrophotometer (Perkin Elmer). Absorption measurements were converted to DNA concentrations using an extinction coefficient of 50μg/ml for double-stranded DNA and 33μg/ml for single-stranded DNA.
2.2.3.6 Cleavage of DNA by restriction endonucleases

DNA cleavage was carried out using enzymes and buffers supplied by Boehringer Mannheim and New England Biolabs under the conditions recommended by the manufacturers. Digests of 0.1 to 20μg DNA were carried out in 20-100μl of the appropriate 1x reaction buffer for 1-12 hours at 37°C.

2.2.3.7 Preparation of concatenated double-stranded oligonucleotides

Concatenated DNA for use as probe in cDNA expression screening (Section 2.2.8.1) was prepared by the method described by Sambrook et al. (1989). The oligonucleotides S1 5'-3' and S1 3'-5', when annealed to each other, form a double-stranded molecule containing site 1 and at each end having a 5' overhang identical to that created by cleavage by BamHI at its target site. 2μg of each were phosphorylated separately by incubation with 1mM ATP and 16 units polynucleotide kinase (both Boehringer Manheim) in 50μl kinase/ligase buffer for 30 minutes at 37°C. The two reactions were mixed and the oligonucleotides annealed by incubation at 85°C for 2 minutes then at 65°C for 15 minutes, at 37°C for 15 minutes, at room temperature for 15 minutes and on ice for 15 minutes. Fresh ATP was added to 1mM and 4 units T7 DNA ligase (New England Biolabs) were added. Ligation was carried out at 20-25°C for 3 hours. After extraction with phenol/chloroform and ethanol precipitation, DNA was resuspended in dH2O, quantified, and labelled by nick translation (see Section 2.2.3.16).

To make a probe lacking the site 1 sequence to test recombinant proteins for DNA binding specificity, oligonucleotides S1 mut 5'-3' and S1 mut 3'-5' were processed as above.

2.2.3.8 Agarose gel electrophoresis

Electrophoresis of DNA was carried out in 0.7-2% MP agarose (Boehringer Mannheim) in TAE or TBE containing 0.5mg/ml ethidium bromide. TBE was used routinely while TAE was used for preparative gels (see Section 2.2.3.9). Prior to loading, DNA samples were mixed with 1/4 volume 4 x agarose gel loading buffer. A potential difference of 1-10V per cm gel was used to separate DNA fragments.
Following electrophoresis DNA was visualised and photographed on a UV transiluminator.

2.2.3.9 Purification of DNA fragments from agarose

Gel slices containing DNA fragments separated by agarose gel electrophoresis were purified using the QIAquick gel extraction kit (Qiagen) according to the manufacturer's instructions. This involves binding of DNA to a silica gel membrane at low pH in the presence of chaotropic salt, followed by washing in a buffer containing ethanol and low salt elution in 10mM Tris.Cl pH8.5.

2.2.3.10 Creation of blunt termini from 3' recessed ends

E. coli DNA polymerase I large fragment (Klenow) was used to fill in recessed DNA 3' termini. DNA at a concentration of 50µg/ml was incubated with 33µg/ml of each of the four dNTPs (Boehringer Mannheim) and one unit of enzyme (New England Biolabs) in 1 x DNA polymerase buffer for 15 minutes at 25°C.

2.2.3.11 Ligation of DNA fragments

For most ligations 50-200ng of linear vector DNA was mixed with an equimolar quantity of the fragment to be inserted and 1 unit of T4 DNA ligase (New England biolabs) in a total volume of 10µl 1 x T4 ligase buffer. The mixture was incubated for 12-16 hours at 16°C. Ligation of PCR products into pGEM®-T was carried out according to the instructions of the manufacturer of the pGEM®-T vector system cloning kit (Promega).

2.2.3.12 Colony and plaque lifts

Following ligation and transformation, colonies containing plasmids with the desired insertions were identified by colony lift hybridisation. A similar procedure was employed for the isolation of lambda phage plaques containing a cDNA insertion of interest. Plates cooled to 4°C were overlayed with Hybond-N filters (Amersham) for 1 minute. The filters were layed, DNA side up, on blotting paper soaked in denaturation solution for 2 minutes then on blotting paper soaked in neutralisation.
buffer for 2 minutes. Filters were rinsed briefly in 2 x SSC before being cross-linked by UV (Section 2.2.3.14).

2.2.3.13 Southern blotting

Following agarose gel electrophoresis, DNA was analysed by hybridisation according to the method of Southern (1975). Gels were soaked first in denaturing solution then in neutralisation buffer, each for 30 minutes with gentle agitation. Following each step gels were rinsed in dH2O. DNA was transferred to hybond-N nylon filters (Amersham) by capillary blotting. A raised platform was placed in a tank and covered with a blotting paper wick, saturated in 20 x SSC. The ends of the wick extended on opposite sides of the platform down into the tank which contained 20 x SSC. The gel was placed on top of the wick on the platform and covered with a sheet of hybond N. This was overlayed with three sheets of blotting paper saturated in 20 x SSC, a stack of dry paper towels and a weight, mounted on a glass plate. This was left for 16 hours for transfer before the membrane was removed and cross-linked with UV prior to hybridisation.

2.2.3.14 UV cross-linking

Filters for hybridisation were fixed by UV cross-linking in a Stratalinker (Stratagene) according to the instructions of the manufacturers.

2.2.3.15 Random primed labelling

Probes for hybridisation were prepared by random-primed labelling using the T7 quickprime kit (Pharmacia) according to the instructions of the manufacturer. 10-100ng of linear DNA were denatured by boiling then cooled and added to a reagent mix containing random oligonucleotides, buffer and dNTPs. 50μCi (α-32P) dCTP and T7 DNA polymerase were added and the mixture incubated for 2 hours at 37°C.

2.2.3.16 Nick translation

Probes for phage λ cDNA expression screening were labelled by incorporation of (α-32P) dCTP using the Promega nick translation system according to the instructions of the manufacturer.
2.2.3.17 End-labelling of DNA fragments using Klenow polymerase

DNA fragments for use as probes in gel retardation were labelled by incorporation of radioactive dCTP during the filling in of recessed 3' ends by Klenow polymerase. 200ng of the DNA fragment were mixed with 2mM each dGTP, dATP and dTTP (all from Boehringer Mannheim), 20μCi (α-32P) dCTP and 5 units of Klenow polymerase (New England Biolabs) in 1 x DNA polymerase buffer. Following a 30 minute incubation at 20-25°C 5μl 5mM dCTP (Boehringer Mannheim) was added. After a further 30 minute incubation (20-25°C) the reaction was stopped by freezing at -20°C.

2.2.3.18 Removal of unincorporated nucleotide following labelling

Following radioactive labelling, probes were passed through Nuctrap® probe purification columns (Stratagene) which remove unincorporated nucleotide by gel filtration.

2.2.3.19 Hybridisation

Hybond-N filters with bound DNA were incubated at 65°C in 15ml prehybridisation solution in Techne hybridisation cylinders for 1-2 hours. Labelled probe was denatured by boiling and added to the prehybridisation solution. Hybridisation was for 3-16 hours at 65°C. The solution was removed and the filters washed once for 15 minutes in 1xSSC, 0.1% SDS then for 15 minutes more in 0.1xSSC, 0.1%SDS, still at 65°C. Filters were removed, dried at room temperature and sealed in plastic bags prior to autoradiography.

2.2.3.20 Autoradiography

Radioactive gels and filters were exposed to X-ray film (HA West) in autoradiographic cassettes for an appropriate time before developing in an X-OGRAPH compact X-2 automatic developer.

2.2.3.21 Sequencing of double-stranded plasmid DNA

DNA sequencing was carried out by the method of Sanger et al. (1977). The T7 Sequenase™ version 2.0 kit (United States Biochemicals) was used. 3-5μg
plasmid DNA was denatured by incubation in 0.2M NaOH, 0.2mM EDTA for 30 minutes at 37°C. Following ethanol precipitation the DNA was resuspended in 7μl dH2O. Annealing of primer to template was carried out by incubation at 37°C for 30 minutes in the reaction buffer provided in the kit. Labelling and termination reactions were as recommended by the manufacturer and used (α-35S)-dATP.

Sequencing reaction products were denatured by boiling prior to separation on 6% polyacrylamide denaturing gels. The Sequi-Gen™ apparatus (Biorad) was used. Electrophoresis was in TBE at 40W for 2-5 hours. Gels were soaked for 5 minutes following electrophoresis in 5% acetic acid, 10% methanol then dried and autoradiographed.

2.2.3.22 Polymerase chain reaction

PCRs were carried out in 100μl of the appropriate PCR buffer with 0.5μM each primer, 0.2mM each dNTP (Boehringer Mannheim), and 1 unit VENT® DNA polymerase (New England Biolabs) or 1 unit Taq DNA polymerase (Promega). Vent polymerase was used for the production of competitor DNA for use in gel retardation assays. Taq was used for all other PCRs. Reactions were incubated at 94°C for 10 minutes followed by 30 cycles of 1 minute at 94°C, 1 minute at the annealing temperature and 1 minute at 72°C. Annealing temperatures differed for the different reactions (see below). Finally reactions were incubated at 72°C for 10 minutes.

For PCR overlap extension mutagenesis the initial PCR using degenerate primers was carried out with an annealing temperature of 57°C. The second PCR using 664C and CAT.R used an annealing temperature of 51°C. Both wild-type and mutant 5'UTR DNA for use as competitor in gel retardation assays was generated from plasmid templates using primers 664C and Y3411 and an annealing temperature of 41°C. Amplification from genomic DNA using the primers 3'opa and 5'opa was carried out with an annealing temperature of 70°C. Amplification from genomic DNA using primers 664C and CATb was carried out with an annealing temperature of 55°C.
2.2.3.23 Cloning of PCR products

PCR products were cloned using the pGEM®-T vector system I kit (Promega). This utilises a pre-cut plasmid vector (pGEM®-T) having a single unpaired deoxythymidine nucleotide at each 3' end. This provides compatible overhangs for ligation to PCR products as thermostable polymerases add an unpaired deoxyadenosine to the 5' end during synthesis (Clark, 1988). Following ligation and transformation, colonies with plasmids containing insertions are detected by blue-white selection (see Section 2.2.1.3).

2.2.5 Manipulation of Drosophila melanogaster flies, cells and tissues

2.2.5.1 Cell culture

Schneider line 2 cells were maintained in Schneider's Drosophila medium with foetal calf serum and antibiotics at a density of 2-4x10^6 cells per ml at 20°C.

2.2.5.2 Maintenance of Drosophila stocks

Drosophila melanogaster strains were maintained at 25°C on French fly food. To maintain the reactivity of stocks, only flies up to seven days old were used for breeding.

2.2.5.3 Preparation of Drosophila genomic DNA

20 flies were frozen for 5 minutes at -70°C then resuspended in 400μl Drosophila DNA extraction buffer. The flies were homogenised using a hand-held Pellet-pestle® motor homogeniser (Kontes). Following incubation at 70°C for 30 minutes, 56μl 8M potassium acetate was added. Samples were incubated 30 minutes on ice. To remove insoluble material, samples were centrifuged at 4°C for 15 minutes at full speed in a microcentrifuge. The supernatant was removed and the centrifugation repeated. The final supernatant was added to 200μl isopropanol and cooled at -70°C for 10 minutes for precipitation of DNA. DNA was recovered by centrifugation, washed with 70% ethanol, dried and resuspended in 40μl TE.
2.2.5.4 *P* element-mediated germline transformation (Rubin and Spradling, 1982)

Several hundred 4-7 day old *JA* flies were allowed to lay eggs for 30-45 minutes at 25°C on plates of apple juice agar spread with yeast paste. Eggs were washed in dH2O and incubated in 50% w/v sodium hypochlorite for 5 minutes to dechorionate. Embryos were washed in dH2O, lined up and stuck onto a coverslip using double-sided sticky tape. Embryos were dessicated for 1-6 minutes in a sealed container containing dried silica gel before being covered with halocarbon oil (KMZ chemicals). DNA for injection was prepared by the Qiagen method (Section 2.2.3.2) and resuspended in injection buffer. A mixture containing 300-500µg/ml of the plasmid to be used in transformation and 100µg/ml of pπ25.7wc which provides transposase was prepared. This was injected into the posterior pole of preblastoderm embryos using a transjector 5246 (Eppendorf). Older embryos were killed by tearing the vitteline membrane. Injected embryos were incubated at 18°C for 36-60 hours then larvae transferred to french fly food for development into adults. Adults were mated with three *JA* virgins of the opposite sex and the progeny examined for flies with coloured eyes. Transformants were mated to three *JA* virgins of the opposite sex to obtain several flies heterozygous for the same chromosome carrying an insertion. Siblings of the progeny were mated to each other to establish homozygous lines.

2.2.5.5 Chromosome mapping of *P* element insertions in *Drosophila* lines

The chromosome on which *white* marked transgenes were located was determined by mating to the balancer strain *Cy/Pl; Dcx/H* followed by crossing to *JA* as follows:
X chromosome:

\[ \begin{array}{c}
G0 & \sigma^{+} & Cy; & Dcxf & \times & \textit{?} \ w' y \ P(w^+); \texttt{?}^{+} \ w' y \ P(w^+) & \texttt{?}^{+} \\
& Y & Pl & H & \downarrow & \texttt{?}^{+} \ w' y \ P(w^+) & \texttt{?}^{+} \\
& & & & \downarrow & \texttt{?}^{+} \ w' y \ P(w^+) & \texttt{?}^{+} \\
\end{array} \]

\[ \begin{array}{c}
G1 & \sigma \ w' y \ P(w^+); & Cy; & Dcxf \texttt{or} & H & \times & \textit{?} \ w' y ; \texttt{?}^{+} \ w' y \ & \texttt{?}^{+} \\
& Y & \texttt{+} & \texttt{+} & \downarrow & \texttt{+} \ w' y & \texttt{+} \\
& \texttt{J}A & & & & \texttt{?}^{+} \ w' y & \texttt{+} \\
\end{array} \]

G2 Examine progeny - all white-eyed flies are male

Chromosome II:

\[ \begin{array}{c}
G0 & \sigma^{+} & Cy; & Dcxf & \times & \textit{?} \ w' y ; \ P(w^+); \texttt{?}^{+} \ w' y \ P(w^+) & \texttt{?}^{+} \\
& Y & Pl & H & \downarrow & \texttt{?}^{+} \ w' y \ P(w^+) & \texttt{?}^{+} \\
& & & & \downarrow & \texttt{?}^{+} \ w' y \ P(w^+) & \texttt{?}^{+} \\
\end{array} \]

\[ \begin{array}{c}
G1 & \sigma \ w' y ; & Cy; & Dcxf \texttt{or} & H & \times & \textit{?} \ w' y ; \texttt{?}^{+} \ w' y \ & \texttt{?}^{+} \\
& Y & \texttt{P(w^+)} & \texttt{+} & \downarrow & \texttt{+} \ w' y & \texttt{+} \\
& \texttt{J}A & & & & \texttt{?}^{+} \ w' y & \texttt{+} \\
\end{array} \]

G2 Examine progeny - no curly flies have red eyes

Chromosome III:

\[ \begin{array}{c}
G0 & \sigma^{+} & Cy; & Dcxf & \times & \textit{?} \ w' y ; \texttt{?}^{+} \ P(w^+) & \texttt{?}^{+} \\
& Y & Pl & H & \downarrow & \texttt{?}^{+} \ w' y \ P(w^+) & \texttt{?}^{+} \\
& & & & \downarrow & \texttt{?}^{+} \ w' y \ P(w^+) & \texttt{?}^{+} \\
\end{array} \]

\[ \begin{array}{c}
G1 & \sigma \ w' y ; & Cy; & Dcxf \texttt{or} & H & \times & \textit{?} \ w' y ; \texttt{?}^{+} \ w' y \ & \texttt{?}^{+} \\
& Y & \texttt{+} & \texttt{P(w^+)} & \downarrow & \texttt{+} \ w' y & \texttt{+} \\
& \texttt{J}A & & & & \texttt{?}^{+} \ w' y & \texttt{+} \\
\end{array} \]

G2 Examine progeny - No Dcxf or hairless flies have red eyes

\[ \begin{array}{c}
\text{2.2.5.6 Ovary dissection} \\
Ovaries were dissected from 3-5 day old flies in PBS at room temperature. \\
\end{array} \]
2.2.6 Protein extraction, purification and analysis

2.2.6.1 Preparation of protein extracts from *E. coli* lysogens

Extracts for gel retardation assays were prepared from *E. coli* lysogens by a method modified from that described by Singh, 1993. Lysogen cultures were grown at 32°C in LB with ampicillin to an optical density of 0.5 at 595nm. They were then incubated at 44°C for 20 minutes to induce phage replication before IPTG was added to a concentration of 10mM and the cultures incubated at 37°C for 1 hour to allow expression of the recombinant protein. Cells were harvested by centrifugation then resuspended in 50mM Tris-HCl (pH7.5), 1mM EDTA, 1mM DTT, 1mM PMSF. The cells were lysed by subjecting them to three cycles of freezing in liquid nitrogen and thawing at 37°C before solid material was removed by centrifugation at full speed in a microcentrifuge for 30 minutes at 4°C. Glycerol was added to the supernatant to 10% before storage of the extracts at -70°C.

2.2.6.2 Preparation of CAT extracts from *Drosophila* tissues

100 pairs of ovaries or 100 female carcasses following ovary dissection were washed in PBS then resuspended in 500μl 250mM Tris.Cl pH7.8. Tissues were homogenised as in Section 2.2.5.2 then subjected to five cycles of freezing in liquid nitrogen followed by thawing in water at 37°C. Following a five minute incubation at 65°C samples were centrifuged for 3 minutes at full speed in a microcentrifuge to pellet solid matter then the supernatant was removed and stored in aliquots at -70°C. Protein concentration was determined as in Section 2.2.6.4.

2.2.6.3 Preparation of nuclear extracts

Nuclear extracts were prepared from *Drosophila* ovaries by a method based on that of Dignam *et al.* (1993). Ovaries were washed once in 5 volumes PBS and once in 5 volumes nuclear extract buffer A. Material was resuspended in two volumes buffer A and homogenised for five minutes (see Section 2.2.5.2). Nuclei and cell debris were collected by centrifugation at full speed in a microcentrifuge for 4 minutes at 4°C and resuspended in a small volume of buffer B. 5M NaCl was added to give a final concentration of 750mM and nuclei were incubated 30 minutes.
on ice. Insoluble material was removed by centrifugation at 50000rpm in a Beckman Ti-100 rotor for 60 minutes at 4°C. Glycerol was added to the nuclear extract to 20% before storage at -70°C.

2.2.6.4 Estimation of protein concentration by Bradford assay

Protein samples were mixed with 1ml H2O and 1ml Coomassie Protein Reagent (Pierce). Absorbance was measured at 595nm after a five minute incubation at room temperature. Concentrations were derived from a standard curve plotted from measurements of bovine serum albumin carried out in identical buffer conditions at the same time. For each sample two different dilutions were measured and the mean value taken as the concentration.

2.2.6.5 Measurement of CAT concentration in tissue extracts

Concentrations of CAT in extracts prepared as in section 2.2.6.2 were measured using a CAT ELISA method (Boehringer mannheim). Samples in a volume of 200μl were incubated for 1 hour at 37°C in the wells of microtitre plates pre-coated with antibodies against CAT. After washing the wells, a solution of anti-CAT antibodies conjugated to dioxigenin was added and the plates incubated for a further hour at 37°C. After a second washing step a solution containing antidioxigenin antibodies conjugated to peroxidase was added and the plates incubated again for one hour at 37°C for antibody binding. The antibody solution was removed by washing and a substrate for peroxidase was added. Peroxidase activity converts this into a green product. Absorbance was measured at 410nm using a Minireader II plate reader (Dynatech) against a blank consisting of a well treated identically to the experimental wells except that 250mM Tris.Cl pH7.8 was added instead of the sample in the initial step. Absorbance readings were converted to CAT concentrations using a standard curve prepared with results from measurements of several dilutions of a CAT standard solution of known concentration carried out at the same time as the experimental measurements.
2.2.7 Protein-DNA interactions

2.2.7.1 cDNA library screen

The λgt22a Tolias ovarian cDNA expression library was screened by a method modified from that of Vinson et al. (1988). 10^6 plaque forming units were plated on 150mm diameter plates at a density of 5 x 10^4 per plate and incubated for 7 hours at 37°C. Supported nitrocellulose filters (Hybond-C extra, Amersham) were rinsed in 10mM IPTG and blotted dry on filter paper before being laid on the plates. Filters and plates were marked with orientation spots. Plates were incubated for 5 hours at 37°C for expression of recombinant protein, cooled to 4°C for 10 minutes and the filters lifted from the plates.

Filters were air-dried, protein side up, on filter paper and stored up to 16 hours before processing. All subsequent steps were carried out at 4°C and using solutions precooled to 4°C.

Filters were incubated in 6M Guanidine Hydrochloride (Sigma) in HEPES binding buffer for 10 minutes with gentle shaking to denature recombinant protein. The filters were then incubated successively in 3M, 1.5M, 0.75M, 0.38M and 0.19M Guanidine Hydrochloride in HEPES binding buffer for 5 minutes each to renature the protein. Filters were rinsed twice in binding buffer then blocked for 60 minutes in block solution. Filters were rinsed twice in binding buffer to remove the milk. Site 1 probe (see Section 2.2.3.7) was added to binding buffer at a concentration of 1x10^6 to 5x10^6 cpm/ml. Poly dI.dC-poly dI-dC (Pharmacia) was added to a concentration of 1μg/ml. Filters were incubated in this binding mixture with 10ml of mixture per 150mm diameter filter at 4°C for 5 hours with gentle agitation. Filters were then washed in binding buffer for 7.5 minutes with gentle agitation. This wash was repeated twice more before the filters were air-dried, protein side up, on filter paper and autoradiographed.

Proteins binding to the probe are visualised as spots on the autoradiograph. Using the orientation spots marked on the filters and plates the positions on the plate corresponding to positives was identified. Agar plugs from these positions were picked and eluted into 1ml of phage buffer for 2 hours at 25°C. These were plated
on 75mm diameter plates which were screened as above. This was repeated once more to obtain plaque-pure phage clones.

To test pure clones for sequence specificity in DNA binding 75mm diameter filters were prepared as above up to the blocking step. They were then cut in half. One half was screened as above. The other was incubated in a binding mixture as above except that a probe lacking the site 1 sequence was used instead of the site 1 probe.

2.2.7.2 Gel retardation assays

A 10μl mixture was prepared containing 1μg of *Drosophila* nuclear extract or *E. coli* lysogen extract, 1μg of double-stranded [poly dI-dC.poly dI-dC] carrier DNA (Pharmacia), 20mM HEPES pH7.76, 40mMKCl, 2mM MgCl₂, 0.1mM EDTA, 1mM DTT and 10% glycerol. In competition assays the specific competitor DNA, prepared by PCR and gel purified, was also included in this mixture. After incubation on ice for 10 minutes, or 1 hour in experiments with antisera, approximately 1ng of end-labelled probe DNA was added and the incubation continued for a further 20 minutes. The probe was a 193bp *EcoRI/XbaI* fragment of p186.T1 which contains the I factor 5'UTR. Complexes were separated by electrophoresis in a 5% acrylamide gel (30:0.8 acrylamide:bis-acrylamide ratio) in 1XTBE buffer at 150V. Gels were dried under vacuum and the labelled DNA detected by autoradiography.
Results
Chapter 3 – A screen for site 1 binding proteins
A screen for site 1 binding proteins

3.1 Introduction

As discussed in Section 1.3.4.2 transposition of the I factor is regulated at the level of transcription. The activity of the promoter, in the first 40bp of the element, is stimulated in ovaries by a downstream enhancer located between base pairs 41 and 186 (Udomkit et al., 1996). The enhancer causes activation to the highest levels in the germ line of reactive females, where transposition occurs at high frequency. To understand the regulation of I factor transcription it will be necessary to identify the proteins that bind within this region.

Udomkit et al. (1996) discovered a binding activity in ovarian nuclear extracts which interacts with the I factor enhancer. This recognises base pairs 137 to 150, an element known as site 1. A reporter construct in which site 1 had been deleted showed a drastic reduction of enhancer activity and the abolition of tissue-specificity in transcription (Udomkit et al., 1996) suggesting that site 1 is an essential component of the enhancer. To identify the protein or proteins that bind to site 1 an ovarian cDNA expression library was screened using a site 1 probe. This chapter describes the characterisation of two clones isolated by this method.

3.2 Results

3.2.1 Isolation and DNA binding specificity of cDNA clones encoding site 1 binding proteins

The sequence-specific site 1 binding activity in ovarian nuclear extracts can be detected in a gel retardation assay (see Figure 3.1). Following incubation of a radiolabelled probe consisting of the 186bp I factor 5'UTR with ovarian nuclear proteins several retarded bands can be separated by electrophoresis (Figure 3.1, lane 3). All of these bands are competed by the addition of an excess of unlabelled 5'UTR DNA (Figure 3.1, lane 4). However a competitor derived from the 5'UTR in which site 1 has been deleted fails to compete for binding (Figure 3.1, lane 5),
demonstrating that binding is sequence-specific and that site 1 is required. A competitor consisting solely of site 1 competes, showing that site 1 is sufficient for binding (Figure 3.1 lane 6). This binding activity is also present in extracts from Schneider line 2 tissue culture cells (Figure 3.1, lane 2), female carcass and males (Udomkit et al., 1996).

![Figure 3.1 Site 1 binding activity in Drosophila nuclear extracts. Gel retardation assays using the 186bp I factor 5' UTR as a probe and nuclear extracts from SL2 tissue culture cells (lane 2) or charolle ovaries (lanes 3-6). Lane 1 is a control with no protein. Unlabelled competitor DNA in 100 fold molar excess was added in the experiments in lanes 4-6 as follows. Lane 4 - 5'UTR, lane 5 - 5'UTR with site 1 deleted, lane 6 - site 1]

In an effort to identify proteins that can bind site 1 an ovarian cDNA expression library was screened with a site 1 probe. The library (Stroumbakis et al., 1994) is cloned in the vector λgt22a and has a complexity of approximately 500,000. 10^6 plaques were screened using a radiolabelled probe consisting of multiple copies of the site 1 sequence. Two positive clones, named Gtol1 and Gtol2, were purified. To test the products of these clones for sequence-specific DNA binding, the phage were incorporated into the genome of E. coli by lysogeny. Lysogens of each were cultured and IPTG added to induce expression of the recombinant protein encoded by the integrated prophage. Whole cell extracts from induced and uninduced cultures were tested for binding to I factor 5' UTR DNA in gel retardation assays (Figure 3.2).
Extracts of Gtoll lysogens give a single retarded band in this assay (Figure 3.2 lanes 2, 3, 4, 5, and 7). Induction with IPTG results in a significant increase in the intensity of this band (lanes 3, 5 and 7), showing that the protein responsible is expressed from the inducible tac promoter of the prophage. No 5'UTR binding was detected in extracts from Gtol2 lysogens (lanes 8-13).

The product of Gtol1 was tested for sequence-specificity in DNA binding using excess unlabelled competitor DNA (Figure 3.3). The 5'UTR sequence competes efficiently for binding (lane 3), while the 5'UTR in which site 1 has been deleted does not (lane 4). This shows that binding is sequence-specific and requires site 1.
Figure 3.3 - Site 1 binding activity in Gtoll lysogen extracts. Gel retardations with extracts from induced Gtoll lysogen cultures. Lane 1 - no protein control. Lane 2 - induced Gtoll extract, no competitor. Lane 3 - induced Gtoll extract plus unlabelled 5'UTR. Lane 4 - induced Gtoll extract plus unlabelled 5'UTR with site 1 deleted.

The negative result obtained with the Gtoll2 lysogen (Figure 3.2 lanes 8-13) may be due to inappropriate DNA binding conditions, lack of expression or insolubility of the protein produced in the lysogen cultures. The assay that was used to screen the library includes denaturation and renaturation steps designed to reconstitute the 3-dimensional structure of expressed proteins (Vinson et al, 1988). To test binding specificity in that assay the probe S1mut was used. This has identical ends to the site 1 probe used in the library screen but no similarity to site 1 in the internal sequence (see Figure 3.4). Filters were prepared from phage plaques expressing the Gtoll1 and Gtoll2 proteins. These were cut in half and one half of each was screened with site 1 probe and the other half with S1mut. As expected the Gtoll1
protein binds only to the site 1 probe confirming that it recognises the site 1 sequence. Both probes are bound by the Gtol2 protein demonstrating that binding is not site 1-specific.

<table>
<thead>
<tr>
<th>Probe</th>
<th>Site 1</th>
<th>S1mut</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gtol1</td>
<td>GATCCGAAAACACAAATAACCGCTAG</td>
<td>GATCCGAGTTTCACACCTTAG</td>
</tr>
<tr>
<td>S1mut</td>
<td>GATCCGAAAACACAAATAACCGCTAG</td>
<td>GATCCGAGTTTCACACCTTAG</td>
</tr>
<tr>
<td>Gtol2</td>
<td>GATCCGAAAACACAAATAACCGCTAG</td>
<td>GATCCGAGTTTCACACCTTAG</td>
</tr>
<tr>
<td>Site 1</td>
<td>GATCCGAAAACACAAATAACCGCTAG</td>
<td>GATCCGAGTTTCACACCTTAG</td>
</tr>
<tr>
<td>S1mut</td>
<td>GATCCGAAAACACAAATAACCGCTAG</td>
<td>GATCCGAGTTTCACACCTTAG</td>
</tr>
</tbody>
</table>

Figure 3.4 - Sequence-specificity of proteins encoded by Gtol1 and Gtol2. Binding assays were carried out using filters prepared from expressed Gtol1 (left) and Gtol2 (right) plaques. The filters were cut in half and the left half of each screened with the site 1 probe while the right half was screened with S1mut. The sequences of the probes are shown.

3.2.2 Gtol1 and Gtol2 contain partial cDNAs of known Drosophila melanogaster genes

Phage DNA from Gtol1 and Gtol2 were purified and digested with NotI and Sall to excise the cDNA inserts. These were subcloned into pBluescript and their sequences partially determined. These were used to search databases of known nucleic acid sequences using the MPsearch program (Sturrock and Dryden, 1997).

Both clones contain sequence from known Drosophila melanogaster genes. Gtol1 encodes a C-terminal fragment of Adult Enhancer Factor-1 (AEF-1, Falb and Maniatis, 1992a, Genbank accession M90755), a transcriptional repressor acting at three enhancers directing transcription in the adult fat body as well as the alcohol dehydrogenase proximal promoter (Falb and Maniatis, 1992a,b, An and Wensink,
part of the cDNA of the large subunit of replication factor C (lsRFC, Allen et al., 1998, U97685), also known as germline transcription factor 1 (Gnf1, Frank and Cohen, 1993, L17340). This protein was originally described as binding to an element within a female germline enhancer at the hsp26 locus (see Section 3.3.2 below).

3.2.3 The Gtol1 insert is a partial Adult Enhancer Factor-1 cDNA

The cDNA insert of Gtol1 is 1.9kb long measured by comparison with size markers in agarose gel electrophoresis. At the 5' end of the coding strand 243bp were sequenced extending into the cDNA insert from the SalI site of λgt22a. These are identical to base pairs 1003 to 1246 of the published cDNA of adult enhancer factor 1 (AEF-1, Falb and Maniatis, 1992a, Genbank accession M90755, see Figure 3.5). At the 3' end 90bp were sequenced. Base pairs 2830-2909 of the published sequence are present in Gtol1 and are followed by a C then 18 As on the coding strand. There is a consensus polyadenylation signal AATAAA from base pairs 2907 to 2912 (Falb and Maniatis, 1992). The portion of Gtol1 that was not sequenced presumably corresponds to base pairs 1246 to 2830 of the AEF-1 cDNA. If this is the case the size of the insert should be 1924bp which agrees with the measured value (1.9kb). Gtol1 therefore includes base pairs 1003 to 2909 and encodes a C-terminal fragment of AEF-1 consisting of amino acids 227 to 321.
3.2.6 The Gto12 insert is a partial cDNA of IsRFC/Gnf1

The clone Gto12 has a cDNA insert of 2.7kb measured by comparison with markers of known size in agarose gel electrophoresis. The insert contains part of the published cDNA sequence of the large subunit of replication factor C (IsRFC, Allen et al., 1998, U97685), also known as germline transcription factor 1 (Gnf1, Frank and Cohen, 1993). The organisation of the Gto12 sequence is not straightforward and is shown schematically in Figure 3.7.
**Figure 3.7 – Sequence organisation of Gtol2 in relation to lsRFC.** A. Nucleotide sequence at the ends of the Gtol2 insert. Only the coding strand is shown. Underlined sequence is present in the λGt22a vector. Sequence in italics is not present in the published lsRFC/Gnf1 genomic DNA. Sequence in bold type is present in the published cDNA. Numbers with arrows refer to the lsRFC/Gnf1 genomic DNA as in L17340. B. Diagram of lsRFC showing the eight motifs conserved between lsRFCs and the region encoded by Gtol2.

At the 5’ end of the coding strand the sequence of 160bp was determined. From base pair 30 to 160 the sequence is identical to base pairs 537-667 of the published lsRFC/Gnf1 cDNA (U97685). At the extreme 5’ end of the coding strand of Gtol2 there is a sequence of 29 base pairs which is not present anywhere in the lsRFC/Gnf1 genomic DNA (L17340). This is shown in Figure 3.7a and encodes the amino acids PRVRGGGTAG in frame with the downstream lsRFC/Gnf1 open reading frame.

At the 3’ end of the coding strand 157bp were sequenced. These are identical to the lsRFC/Gnf1 genomic sequence (L17340) from base-pairs 4328 to 4485 where

the clone terminates with a polyA tail of which 23 deoxyadenosines were sequenced (see Figure 3.7a). This site is in the 3' untranslated region of the gene but does not correspond to a consensus polyadenylation signal.

Assuming that the sequence between the two regions that have been determined is identical to the published cDNA, Gtol2 encodes amino acids 180 to 986 of the 986 amino acid protein. If this is correct the size of the insert should be 2609bp compared with 2.7kb determined by electrophoresis. The difference is within the range of the error expected with this technique.

The 5' 29bp insertion of Gtol2 may be present in a natural *Drosophila* RNA or may be the result of a cloning artifact. In an attempt to resolve this question the Gtol2 insert was used as a probe to screen another ovarian cDNA library (Jongens et al., 1992) by hybridization. Six clones were purified labelled Gtflan1-6. The library is cloned into the λZap vector (Short et al., 1988) enabling a plasmid containing the cDNA insert to be excised *in vivo* by site-specific recombination. Plasmids containing the inserts of Gtf Jan 1-6 were purified. None had a cDNA insert as long as that of Gtol2 showing that a full length IsRFC/Gnf1 cDNA had not been obtained.

### 3.3 Discussion

The data presented in this chapter show that site 1, an element believed to be essential for correct expression of the *I* factor, is bound specifically by a fragment containing part of the DNA binding domain of AEF-1, a transcription factor originally described as regulating the alcohol dehydrogenase gene (Falb and Maniatis, 1992a). There is no evidence that IsRFC/Gnf1 protein shows sequence-specific binding to site 1.

#### 3.3.1 Adult Enhancer Factor-1

AEF-1 was discovered by virtue of its binding to a site in the alcohol dehydrogenase (Adh) adult enhancer (AAE, Falb and Maniatis, 1992b). The AAE activates transcription in the adult fat body and AEF-1 represses this activation function (Falb and Maniatis, 1992a, b). AEF-1 also binds to elements in two other adult fat body enhancers, the yolk protein 1 and 2 fat body enhancer 1 (FBE1, Falb...
and Maniatis, 1992a) and the upstream region of the fat body protein-1 gene (Lapie et al., 1993). There is also an AEF-1 binding site overlapping the Adh proximal promoter (Ren and Maniatis, 1998). AEF-1 binding represses transcription at all of these loci.

The known AEF-1 binding sites have similar sequences (see Figure 3.9). Comparisons between them, and studies of the requirements for binding at the yolk protein fat body enhancer, produced the consensus binding site shown in Figure 3.9 (An and Wensink, 1995a). Site 1 can be aligned with this consensus in several ways, one of which is shown. Using this alignment site 1 matches the consensus at 9 positions. The merits of each of the possible alignments of site 1 with the other binding sites and the sequence requirements for AEF-1 binding will be discussed in Chapter 4.

Consensus

<table>
<thead>
<tr>
<th>C</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>CACAACNA</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>melanogaster Adh (AAE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCAGCAACAAACACGATC</td>
</tr>
<tr>
<td>ACCACAAAACAAAATAAC</td>
</tr>
<tr>
<td>TGCAACTACAATAATGTT</td>
</tr>
<tr>
<td>AAACCAAACCTAAACCG</td>
</tr>
<tr>
<td>TACACAAGCAACAAAAA</td>
</tr>
<tr>
<td>CCCCCAGCAATAAAAATCT</td>
</tr>
<tr>
<td>AACAACAAACAAACAAT</td>
</tr>
</tbody>
</table>

Figure 3.9 - Alignments of published AEF-1 binding sites and site 1. The consensus was defined by An and Wensink (1995a). References for the other binding sites are as follows: AAE, Drosophila mulleri Adh and YP1 and YP2 - Falb and Maniatis, 1992a. Adh proximal initiator element (pie) - Ren and Maniatis, 1998. Human Adh - Falb and Maniatis, 1992b. Rat Adh - Potter et al., 1994.

It has not been demonstrated that the site 1 binding activity present in ovarian extracts is due to AEF-1. However it seems likely that this is the case given the binding specificity of the protein encoded by Gtoll and the similarities between site 1 and the known AEF-1 binding sites. Like the site 1 binding activity, AEF-1 is...
abundant in ovaries and is also found in other adult tissues in both male and female flies (Falb and Maniatis, 1992a).

AEF-1 has three identifiable motifs (see Figure 3.5a), a zinc finger DNA binding domain, an alanine-rich region and a glutamine-rich domain (Falb and Maniatis, 1992b, see Figure 3.10). The DNA binding domain has four consecutive zinc fingers. The discovery that the product of Gtoll binds DNA demonstrates that the two C-terminal zinc fingers are sufficient for binding in vitro. The alanine-rich region extends from amino acids 103 to 127. A similar stretch in the Drosophila Krüppel repressor confers transcriptional repression in mammalian cells (Licht et al., 1990). Glutamine-rich domains are believed to be involved in protein-protein interactions. In certain other transcription factors, glutamine-rich domains have been implicated in transcriptional activation (Courey et al., 1988). At the DNA level the glutamine-rich domain of AEF-1 is encoded by a $(CAX)_2$ trinucleotide repeat of the type known as opa (Wharton et al., 1985). A possible role for this motif in $I$ factor regulation is addressed in Chapter 7.

The discovery that site 1 is bound by AEF-1, a known transcriptional repressor, was surprising. The effect of deleting site 1 is to reduce transcription from the $I$ factor promoter (Udomkit et al., 1996) so it was predicted that any site 1 binding proteins would be transcriptional activators. The question of whether AEF-1 is an activator or a repressor of $I$ factor transcription is addressed by the experiments described in Chapters 4, 5 and 6.

### 3.3.2 IsRFC and germline transcription factor 1

The product of the IsRFC/Gnfl gene was originally described as a protein binding to the CAACAA element (Frank and Cohen, 1993). This forms part of the regulatory region of the hsp26 gene which is expressed at high levels in the female germ line (Frank et al., 1992). A 171bp segment containing this element forms a germ line-specific enhancer when placed in two or more tandem copies upstream of a basal promoter (Frank et al., 1992). Mutation of the CAACAA elements within this enhancer abolishes activation. So this element, like site 1 of the $I$ factor, is an
essential part of a tissue-specific enhancer active in the female germ line. The sequence of the CAACAA element can be aligned with site 1 giving identical base pairs at 7 consecutive positions out of the 14bp site 1 footprint (see Figure 3.10). Given this functional and structural similarity it is plausible that the same ovarian protein may recognise each of these sequences. At the time it was first isolated no lsRFCs from other organisms had been described. Consequently the protein was named germline transcription factor 1 (Gnfl) and was believed to be an activator of hsp26 transcription (Frank and Cohen, 1993).

CAACAA element (hsp26):  
G A C A A C A A C T A C  
Site 1:  
A C A A A A A C A A C A A T A C

Figure 3.10 - allignment of site 1 and the CAACAA element (Frank et al., 1992)

When the gene for the murine large subunit of replication factor C (IsRFC) was cloned it was found to be homologous to Gnfl (Luckow et al. 1994). The amino acid sequences show 44% identity and 61% similarity. Also known as activator 1, RFC is a multisubunit complex which is a cofactor for DNA polymerases delta and epsilon (Tsurimoto and Stillman, 1989, Podust et al., 1992, Burgers, 1993). The complex recognises the junction of the primer and template at a replication fork (Tsurimoto and Stillman, 1991), recruiting proliferating cell nuclear antigen (PCNA) and polymerase to initiate DNA synthesis (Lee and Hurwitz, 1990, Lee et al., 1990). Two DNA binding domains have been identified in lsRFC proteins. The C-terminal half of the human lsRFC binds in a structure-specific manner to junctions between single and double-stranded regions having a recessed 3' hydroxyl group (Uhlmann et al., 1997). This activity is consistent with the proposed role of the complex in vivo. An N-terminal fragment of the Drosophila lsRFC/Gnfl, including a conserved domain similar to prokaryotic ligases (labelled I in Figure 3.7b), shows a different specificity requiring a duplexed 5' phosphate (Allen et al., 1998).

In at least three previous expression screens of the type employed here, clones encoding lsRFC have been isolated by due to structure-specific DNA binding (Luckow et al., 1994, Stuempfle and Floros, 1997, Allen et al., 1998). The probes
used in such screens often have 5' single-stranded overhangs at the ends, creating a structure analogous to a primer-template junction which may be recognised by IsRFC. The site 1 probe used here had overhanging ends to facilitate ligation to produce concatemers. The discovery that the protein encoded by Gtol2 does not show specificity for site 1 suggests that it may have been isolated as a result of this artefact.

It should be noted that the protein encoded by Gtol2 lacks 180 amino acids of Gnf1 at the N-terminus. It may be that the full length protein does have a sequence-specific DNA binding activity. There are reports of sequence-specific DNA binding proteins that are homologous or identical to IsRFC (Jin et al., 1994, Halligan et al., 1995, McGehee and Habener, 1995). These include a mouse protein, VDJP, which binds specifically to the nonamer repeat element involved in VDJ recombination (Halligan et al., 1995). This contains an N-terminal fragment including the ligase homology domain and is apparently produced from the same gene as the mouse IsRFC by alternative splicing. It remains a possibility that Drosophila IsRFC or a different protein product of the same gene, does bind specifically to the CAACAA element (and perhaps also to site 1) but there is no published evidence for this.
Chapter 4 - Effect of site 1 mutations on AEF-1 binding in vitro
4.1 Introduction

4.1.1 Is AEF-1 a repressor or an activator of I factor transcription?

The results presented in Chapter 3 suggest that site 1, an element in the I factor 5' untranslated region, is bound specifically by AEF-1 present in ovarian nuclear extracts. At the other loci that AEF-1 is known to regulate, the protein acts as a repressor of transcription. Mutagenesis of the AEF-1 site in the Adh adult enhancer (AAE) resulted in a five to ten fold higher level of Adh expression in the adult fat body (Falb and Maniatis, 1992a). In addition expressing AEF-1 in tissue culture cells reduced expression of a reporter gene construct containing the AAE (Falb and Maniatis, 1992a). At the AAE there is a site adjacent to the AEF-1 binding site which is required for enhancer function and which can be bound by the mammalian CAAT/enhancer binding protein (C/EBP, Falb and Maniatis, 1992b). AEF-1 can displace C/EBP from this binding site in vitro suggesting that the displacement of a Drosophila C/EBP-like transcriptional activator may be the mechanism by which AEF-1 represses transcription (Falb and Maniatis, 1992b).

Studies of a regulatory element present in the fat body enhancer 1 (FBE1) of the yolk protein 1 and 2 genes also identify AEF-1 as a repressor. This element, o, consists of overlapping binding sites for AEF-1, DSX and C/EBP. An artificial enhancer consisting of four copies of o activates transcription in the female fat body (An and Wensink, 1995b). Mutagenesis of the AEF-1 binding sites in this enhancer resulted in an increase in reporter gene expression in ovaries (An and Wensink, 1995b). AEF-1 and DSX cannot bind simultaneously to o in vitro suggesting that AEF-1 may repress transcription at this enhancer by excluding DSX (An and Wensink, 1995b).

At the Adh proximal promoter an AEF-1 binding site overlaps the initiator and AEF-1 represses transcription in adult flies, presumably again by competing with transcription factors for their binding sites on the DNA (Ren and Maniatis, 1998). A further example occurs in the regulatory region of the Fat body protein 1 gene where
deletion of an AEF-1 binding site resulted in an increase in expression (Lapie et al., 1993).

In the case of the I factor, reporter gene experiments have shown that the complete 5'UTR confers enhanced transcription in ovaries to a level approximately 20 fold greater than that in female carcass (McLean et al., 1993, Udomkit et al., 1996). Deletion of site 1 resulted in a dramatic reduction in expression and the abolition of ovary specificity (Udomkit et al., 1996), which might suggest that site 1 is bound by a positive regulator of transcription. However site 1 is a binding site for AEF-1, which acts as a repressor of other genes. It is possible that AEF-1 may act as a repressor at some loci and as an activator at others. Alternatively the deletion of site 1 may affect reporter gene expression by a mechanism other than the prevention of AEF-1 binding.

4.1.2 Dual-function regulators

There are several examples of DNA binding proteins that can activate or repress transcription under different circumstances. One mechanism for this is that the binding of additional proteins to nearby sites can alter the effect of a transcription factor on a promoter. The Dorsal protein activates transcription of twist and snail, genes expressed in the ventral region of the Drosophila embryo, but represses the promoters of zerknült and decapentaplegic (Ray et al., 1991). Activation by Dorsal requires only the presence of Dorsal binding sites (Thisse et al., 1991, Pan and Courey, 1992). In contrast repression by Dorsal at the zerknült gene requires binding sites for other transcription factors that are necessary to convert Dorsal from an activator to a repressor (Pan and Courey, 1992, Jiang et al., 1992, Jiang et al., 1993, Kirov et al., 1993). The binding of proteins to these sites results in the formation of a repressive complex which probably includes the groucho corepressor (Dubnicoff et al., 1997, Valentine et al, 1998).

The glucocorticoid receptor (GR) regulates transcription positively and negatively in a variety of ways depending on the context of its binding site. Activation of transcription results from hormone-bound GR binding to glucocorticoid response elements (GREs) present at the target genes. GREs interact synergistically with the binding sites for other transcription factors to promote
activation (Strahle et al., 1988). Repression by GR can occur by a variety of mechanisms. Certain transcriptional activators, including AP-1 and NFκB are repressed by GR in a hormone-dependent manner that does not require contacts between GR and DNA, apparently through direct protein-protein interactions (König et al., 1992, Caldenhoven et al., 1995). At other genes DNA binding by GR is required for repression. At the proliferin gene a GRE is located next to a binding site for the dimeric transcription factor AP-1 which can consist of a c-Jun-c-Jun homodimer or a c-Jun-c-Fos heterodimer. In the absence of AP-1 binding GR has no effect on transcription. When the AP-1 site is occupied by a homodimer, GR acts as an activator of transcription but when a heterodimer is bound GR acts as a repressor (Diamond et al., 1990). One final mode of repression by GR that has been proposed involves direct competition for DNA binding with basal or regulatory transcription factors. At the c-fos and bovine prolactin genes, GREs overlap binding sites for activating proteins (Karagianni and Tsawdaroglou, 1994, Sakai et al., 1988). The human osteocalcin gene has a GRE overlapping the TATA box and there is evidence that GR represses transcription by competition for binding with the basal transcription factor TBP (Stromstedt et al., 1991, Meyer et al., 1997). As discussed above, the same mechanism of competition for binding at overlapping sites is thought to be the means by which AEF-1 represses transcription at the alcohol dehydrogenase and yolk protein enhancers.

These and other examples demonstrate that DNA binding proteins can participate in numerous different interactions with other proteins and have different effects on transcription depending on the context. It is possible therefore that AEF-1 could act as an activator in the context of the I factor enhancer while repressing transcription at other loci.

4.1.3 Deletion of site 1 may affect expression by a mechanism not directly related to AEF-1 binding

It is possible that the result obtained when site 1 was deleted may not be due to prevention of AEF-1 binding. It may be that in the creation of the site 1 deleted construct sequences from either side were brought together to form a new binding site for a transcriptional repressor. Alternatively the deletion may affect nucleosome
positioning in the enhancer or the translation of the *I-CAT* RNA. If any of these possibilities is correct then AEF-1 may be a repressor of *I* factor transcription, or may have no effect at all.

A related idea is that the deletion of site 1 disrupts one or more binding sites for proteins other than AEF-1. This is supported by the arrangement of transcription factor binding sites at other loci, where AEF-1 sites overlap or are adjacent to binding sites for transcriptional activators (see above). In order to investigate these possibilities and to gain more information on the role of the AEF-1 binding site in *I* factor regulation a more detailed mutagenesis was carried out in the vicinity of site 1. The effects of point mutations within and around site 1 on AEF-1 binding *in vitro* and on *I* factor enhancer activity *in vivo* have been investigated. This chapter describes the creation of mutants and *in vitro* binding studies while the effects on expression *in vivo* are addressed in Chapter 5.

### 4.2 Results

#### 4.2.1 Mutagenesis of site 1

Alterations were made in the vicinity of site 1 with the aim of identifying two classes of mutations. Firstly, mutations that reduce AEF-1 binding will give direct information as to the role of AEF-1 in regulating the *I* factor. Secondly, mutations adjacent to the AEF-1 site that do not affect AEF-1 binding may affect the binding of other proteins involved in transcriptional regulation.

Mutagenesis was carried out by overlap extension PCR (Ho *et al.*, 1989). Figure 4.1 illustrates the method using mutagenesis of base pairs 137 and 138 as an example. The plasmid p186.T1, which contains the wild type *I* factor 5'UTR fused upstream of the *CAT* open reading frame was used as a template. Oligonucleotides 664C and CAT.R prime polymerase synthesis on p186.11 in opposite directions across the 5'UTR. Pairs of mutagenic primers were designed that direct synthesis outward from site 1 and contain degenerate nucleotides at the sites to be mutated. The sequences of these are shown in Section 2.1.2.11. In conjunction with 664C and CAT.R, each pair was used to amplify two overlapping fragments which together cover the entire 5'UTR. The products produced using primers R137.138 and
L137.138 are shown in Figure 4.2 lanes 2 and 3. These products were gel purified and a small amount of each was included in a final PCR using 664C and CAT.R to amplify across the entire 5'UTR (Figure 4.2 lane 4). The mutant products generated were subcloned and sequenced across the entire length of the final PCR in each case. No changes from the original p186.T1 sequence were observed in any of the mutant clones obtained at any positions other than those deliberately mutated. An example of a sequencing gel showing the results of mutagenesis using primers R137.138 and L137.138 is shown in Figure 4.4. The wild type sequence from nucleotides 130-150 of the I factor is CTTAACAACAAAAACAACAA. The clones shown in Figure 4.2 have alterations at positions 137 and 138. Clone 1 has 137C138G, clone 2 137C138T, clone 3 137G138G and clone 4 137G138T.
Figure 4.1. Site-directed mutagenesis of site 1 by overlap extension PCR. Nucleotides of the template p186T1 are shown in black type. Nucleotides of the two initial left and right PCR products are shown in blue and green respectively. Degenerate nucleotides that constitute substitutions at the nucleotides to be mutated are shown in red. B = T, C or G; D = T, G or A; V = C, G or A; H = T, C or A.
Figure 4.2 PCR products generated during mutagenesis of base pairs 137 and 138 of the I factor 5'UTR. 5µl of the PCR product was loaded in each lane. Lane 1 - Molecular weight markers, Lane 2 - Left PCR product, Lane 3 - Right PCR product, Lane 4 - Final mutant PCR product, Lane 5 - PCR using primers 664C and CAT.R with p186.T1 template, Lane 6 - Negative control PCR with primers 664C and CAT.R and no template.

Figure 4.3 Sequencing of the products of mutagenesis of base pairs 137 and 138. Clone 1 - 137C138T, Clone 2 137G138G, clone 3 - 137G138T.
4.2.2 Effect of site 1 mutagenesis on binding by AEF-1 in ovarian nuclear extracts

Initially three pairs of adjacent bases within site 1 were chosen for mutagenesis. These are A137C138, A143C144 and T150A151. Three different products were obtained with alterations at positions 137 and 138, two with changes at positions 143 and 144 and four with mutations at positions 150 and 151. Their sequences are shown in Figure 4.4. All of these mutants were tested for binding to AEF-1 in an ovarian nuclear extract. Gel retardation assays were carried out using the wild type 5'UTR as a probe and competitors consisting of the different mutant versions of the 5'UTR (Figure 4.5).

In this assay the wild-type 5'UTR competes efficiently for binding so that no retarded probe can be seen (Figure 4.5, lane 3). The mutations at positions 137 and 138 reduced the ability of the 5'UTR to compete for AEF-1 binding (lanes 5-7). The same is true of the mutations at positions 143 and 144 (lanes 8 and 9). Altering the sequence at bases 150 and 151 had no detectable effect on competition for binding by the 5'UTR (lanes 10-13).
Figure 4.5 Binding of AEF-1 in ovarian nuclear extract to mutant versions of the / factor 5'UTR. Gel retardation assays were carried out using the 186bp / factor 5' UTR as a probe and nuclear extract from JA ovaries. Lane 1 - control with no protein, Lane 2 - no competitor. Lanes 3-13 - 100ng competitor DNA included - approximately 100 fold molar excess over probe. Competitors: Lane 3 - Wild-type 5'UTR, lane 4 - 5'UTR with site 1 deleted, lane 5 - 137G138G, lane 6 - 137G138T, lane 7 - 137C138T, lane 8 - 143C144G, lane 9 - 143C144A, lane 10 - 150G151C, lane 11 - 150C151C, lane 12 - 150G151G, lane 13 - 150A151G
With the aim of obtaining a mutant that showed no binding to AEF-1, primers were designed with degeneracies at four positions: 137, 138, 153 and 154. Using these primers two mutants were obtained, with the sequences 137G138G143C144G and 137C138T143C144G (see Figure 4.4). In a gel retardation assay these showed no detectable binding to ovarian AEF-1 (Figure 4.6 lanes 7 and 8).

As discussed above, one aim of these experiments was to investigate the possibility that the AEF-1 site overlaps binding sites for other transcription factors. For this purpose mutations at either side of site 1 are required that have no effect on AEF-1 binding. The mutations at positions 150 and 151 lie on the 3' side of site 1 on the coding strand and do not affect binding (Figure 4.5 lanes 10-13). To extend the mutagenesis at the 5' end primers were designed to alter base pairs C135 and A136. Two mutants were obtained using these primers with the sequences T135C136 and G135C136 (Figure 4.4). In a gel retardation assay both of these mutants competed efficiently for binding (Figure 4.6, lanes 5 and 6).

For studies of the effects of mutations in and around site 1 on expression from the I factor promoter in vivo one example of each set of mutations was chosen. These are 135T136C, 137G138G, 143C144G, 150G151C and 137G138G143C144G. Figure 4.7 shows a direct comparison between the binding affinities of these mutants for ovarian AEF-1.

4.2.3 Measurement of relative affinities for ovarian AEF-1 of mutant versions of the I factor 5'UTR

With the aim of estimating the relative affinities of the different mutants for ovarian AEF-1, gel retardation assays were carried out using a range of competitor concentrations. To control for possible differences in the specific activity of the probe between experiments, a comparison was made in each case between the mutant and the wild-type 5'UTR. Figure 4.8 shows an experiment in which the wild-type 5'UTR and the 137G138G mutant were compared for efficiency of competition. As the amount of wild-type competitor is increased, the major retarded band becomes undetectable when 16ng is added (Figure 4.8, lane 6) but can be seen faintly when 8ng is added (lane 5). Using the mutant DNA the major band is not seen when 600ng of competitor is added (lane 12) but can be seen at 400ng (lane 11). The wild-
Figure 4.6 Comparison of binding of AEF-1 in ovarian nuclear extract to mutant versions of the \(\text{IFactor}\) 5'UTR. Gel retardation assays were carried out as in figure 4.4. Lane 1 - control with no protein, Lane 2 - no competitor. Lanes 3-13 - 100ng competitor DNA included - approximately 100 fold molar excess over probe. Competitors: Lane 3 - Wild type 5'UTR, lane 4 - 5'UTR with site 1 deleted, lane 5 - 135T136C, lane 6 - 135G136T, lane 7 - 137G138G143C144G, lane 8 - 137C138T143C144G.
Figure 4.7 Comparison of binding of AEF-1 in ovarian nuclear extract to mutant versions of the $I$ factor 5'UTR. Gel retardation assays were carried out as in figure 4.5. Lane 1 - control with no protein, Lane 2 - no competitor. Lanes 3-13 - 100ng competitor DNA included - approximately 100 fold molar excess over probe. Competitors: Lane 3 - Wild type 5'UTR, lane 4 - 5'UTR with site 1 deleted, lane 5 - 135T136C, lane 6 - 137G138G, lane 7 - 143C144G, lane 8 - 150G151C, lane 9 - 137G138G143C144G
Figure 4.8 Gel retardation with titrations of wild-type and 137G138G competitors. Gel retardation assays were carried out using the 186bp I factor 5' UTR as a probe and nuclear extract from JA ovaries. Lane 1 is a control with no extract. Lane 2 - no competitor. Lanes 3-7 - wild type 5'UTR competitor. Lane 3 - 2ng, lane 4 - 4ng, lane 5 - 8ng, lane 6 - 16ng, lane 7 - 32ng. Lanes 8-12 - 5'UTR with 137G138G
Type competitor is therefore between 25 (400/16) and 75 times more efficient as a competitor than the 137G138G mutant. The intensity of the retarded band in lane 5 (8ng of wild-type competitor) is approximately the same as that in lane 11 (400ng of mutant competitor) so it can be estimated that the wild-type competitor is 50 (400/8) times as efficient as the mutant.

Similar experiments were carried out for the other mutants to be used in \textit{in vivo} assays. Figure 4.9 shows the experiment for 143C144G. The result is the same as with the 137G138G competitor in that the major retarded band can be seen faintly in lanes 5 and 11 but is absent in lanes 6 and 12 showing that the mutant competitor is again around 50 times less efficient than the wild-type. A comparison of lanes 6 and 7 in Figure 4.7 confirms that 137G138G and 143C144G are similar in the efficiency with which they compete for AEF-1 binding.

Figure 4.10 shows a titration experiment using the 150G151C mutant. The major band can be seen when 4ng of wild-type competitor are used (lane 4) but not when 8ng are added (lane 5). When the 150G151C mutant is used the band is present at 8ng (lane 9) but absent at 16ng (lane 10). The mutant binds AEF-1 with approximately half the affinity of the wild type.

Finally a titration was carried out for the 135T136C mutant (Figure 4.11). On this gel the major band is seen when 8ng of wild-type competitor is added (lane 3) but not when 16ng is added (lane 4). In the case of the 1365T136C competitor the band is visible faintly when 8ng is added (lane 8) but not when 16ng is added (lane 9). There is less probe in the major retarded complex in lanes 6, 7 and 8 than in lanes 1, 2 and 3 respectively suggesting that the mutant is slightly more efficient at competing for AEF-1 binding than the wild type 5'UTR.
Figure 4.9 Gel retardation with titrations of wild-type and 143C144G competitors. Gel retardation assays were carried out as in figure 4.8. Lane 1 is a control with no extract. Lane 2 - no competitor. Lanes 3-7 - wild type 5'UTR competitor. Lane 3 - 2ng, lane 4 - 4ng, lane 5 - 8ng, lane 6 - 16ng, lane 7 - 32ng. Lanes 8-12 - 5'UTR with 143C144G mutation as competitor. Lane 8 - 50ng, lane 9 - 100ng, lane 10 - 200ng, lane 11 - 400ng, lane 12 - 600ng
Figure 4.10 - Gel retardation with titrations of wild-type and 150G151C competitors. Gel retardation assays were carried out as in figure 4.8. Lane 1 is a control with no extract. Lane 2 - no competitor. Lanes 3-7 - wild type 5'UTR competitor. Lane 3 - 2ng, lane 4 - 4ng, lane 5 - 8ng, lane 6 - 16ng, lane 7 - 32ng. Lanes 8-11 - 5'UTR with 150G151C mutation as competitor. Lane 8 - 4ng, lane 9 - 8ng, lane 10 - 16ng, lane 11 - 32ng.
Figure 4.11 Gel retardations with titrations of wild-type and 136T137C competitors. Gel retardation assays were carried out as in figure 4.8. Lanes 1-5 - wild type 5'UTR competitor. Lane 1 - 2ng, lane 2 - 4ng, lane 3 - 8ng, lane 4 - 16ng, lane 5 - 32ng. Lanes 6-10 - 5'UTR with 136T137C mutation as competitor. Lane 6 - 2ng, lane 7 - 4ng, lane 8 - 8ng, lane 9 - 16ng, lane 10 - 32ng
4.3 Discussion

This chapter has described the creation and characterisation of several mutant versions of the I factor 5'UTR with alterations in the vicinity of site 1. These mutations have various effects on binding to AEF-1 from ovarian nuclear extract. Altering base pairs 135 and 136 from CA to TC causes a slight increase in binding. Altering base pairs 137 and 138 or bases 143 and 144 reduces binding around 50 fold. Changing base pairs 150 and 151 reduces binding approximately two fold. Altering the four bases 137, 138, 143 and 144 abolishes all detectable binding.

4.3.1 Allignment of published AEF-1 binding sites with site 1

Udomkit et al., 1996 defined the protein binding site at site 1 by DNAse I footprinting. They found that the protected area covered base pairs 137 to 150 with the sequence ACAAAAACAACAAT. This sequence can be aligned with published AEF-1 binding sites as shown in Figure 4.12 (alignment A). A consensus binding site derived by An and Wensink (1995a) is also shown.

It is difficult to define the precise sequence requirements for AEF-1 binding due to the A and C-rich nature of the binding site. In fact there are alternative alignments of site 1 that match the consensus as well or better than alignment A, shown as alignments B and C in Figure 4.12. Alignment A has 9 matches with the consensus, alignment B has 10 and alignment C has 9. However the footprinting results (Udomkit et al, 1996) and studies of AEF-1 binding to the FBE-1 and Adh regulatory elements (An and Wensink, 1995a, Falb and Maniatis, 1992a, Ren and Maniatis, 1998) suggest that alignment A is correct. Presumably the affinity of binding is affected by the bases at the degenerate positions in the consensus or by flanking sequence.
1. Consensus

2. melanogaster Adh (AAE)

3. mulleri Adh

4. melanogaster FBE1

5. melanogaster Adh (pie)

6. Human Adh

7. Rat Adh

8. Site 1 - Alignment A

9. Site 1 - Alignment B

10. Site 1 - Alignment C

11. 137G138G - Alignment A

12. 143C144G - Alignment A

13. 103-114

C C
CACAACNA A
A A
CAGCAACAAACACGATC
ACCACAACAAAATAAAC
TGCAACAACCTACAATGTT
AAACCAACCTAAACG
TCAACACGCAACAAACAA
CCCCAGCAAAATAAACATCT
ACAAACAAAAACAAAACAT
ACAAACAAAAACAAAACAT
ACAAACAAAAACAAAACAT
ACAAACAAAAACAAAACAT
ACAAACAAAAACAAAACAT
ATCAAACAAAAACAAAACAT

The results obtained with mutant versions of the 5'UTR support alignment A. Mutagenesis of positions 137 and 138 or positions 143 and 144 change consensus bases in this alignment (Figure 4.12) and cause a substantial reduction in binding. It is possible that the weak binding observed with these mutants represents AEF-1 binding to an alternative binding site. In this context it may be significant that mutating positions 137 and 138 does not affect the consensus bases in alignment C while mutating 143 and 144 has no effect on the consensus bases of alignment B. It is possible that AEF-1 binds to these sites with a lower affinity than the site represented by alignment A. Where the four bases 137, 138, 143 and 144 are mutated there is no detectable binding. There is a sequence upstream of site 1 from positions 103-114 that matches the consensus in 9 consecutive positions (Figure 4.12). Apparently this is not sufficient for AEF-1 binding.

Experiments aimed at determining the effects of the mutations described in this chapter on on transcription directed by the I factor 5'UTR in Drosophila ovaries are described in Chapter 5.
Chapter 5 - Effect of mutagenesis of site 1 on the activity of the /factor enhancer
Effect of mutagenesis of site 1 on the activity of the \textit{I} factor enhancer

\subsection*{5.1 Introduction}

As described in Chapter 4, mutagenesis of site 1 affects the binding of \textit{I} factor 5'UTR by AEF-1 \textit{in vitro}. Mutations are available that reduce, abolish or have little effect on binding. This chapter describes experiments designed to establish the effect of such mutations on expression from the \textit{I} factor promoter in \textit{Drosophila} ovary and carcass tissues with the aim of elucidating the role of AEF-1 in \textit{I} factor regulation.

\subsection*{5.2 Results}

\subsubsection*{5.2.1 \textit{P} element transformation vectors containing mutant \textit{I-CAT} reporter genes}

Five mutant versions of the \textit{I} factor 5'UTR were selected for investigation of enhancer activity (see Section 4.2.2). In addition, control experiments were carried out using the wild-type 5'UTR and the 5'UTR with a deletion at site 1 described by Udomkit \textit{et al.} (1996). \textit{P} element-mediated transformation vectors containing a \textit{CAT} reporter gene transcribed from the \textit{I} factor promoter and under the control of \textit{I} factor regulatory sequences were constructed in a two stage process based on that of McLean \textit{et al.} (1993, see Figure 5.1). In each of the mutants the 5'UTR is flanked by recognition sites for \textit{XbaI} and \textit{BamHI}. Using these enzymes a 193bp fragment was excised and inserted into pCAT.1. The resulting plasmid was cut with \textit{PstI}, liberating a 1.8kb fragment containing the mutant 5'UTR, the \textit{CAT} open reading frame and SV40 sequences corresponding to the small \textit{t} intron and the polyadenylation signal which ensures efficient 3' processing of the transcripts (Thummel \textit{et al.}, 1988). This fragment was inserted into \textit{PstI}-digested pW8 and a clone selected in which the orientation of the insert was such that transcription of \textit{CAT} was in the opposite direction from that of the \textit{white} gene and also the \textit{P} element promoter which is present in the \textit{P} element 5' end sequence. This is to avoid the possibility of transcription from the \textit{I} factor promoter being affected by transcription
Figure 5.1 Construction of reporter plasmids for germ line transformation containing mutant versions of the \( l \) factor 5'UTR. pCAT.1 is described in McLean et al. (1993) and pW8 in Klemenz et al. (1987). Genes are shown by the open arrows pointing in the direction of transcription, mutant \( l \) factor 5'UTR sequence by the crossed box, \( P \) element ends by the filled black arrows, \( hsp70 \) promoter sequence by the diagonally striped box and SV40 sequences by the open box. In the plasmid names, M stands for the name of the mutant. MCS = multiple cloning site.
from the other promoters. The construct with site 1 deleted was made by digestion of p186Δ.T1 (Udomkit et al., 1996) with PstI and ligation of the resulting fragment into pW8 as described above. The wild type control was plasmid p186.W8 which was constructed by Carol McLean by a similar method (McLean et al., 1993).

5.2.2 Drosophila lines containing mutant l-CAT reporter constructs

The seven reporter gene plasmids, representing five mutants described in Chapter 4 as well as site 1 deleted and the wild-type 5'UTR, were used in \( P \) element-mediated transformation of the reactive strain \( JA \). Transformants were obtained from injections of each of the seven constructs. Southern blotting was carried out on restriction digests of genomic DNA from each transformant line to establish which contain independent insertions which contain a single copy of the transgene. For some lines the chromosome on which the transgene was inserted was determined. Details of transgenic flies are given in the appendix.

5.2.3 CAT expression in flies containing wild-type and site 1 deleted l-CAT reporter constructs

Male flies containing reporter constructs were crossed to \( JA \) virgin females to produce progeny heterozygous for the transgene. Newly-eclosed flies were aged for five days before females were dissected and extracts of ovary and carcass tissue prepared for measurement of CAT levels by enzyme linked immunosorbent assay (ELISA). Except where stated, two measurements of CAT concentration were made on separate occasions for each extract. Overall protein concentrations were measured by Bradford assay.

Previous studies have revealed that expression directed by the wild-type 5'UTR is greater in ovaries than in carcass tissue (McLean et al., 1993, Udomkit et al., 1996). The amount of CAT enzyme per microgram of total protein for the three lines containing the wild-type 5'UTR are shown in Table 5.1. The results show from 10 to 74 fold higher expression in ovaries than in carcass which is consistent with the earlier studies. Udomkit et al. (1996) measured five homozygous lines and found ovary/carcass ratios ranging from 5:7:1 to 37:1.
The results of Udomkit et al. (1996) suggested that deletion of site 1 caused a dramatic reduction in expression and also affected the tissue-specificity, giving slightly greater levels of CAT in carcass tissue than in ovaries. The results for the three site 1 deleted lines are shown in Table 5.1. Here there is no dramatic reduction in expression on deletion of site 1 with the figures for ovary expression in strains 471 and 472 not being significantly lower than with the wild-type construct. Strain 469 did give lower expression than the wild type strains but even here there is greater expression in ovary than in carcass tissue with a ratio of nearly 3:1. The lower expression in ovaries in this line could be due to a position effect resulting from the site of the transgene insertion.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Strain</th>
<th>Ovary</th>
<th>Carcass</th>
<th>O/C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>466</td>
<td>4.72±0.17</td>
<td>0.256±0.014</td>
<td>18.5</td>
</tr>
<tr>
<td></td>
<td>467</td>
<td>5.63±0.70</td>
<td>0.075±0.0018</td>
<td>74.6</td>
</tr>
<tr>
<td></td>
<td>468</td>
<td>2.12±0.09</td>
<td>0.19±0.038</td>
<td>10.9</td>
</tr>
<tr>
<td>Site 1 deleted</td>
<td>469</td>
<td>0.66±0.23</td>
<td>0.23±0.031</td>
<td>2.86</td>
</tr>
<tr>
<td></td>
<td>471</td>
<td>3.12±0.59</td>
<td>0.22±0.030</td>
<td>13.6</td>
</tr>
<tr>
<td></td>
<td>472</td>
<td>6.07±2.03</td>
<td>0.65±0.141</td>
<td>9.27</td>
</tr>
</tbody>
</table>

Table 5.1 Expression from wild type and site 1 deleted I-CAT reporter constructs. CAT levels were measured in five day old female progeny of males carrying I-CAT reporter constructs and JA virgin females. Results are expressed as pg CAT per µg total protein. The results for ovary and carcass from strains 466, 467 and 468 and the 469 carcass result are the means and standard errors on two measurements of the same extract. Results for both 471 and 472 are from four measurements, two on each of two extracts made from the progeny of crosses carried out separately. The pooled means and standard errors are given. The 469 ovary result is a single measurement.

5.2.4 Expression from I-CAT reporter constructs containing point mutations

The effect of the point mutations described in Chapter 4 on expression from I-CAT reporter constructs was investigated. Males carrying transgene constructs were crossed to JA virgin females as for the wild-type and site 1 deleted lines described above and CAT levels were measured in extracts of ovary and carcass tissue of the female progeny. At least two lines were tested carrying each of the mutant constructs with the exception of 135T136C. The results are shown in Table 5.2 and the
Table 5.2 Effect of point mutations in site 1 on expression from the I-CAT reporter. Measurements were carried out on ovary and carcass extracts of the female progeny of males carrying reporter constructs and JA virgin females. Results are expressed as pg CAT per μg total protein. Results are the means and standard errors of two measurements on the same extract with the exception of the result for 493 carcass which is a single measurement.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Strain</th>
<th>Ovary</th>
<th>Carcass</th>
<th>O/C</th>
</tr>
</thead>
<tbody>
<tr>
<td>137G138G</td>
<td>475</td>
<td>8·02±2·09</td>
<td>0·285±0·006</td>
<td>28·1</td>
</tr>
<tr>
<td></td>
<td>474</td>
<td>10·2±1·1</td>
<td>1·09±0·22</td>
<td>9·36</td>
</tr>
<tr>
<td>143C144G</td>
<td>480</td>
<td>16·2±1·1</td>
<td>0·388±0·017</td>
<td>41·7</td>
</tr>
<tr>
<td></td>
<td>481</td>
<td>2·49±0·31</td>
<td>0·202±0·086</td>
<td>12·3</td>
</tr>
<tr>
<td>150G151C</td>
<td>482</td>
<td>4·98±0·18</td>
<td>0·427±0·037</td>
<td>11·7</td>
</tr>
<tr>
<td></td>
<td>483</td>
<td>6·74±0·03</td>
<td>0·720±0·136</td>
<td>9·36</td>
</tr>
<tr>
<td></td>
<td>484</td>
<td>8·33±1·13</td>
<td>0·646±0·130</td>
<td>12·9</td>
</tr>
<tr>
<td></td>
<td>485</td>
<td>2·57±0·05</td>
<td>0·499±0·004</td>
<td>5·15</td>
</tr>
<tr>
<td>137G138G143C144G</td>
<td>492</td>
<td>8·00±0·07</td>
<td>0·341±0·049</td>
<td>23·5</td>
</tr>
<tr>
<td></td>
<td>493</td>
<td>1·08±0·29</td>
<td>0·0322</td>
<td>33·5</td>
</tr>
</tbody>
</table>

Figure 5.2 Effect of mutagenesis of site 1 on expression from the I-CAT reporter in ovaries. The results shown in tables 5.1 and 5.2 are plotted.
measurements from ovary extracts are shown graphically, along with those for the wild type and site 1 deleted I-CAT strains (see Table 5.1), in Figure 5.2. Experiments in which flies carrying the 135T136C construct were crossed to \( W^X \) will be described below.

As with the experiment using the site 1 deleted constructs these results provide no convincing evidence that expression is affected by mutations that reduce AEF-1 binding. CAT expression varies in the different strains, presumably as a result of position effects, but no consistent pattern emerges relating the level of CAT to the mutations present in the reporter constructs.

The results presented here contradict the conclusion of Udomkit et al. (1996) that site 1 is an important component of the \( I \) factor enhancer. There were several differences between the methods employed by McLean et al. (1993) and Udomkit et al. (1996) and those used in the present study, which might potentially explain the different results. CAT levels were measured by an enzyme activity assay as opposed to the immunological detection method used here. In the previous studies expression was measured in homozygous flies while here heterozygotes were tested, and the reactive strain into which the reporter constructs were introduced was \( W^X \) not \( JA \) as in the present experiments. The possible effects of these differences on the results have been investigated.

**5.2.5 CAT expression in existing site 1 deleted I-CAT reporter strains**

If a large proportion of the total CAT protein in an extract is in an inactive form, unable to catalyze the acetylation of chloramphenicol, then the results of the enzyme activity assay might differ from those of the ELISA, which measures the total protein level. To establish if there is a significant difference between the results obtained from site 1 deleted strains using the different methods of measurement, extracts were made from the strains 230, 231, 232 and 234 which were tested by Udomkit et al. (1996) and CAT levels measured by the ELISA method (see Table 5.3a). These strains (referred to as 9, 11, 13 and 17 in Udomkit et al., 1996) contain the \( I \) factor 5'UTR with site 1 deleted upstream of the CAT reporter. Activity assays of CAT gave ovary:carcass ratios ranging from 0.34 to 0.61 (Udomkit et al., 1996, see Table 5.3b). The ELISA measurements gave similar results for strains 231, 232
and 234 (Table 5.3a), with lower CAT expression in ovaries than in carcass, suggesting that measurements of CAT expression in the same fly stocks are consistent between the two assay methods. Strain 230 gave higher expression in ovaries than in carcass contrary to the results of Udomkit et al. There is evidence, which will be presented below, that this strain may have been contaminated by flies carrying a wild-type I-CAT reporter construct.

A.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Ovary (pg/μg)</th>
<th>Carcass (pg/μg)</th>
<th>O/C</th>
</tr>
</thead>
<tbody>
<tr>
<td>230</td>
<td>0.422±0.066</td>
<td>0.111±0.018</td>
<td>3.8</td>
</tr>
<tr>
<td>231</td>
<td>0.0699±0.0074</td>
<td>0.138±0.021</td>
<td>0.51</td>
</tr>
<tr>
<td>232</td>
<td>&lt;0.0148</td>
<td>0.0235</td>
<td>&lt;0.63</td>
</tr>
<tr>
<td>234</td>
<td>0.0710±0.0152</td>
<td>0.140±0.018</td>
<td>0.51</td>
</tr>
</tbody>
</table>

B.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Ovary (pmol/min/mg)</th>
<th>Carcass (pmol/min/mg)</th>
<th>O/C</th>
</tr>
</thead>
<tbody>
<tr>
<td>230</td>
<td>3.2</td>
<td>9.4</td>
<td>0.34</td>
</tr>
<tr>
<td>231</td>
<td>1.8</td>
<td>5.0</td>
<td>0.36</td>
</tr>
<tr>
<td>232</td>
<td>2.9</td>
<td>7.7</td>
<td>0.38</td>
</tr>
<tr>
<td>234</td>
<td>0.58</td>
<td>0.95</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Table 5.3 Measurements of CAT expression in strains 230, 231, 232 and 234 by ELISA (A) and CAT activity assay (B). The units are pg CAT per microgram total protein (ELISA) and pmol chloramphenicol acetylated per minute per miligram total protein (activity assay). The 232 ovary ELISA gave a lower signal than 11.2pg of purified CAT when 760 μg of extract was used. The figure for 232 carcass is a single measurement. All other ELISA measurements are the means with standard errors of two measurements on the same extract. Measurements by the activity assay (B) are taken from Udomkit et al., 1996.

5.2.6 Measurements of I-CAT reporter gene expression in homozygous or heterozygous flies

A further difference between the methods used in this study and those of the previous studies is that measurements were made of CAT expression in heterozygous, rather than homozygous flies. There is evidence that the number of copies of the I factor 5'UTR can influence the level of expression from an I-CAT reporter (Chaboissier et al., 1998) with increasing numbers of copies resulting in a reduction in expression in ovaries. Table 5.4 shows measurements of expression in...
homozygotes from strains 468, 471 and 472. The results follow the same pattern as those from heterozygous flies (compare Table 5.1) with the highest expression in 472 followed by 471 then 468. In strain 471 expression in both ovary and carcass is approximately double that in the heterozygote as might be expected as the gene dosage has been doubled. In strains 468 and 472 expression in ovaries is increased more than twofold while that in carcass is increased to a lesser extent, resulting in an increase in the ovary:carcass ratios. There is therefore some evidence that the results from different strains are affected differentially by measurement of the homozygote rather than the heterozygote. However there is again no evidence of a dramatic reduction in expression on deletion of site 1 so measurement in heterozygotes cannot account for the different results of this study and Udomkit et al.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Strain</th>
<th>Ovary</th>
<th>Carcass</th>
<th>O/C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>468</td>
<td>5.94±0.04</td>
<td>0.25±0.014</td>
<td>23.1</td>
</tr>
<tr>
<td>Site 1 deleted</td>
<td>471</td>
<td>6.50±1.53</td>
<td>0.466±0.039</td>
<td>13.9</td>
</tr>
<tr>
<td></td>
<td>472</td>
<td>21.9±1.9</td>
<td>0.812±0.349</td>
<td>27.0</td>
</tr>
</tbody>
</table>

Table 5.4 CAT levels in homozygous females containing wild type and site 1 deleted I-CAT reporter constructs. Results are expressed as pg CAT per μg total protein. The results for strains 468 and 472 are the means and standard errors on two measurements of the same extract. Results for 471 are the pooled means and standard errors from four measurements, two on each of two extracts made from the progeny of crosses carried out separately.

### 5.2.7 Effect of genetic background on expression from I-CAT reporter genes

In McLean et al. (1993) and Udomkit et al. (1996) I-CAT reporter constructs were introduced into $W^X$ not $JA$ as in the present study. This difference in the strain background may affect CAT expression and the ratio of CAT levels in ovaries and carcass. To investigate this possibility, males from several of the strains carrying I-CAT reporter constructs were mated to $W^X$ virgin females and CAT levels in the female progeny measured as before. The strains tested include examples carrying the $135T136C$ mutation. The results are shown in Table 5.5 and ovary results are compared graphically in Figure 5.3. As with the results from crosses to JA there is no clear relationship between the mutations and the level of expression. The $135T136C$ mutation has no clear effect on expression.
<table>
<thead>
<tr>
<th>Construct</th>
<th>Strain</th>
<th>Ovary</th>
<th>Carcass</th>
<th>O/C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>467</td>
<td>0.739±0.148</td>
<td>0.045±0.0318</td>
<td>16.4</td>
</tr>
<tr>
<td>Site 1 deleted</td>
<td>471</td>
<td>0.401±0.111</td>
<td>0.137±0.008</td>
<td>2.92</td>
</tr>
<tr>
<td></td>
<td>472</td>
<td>0.818±0.013</td>
<td>0.490±0.007</td>
<td>1.67</td>
</tr>
<tr>
<td>137G138G</td>
<td>474</td>
<td>0.197±0.029</td>
<td>0.763±0.007</td>
<td>0.259</td>
</tr>
<tr>
<td>143C144G</td>
<td>478</td>
<td>1.44±0.280</td>
<td>0.396±0.207</td>
<td>3.63</td>
</tr>
<tr>
<td></td>
<td>479</td>
<td>0.589±0.059</td>
<td>0.203±0.006</td>
<td>2.90</td>
</tr>
<tr>
<td>150G151C</td>
<td>494</td>
<td>1.17±0.17</td>
<td>0.264±0.028</td>
<td>4.43</td>
</tr>
<tr>
<td></td>
<td>483</td>
<td>0.681±0.209</td>
<td>0.912±0.184</td>
<td>0.747</td>
</tr>
<tr>
<td>135T136C</td>
<td>486</td>
<td>0.721±0.097</td>
<td>0.834±0.202</td>
<td>0.747</td>
</tr>
<tr>
<td></td>
<td>489</td>
<td>0.125±0.063</td>
<td>0.572±0.187</td>
<td>0.219</td>
</tr>
</tbody>
</table>

Table 5.5 Expression from mutant *I-CAT* reporter constructs in the hybrid *W^K/JA* genetic background. Measurements were carried out on ovary and carcass extracts of the female progeny of males carrying reporter constructs and *W^K* virgin females. Results are expressed as pg CAT per µg total protein. The results are the means and standard errors of four measurements, two on each of two extracts made from the progeny of crosses carried out separately, except for the results from strains 478, 483 and 489, each of which represent two measurements on the same extract.

![Figure 5.3 Expression in ovaries from *I-CAT* reporter constructs in the hybrid *W^K/JA* genetic background. The results in table 5.5 are plotted.](image-url)
Six of the strains have been tested by crossing to both JA and \( W^K \); 467, 471, 472, 474, 482 and 483. The results are compared in Table 5.6. In all cases crossing to \( W^K \) instead of JA resulted in a substantial reduction in expression in ovaries, indicating that there are differences between the two strains that affect expression. \( W^K \) has a lower reactivity than JA, that is the progeny of dysgenic crosses are more fertile (Marie-Christine Chaboissier, personal communication), presumably reflecting a lower \( I \) factor transposition frequency. The results are consistent with earlier studies that showed that the reactivity level affects the expression of transgenes transcribed from the \( I \) factor promoter (Lachaume and Pinon, 1993).

<table>
<thead>
<tr>
<th>Construct</th>
<th>Strain</th>
<th>( JA )</th>
<th>( JAW^K ) hybrid</th>
<th>JA/hybrid</th>
</tr>
</thead>
<tbody>
<tr>
<td>467</td>
<td>5.63±0.70</td>
<td>0.739±0.148</td>
<td>7.61</td>
<td></td>
</tr>
<tr>
<td>471</td>
<td>3.12±0.59</td>
<td>0.401±0.011</td>
<td>7.79</td>
<td></td>
</tr>
<tr>
<td>472</td>
<td>6.07±2.03</td>
<td>0.818±0.013</td>
<td>7.41</td>
<td></td>
</tr>
<tr>
<td>474</td>
<td>10.2±1.1</td>
<td>0.197±0.029</td>
<td>51.9</td>
<td></td>
</tr>
<tr>
<td>474</td>
<td>10.2±1.1</td>
<td>0.197±0.029</td>
<td>51.9</td>
<td></td>
</tr>
<tr>
<td>482</td>
<td>6.74±0.03</td>
<td>0.681±0.209</td>
<td>9.88</td>
<td></td>
</tr>
<tr>
<td>483</td>
<td>6.74±0.03</td>
<td>0.681±0.209</td>
<td>9.88</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.6 Comparison between expression levels in pure JA and JAW\(^K\) hybrid genetic backgrounds. The results are from Tables 5.1, 5.2 and 5.5

The amount by which expression was reduced in ovaries varied between the different transgene insertions. For 474 there is a 52 fold difference while for the other strains the difference ranges from five fold for strain 482 to 10 fold for strain 474 (see Table 5.6). The carcass results also show an effect of genetic background on expression. Most strains show a reduction in expression in the hybrid background compared to the JA background ranging from 1.3 to 2.4 fold. The transgene of strain 482 shows 1.3 fold greater expression in the hybrid background than in the pure JA.

101
background. Because the effect of strain background is greater on expression in ovaries, the ovary carcass ratios are lower in the hybrid background than in the pure JA background.

The results suggest that the different transgene insertions are affected to different extents by the change in genetic background. However these differences do not appear to be related to the sequence of the transgene and are most likely to be the result of differences in the location of the transgenes in the genome. The data present no evidence that the wild type and site 1 deleted constructs are affected by the strain background in different ways which might have explained the discrepancy between the site 1 deleted results of Udomkit et al. and those obtained here. A comparison of the ovary results for the wild type and site 1 deleted lines shows that the relative values are not much different in the hybrid background with 472 still giving the highest expression in ovaries followed by 467 then 471.

Udomkit et al. measured expression in homozygotes in a pure \( W^X \) background. It remains possible that there could be a difference in the relative values between the different constructs in that case as here only the hybrid \( JAW^X \) background was tested. To investigate the expression of a transgene tested by Udomkit et al. in a hybrid genetic background, males from strain 232 were crossed with JA virgin females and CAT levels in the progeny measured. In both ovary and carcass the level of CAT was too low to be detected which might suggest that the genetic background cannot explain the low expression in this line. However there is evidence, discussed below, that strain 232 may not actually contain the site 1 deleted transgene so this experiment should be interpreted with caution.

5.2.8 Characterization of site 1 deleted I-CAT reporter constructs in Drosophila strains

A possible explanation for the different results obtained with site 1 deleted constructs between this study and Udomkit et al. (1996) is that the plasmids used in the transformation of Drosophila may not have been identical. It is possible that one of them carried a further mutation or rearrangement that affects expression, in addition to the deletion of site 1. Both plasmids were made independently by the same method, the insertion of a \( PstI \) fragment of p186Δ.T1 into pW8 (see Section
5.2.1) but a mutation could have occurred during replication of the plasmids in *E. coli*. The plasmid pW8.1-Δ186 used by Udomkit *et al.* was not available for characterization. Southern blotting assays were therefore carried out on genomic DNA isolated from transgenic flies to investigate the structure of the transgene.

Genomic DNA was isolated from the eight strains carrying the site 1 deleted reporter construct; 469, 471 and 472 from this study and 230, 231, 232, 233 and 234 created by Udomkit *et al.* (1996, the strains are referred to as 9, 11, 13, 14 and 17 respectively in that paper). In addition DNA was prepared from strain 137 (McLéan *et al.*, 1993) which contains the wild type 5'UTR upstream of the CAT reporter, and JA for use as a negative control. Figure 5.4a shows a map of the expected *P* element insertion. Genomic DNA was cleaved with *EcoRI* which should cut twice in the insert DNA to give an internal fragment of 1.5kb. After electrophoresis and blotting the filter was incubated with a probe made from a 1.7kb *PstI* fragment of pCAT.1 (see Figure 5.1) containing the CAT and SV40 sequences present in the transgene. This should hybridise with the expected 1.5kb product of *EcoRI* digestion. The result is shown in Figure 5.5. The expected 1.5kb fragment is seen in the digests of DNA from strains 137, 469, 471, 472 and 230 (Figure 5.5 lanes 2-6) but not in the DNA from strains 231, 232, 233 and 234. Strains 231, 233 and 234 do show bands that hybridise with the probe but they are not of a uniform size and are all larger than 1.5kb (lanes 7, 9 and 10).

One possible explanation for the results seen with strains 231, 233 and 234 is that in the creation of the pW8.Δ.CAT plasmid used by Udomkit *et al.*, the *PstI* fragment of p186Δ.T1 was inserted into pW8 in the opposite orientation from that shown in Figure 5.1. This would give the transgene shown in Figure 5.4b, which on digestion with *EcoRI* would produce a hybridising fragment extending across the white gene and into the flanking genomic sequence. Digestion of the transgene with *PstI* and *EcoRI* should give a 1.4kb fragment that hybridises with the probe regardless of the orientation of the *PstI* fragment. This was tested in the experiment shown in Figure 5.6. All of the samples from transgenic flies gave a band around 1.4kb with the exception of the 232 sample, which again gave no detectable hybridization.
Figure 5.4 Restriction maps of I-CAT reporter transgene insertions. The dotted line represents flanking genomic DNA. The distances shown in base pairs assume that the 5'UTR does not have site 1 deleted. A. Map of expected insertion following injection of pW8.M.CAT plasmids (see figure 5.1) B. Map of expected insertion following injection of a plasmid in which the PstI fragment of pM.T1 has been inserted in the opposite orientation from that shown in figure 5.1
Figure 5.5 Southern blot of genomic DNA from I-CAT reporter strains digested with EcoRI. Samples were run in a 1% agarose gel. The probe was the 1.7kb PstI fragment of pCAT.1. Lines with numbers to the left of the blot represent the mobility of λ.HindIII DNA size markers. Lane 1 - JA, lane 2 - 137, lane 3 - 469, lane 4 - 471, lane 5 - 472, lane 6 - 230, lane 7 - 231, lane 8 - 232, lane 9 - 233, lane 10 - 234.
Figure 5.6 Southern blot of genomic DNA from *I-CAT* reporter strains digested with *EcoRl* and *PstI*. Samples were run in a 1% agarose gel. The probe was the 1.7kb *PstI* fragment of pCAT.1. Lines with numbers to the left of the blot represent the mobility of λ*HindIII* DNA size markers. Lane 1 – JA, lane 2 – 137, lane 3 - 469, lane 4 - 471, lane 5 - 472, lane 6 - 230, lane 7 - 231, lane 8 - 232, lane 9 - 233, lane 10 - 234
To test whether the 5'UTR is present upstream of the \textit{CAT} reporter in the transgenes, PCR was carried out using primers 664C and CATb. This should amplify a fragment extending from the start of the \textit{I} factor sequence into the \textit{CAT} open reading frame. The size of the predicted product is 432bp if site 1 is present and 413bp if the template has the site 1 deletion. Figure 5.7 shows the products of PCR using these primers and each of the genomic DNA samples as a template. Apart from strains \textit{JA} and 232 all the reactions gave a strong band with a mobility between the 310bp and the 603bp markers. As expected the product amplified from 137, has a lower mobility than the products amplified from the site 1 deleted strains. The only exception to this is strain 230 which appears to give a band of the same mobility as 137. Taken together the results of southern blotting, PCR and CAT assays (Section 5.2.5) suggest that strain 230 may have been contaminated by flies carrying the wild-type construct present in 137 some time after the original CAT assay measurements of Udomkit \textit{et al.} (1996). Strain 232 gave no amplification product and also gave no hybridising bands on the Southern blot. This strain gave very low levels of CAT expression when measured by ELISA (see Section 5.2.5). The evidence suggests that 232, as well as 230, may have been contaminated by flies with a different genotype, in the latter case by flies lacking the \textit{I-CAT} reporter.

The PCR results for the other site 1 deleted strains created by Udomkit \textit{et al.} (1996) suggest that they do contain the \textit{CAT} reporter gene downstream of the \textit{I} factor 5'UTR with site 1 deleted. The evidence is consistent with the view that the injected plasmid had the \textit{PstI} fragment from p186Δ.T1 in the opposite orientation from that shown in Figure 5.1. Strain 137, and presumably the other wild type \textit{I-CAT} strains tested by Udomkit \textit{et al.} and McLean \textit{et al.}, as well as all of the strains created in this study, did have transgenes with the structure shown in Figure 5.1. This may provide an explanation for the difference in CAT expression. It is also possible that there may be differences in the sequence of the transgenes that were not detected by the southern bloting and PCR experiments described here. To resolve that question it would be necessary to clone and sequence the transgenes.
Figure 5.7 Agarose gel of PCR products amplified from genomic DNA using primers 664C and CATb. Samples were run in a 2% agarose gel. Lane 1 - φX174 Haelll size markers, lanes 2-11 PCR products. Template DNA: lane 2 – JA, lane 3 – 137, lane 4 – 469, lane 5 – 471, lane 6 – 472, lane 7 – 230, lane 8 – 231, lane 9 – 232, lane 10 – 233, lane 11 – 234.
5.3 Discussion

The experiments described in this chapter fail to provide any evidence that site 1 is important in the function of the \( I \) factor enhancer. Deletion of site 1, and other mutations that affect AEF-1 binding, did not have a dramatic effect on expression directed by the \( I \) factor regulatory sequences. These results contradict those of Udomkit et al. (1996) who found that expression was reduced around 30 fold in constructs having a deletion of site 1. There are a number of plausible explanations for the differences in the results.

For each construct only a small number of independent insertions were tested. It is therefore possible that variation in the results could have masked real differences in expression between the different transgene constructs. There are several possible sources of variation that could affect the measurements of CAT expression, the most important of which are probably position effects. The expression of a transgene introduced by \( P \) element-mediated transformation is affected by its location in the genome as a result of local differences in chromatin structure, and in some cases the proximity of enhancer or repressor elements that regulate transcription (Spradling and Rubin, 1983, Levis et al., 1985). Other sources of variation in the experiments include sampling errors, as presumably not all of the flies of a given genotype express CAT to the same level, and errors in the process of measurement which could arise during pipetting of samples and taking spectrophotometer readings. It seems unlikely however that variation in the results could explain the lack of a dramatic reduction in expression on deletion of site 1. All three of the site 1 deleted lines would have to have the transgene inserted in a position conferring strong ovarian expression. Other studies of expression from the 5'UTR have found that different insertions of the same transgene in reactive flies can have a different overall level of expression but the overall tissue-specificity and developmental pattern of expression are similar (Lachaume et al., 1992, McLean et al., 1993, Udomkit et al., 1996). It is likely however that any more subtle differences in expression levels between the different mutant constructs may not have been detected due to the low numbers of transgenic lines tested.
There was significantly lower CAT expression in ovaries from several transgenes in a \( JA^{W^k} \) hybrid genetic background compared with a pure \( JA \) background. As discussed in Section 5.2.7 the observed differences do not account for the failure to observe a reduction on deletion of site 1, but there remains the possibility that the results may have been different in a pure \( W^k \) background. However there is another way in which the strain background could have affected the results. If expression of CAT in ovaries is detrimental to the fertility or viability of the flies, it may be impossible to obtain transformants showing very high levels of expression or to establish homozygous lines from such flies. In that case the lines obtained in this study could only have the transgenes inserted at small subset of the possible insertion sites, those at which the local chromatin structure results in a low level of expression. This would tend to make the results more homogeneous and minimize any differences between the site 1 deleted and the wild type lines. If this is the case then it would perhaps be a greater problem when \( JA \) is used as the recipient for injection than when \( W^k \) is used because expression is greater in the \( JA \) background. There was some evidence during this study that the size of the ovaries is correlated with the level of CAT expression which might be expected if high CAT levels affect the fertility of the flies. To test this possibility the viability of flies expressing different levels of CAT could be measured by counting their progeny, to see if there is a correlation. The hypothesis predicts that expression from the white marker gene would be lower in flies carrying \( I\text{-CAT} \) reporter constructs than in control flies transformed with a \( P \) element carrying only the marker and this could be tested.

Probably the most likely explanation for the differences in the site 1 deleted results between this study and Udomkit et al. (1996) relates to the structure of the reporter gene constructs. The transgenes created by Udomkit et al. have a different structure from that described and from that of the wild-type \( I\text{-CAT} \) constructs which provided the positive control for enhanced germ line expression. It should therefore be considered doubtful that the reduced expression observed by Udomkit et al. was a consequence of deleting site 1.

110
PCR analysis of the transgenes of strains 231, 233 and 234 suggested that they do contain the I factor 5'UTR with site 1 deleted upstream of the CAT reporter gene. The results are consistent with a construct in which the I-CAT reporter cassette is oriented in the opposite direction with respect to the P element vector from that in the other transgenic strains that were investigated. It is not immediately obvious how this could affect expression from the I factor promoter but there are some possible mechanisms. In that orientation the P element promoter would initiate transcription upstream of and in the same direction as the I factor promoter. This could affect the initiation of transcription from the I factor promoter. Alternatively the upstream sequences could have some local effect on the chromatin structure that might affect initiation. It is also possible that in the orientation used in this study and with the wild-type 5'UTR in McLean et al. and Udomkit et al. there is a positive influence from surrounding sequences. The hsp70 promoter is in close proximity to the I factor promoter in this transgene and it is possible that this could enhance transcription, perhaps by altering the chromatin structure in the vicinity. However other constructs with different structures have also shown enhanced I factor expression in ovarian tissue (Lachaume et al., 1992, Udomkit et al., 1996, Seleme et al., 1999) suggesting that this is not simply an artifact of the particular construct used in the I-CAT experiments. This would favour an explanation based on inhibition of normal transcription in the orientation seen in strains 231, 233 and 234.

In conclusion, the results presented in this chapter suggest that site 1 is not required for the germ line specific activity of the I factor enhancer. Therefore it is doubtful whether AEF-1 is involved in I factor regulation.
Chapter 6 - Effect of varying AEF-1 levels in the female germ line on expression from the I factor promoter
Effect of varying AEF-1 levels in the female germ line on expression from the I factor promoter

6.1 Introduction

The evidence presented in Chapter 3 suggests that site 1, an element in the I factor 5' untranslated region, is bound by AEF-1 present in Drosophila ovaries. Udomkit et al. (1996) showed that a reporter construct in which site 1 was deleted gave a dramatically reduced level of expression, which suggested that either AEF-1 or a protein with an overlapping binding site is a strong activator of transcription. However the data presented in Chapter 5 has cast doubt on the interpretation of that experiment by showing that the transgene construct with site 1 deleted was rearranged with respect to the wild type construct used for comparison. Experiments using new constructs with site 1 deleted and other mutations in site 1 have failed to show any effect of disrupting AEF-1 binding on expression from the I factor promoter. It is therefore unclear if AEF-1 plays any role in I factor regulation.

If AEF-1 is involved in regulating the I factor enhancer then expression from the I-CAT reporter is likely to be sensitive to the concentration of AEF-1 protein in the germ line. This chapter describes an attempt to test this proposition by expressing sense and antisense AEF-1 RNA from a female germ line-specific promoter. The effect on expression from the I-CAT reporter was investigated.

6.2 Results

6.2.1 Constructs for expression of AEF-1 and antisense AEF-1 in the female germ line

The plasmid pW8.BBS.AEF1.SV2 (see Figure 6a) is designed to direct expression of an epitope-tagged AEF-1 protein in the Drosophila female germ line. The plasmid pW8.BBS.αAEF1.SV2 (Figure 6b) should cause expression of an antisense AEF-1 RNA in the same tissue. They were constructed as follows. An artificial promoter region (BBS), consisting of two copies of the B box element from the hsp26 gene upstream of the SGS3 promoter is transcribed specifically in the
Figure 6.1 Plasmids for expression of sense (A) and antisense (B) AEF-1 RNA in the female germ line. $P$ element ends are shown as filled black arrows, genes as open arrows pointing in the direction of translation of their open reading frames. Promoters are shown as shaded boxes with the direction of transcription of AEF-1 from the $SGS3$ promoter indicated by the arrows.
female germ line (Frank et al., 1992). The plasmid pBBS contains a 572bp BamHI + NheI fragment containing this region. This was inserted into the P element-mediated germ line transformation vector pW8, between the BamHI and XbaI sites, to create pW8.BBS. An XbaI + HindIII fragment derived from the plasmid pBS35 contains the AEF-1 open reading frame downstream of codons 408-439 of the human c-myc gene encoding a myc epitope tag. This fragment was blunt ended using klenow polymerase and inserted into HpaI-digested pW8.BBS. Clones were obtained with the fragment in both orientations, designated pW8.BBS.AEF1, with the coding strand of the AEF-1 open reading frame in the same orientation as the BBS promoter, and pW8.BBS.αAEF1, with the insert in the opposite orientation. The SV40 small t intron and late polyadenylation signals provide correct 3' processing of recombinant transcripts in Drosophila melanogaster (Thummel et al., 1988). A 1.6kb BglII + EcoRI fragment derived from the plasmid pSV2βglobin contains these sequences. This fragment was blunt-ended and inserted into XhoI-digested, blunt-ended, pW8.BBS.AEF1 and pW8.BBS.αAEF1 to obtain the plasmids pW8.BBS.AEF1.SV2 and pW8.BBS.αAEF1.SV2 (Figure 6.1). Clones with the insertion in the correct orientation were identified by DNA sequencing.

6.2.2 Measurement of CAT expression

The plasmids pW8.BBS.AEF1.SV2 and pW8.BBS.αAEF1.SV2 were used in P element-mediated germ line transformation of the reactive strain JA. Several homozygous transformant lines which are listed and described in the appendix were obtained using each plasmid. Lines containing single insertions of the transgene were identified by southern blotting of genomic DNA. Of these, six lines containing the sense construct and five containing the antisense construct were selected to test the effect of the transgene on I factor expression.

Strain 137 carries a reporter gene with the wild type I factor 5'UTR upstream of the CAT ORF in the W* strain background and gives approximately 20 fold greater CAT expression in ovaries than in female carcass (McLean et al., 1993, Udomkit et al., 1996). Virgin 137 females were crossed to males from the strains containing the BBS.AEF-1 transgenes to produce heterozygotes in which the reporter and the sense
or antisense AEF-1 expression constructs were present in the same individuals. Ovarian and carcass extracts were made from five day old female progeny and CAT expression measured by ELISA. Two measurements were taken from each extract. For each strain, with the exceptions of 443, 449, 457 and 460, the crosses were repeated and a second extract made and measured in the same way as the first. As a control, JA males were crossed to 137 virgin females and CAT levels measured in the same way, except that for one of the two extracts only a single measurement was taken. The results are shown in Table 6.1 and are displayed graphically in Figure 6.2. All of the experiments using lines carrying either the sense or the antisense construct gave lower expression of CAT than the control experiment with JA. The results were analysed statistically to assess the significance of this finding.

6.2.3 Statistical analysis

The aim of this experiment was to compare the results obtained with the two transgenes, sense and antisense, and the control. The mean level of expression obtained with each transgene can be calculated from the results for each of the individual strains shown in Table 6.1. These are 0.463 for the sense strains, 0.836 for the control and 0.434 for the antisense strains. These figures are subject to variation from several sources at different levels. There is variation between the different strains, which could be attributed to position effects that alter the level of expression of the sense or antisense AEF-1 RNA. There is also variation between the extracts made from the progeny of different crosses involving each strain. This could result from environmental effects on expression levels, from the sampling error involved in making extracts from a limited number of flies and also from errors in the measurement of the protein concentration of each extract. Finally there is some variation between the two different measurements carried out on each extract, which will result from errors in measurements of the volumes of extract and CAT standard solutions and the measurement of the absorbance in the ELISA.

An analysis of variance can be carried out to estimate the variation within and between the three groups representing the sense, antisense and control experiments (Table 6.2). The mean square within groups is an estimate of the variance associated
<table>
<thead>
<tr>
<th>Sense, antisense or control</th>
<th>Strain crossed to 137 females</th>
<th>Ovary</th>
<th>Carcass</th>
<th>O/C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sense</td>
<td>443</td>
<td>0.576±0.108</td>
<td>0.0218±0.0035</td>
<td>26.5</td>
</tr>
<tr>
<td></td>
<td>445</td>
<td>0.654±0.086</td>
<td>0.0237±0.0076</td>
<td>27.6</td>
</tr>
<tr>
<td></td>
<td>449</td>
<td>0.481±0.001</td>
<td>0.0319±0.0007</td>
<td>15.1</td>
</tr>
<tr>
<td></td>
<td>452</td>
<td>0.304±0.072</td>
<td>0.0203±0.0020</td>
<td>15.0</td>
</tr>
<tr>
<td></td>
<td>453</td>
<td>0.185±0.016</td>
<td>0.0271±0.0033</td>
<td>6.83</td>
</tr>
<tr>
<td></td>
<td>454</td>
<td>0.580±0.008</td>
<td>0.0253±0.0075</td>
<td>22.9</td>
</tr>
<tr>
<td>Control</td>
<td>JA</td>
<td>0.836±0.239</td>
<td>0.0251±0.0023</td>
<td>33.3</td>
</tr>
<tr>
<td>Antisense</td>
<td>456</td>
<td>0.571±0.020</td>
<td>0.0299±0.0040</td>
<td>19.1</td>
</tr>
<tr>
<td></td>
<td>457</td>
<td>0.349±0.012</td>
<td>0.0212±0.0001</td>
<td>16.5</td>
</tr>
<tr>
<td></td>
<td>458</td>
<td>0.594±0.058</td>
<td>0.308±0.0056</td>
<td>19.3</td>
</tr>
<tr>
<td></td>
<td>460</td>
<td>0.391±0.000</td>
<td>0.0281±0.0009</td>
<td>13.9</td>
</tr>
<tr>
<td></td>
<td>463</td>
<td>0.266±0.072</td>
<td>0.0211±0.0025</td>
<td>12.7</td>
</tr>
</tbody>
</table>

Table 6.1 Effect of AEF-1 sense and antisense germ line expression constructs on the I-CAT reporter. The strains shown were crossed to 137 virgin females and CAT levels measured in ovary and carcass extracts of the female progeny. Figures are given as pg CAT/µg total protein. With the exceptions stated in the text they are pooled means and standard errors for 4 measurements, two on each of two extracts made from the progeny of two crosses carried out separately.

Figure 6.2 Effect of sense and antisense AEF-1 germ line expression constructs on expression from the I-CAT reporter in ovaries. The results from table 6.1 are plotted.
with the group means. It can be used to estimate the standard error associated with
the differences between the group means and therefore the \( t \) statistic for each of the
pairwise comparisons between them (Table 6.3). With 9 degrees of freedom the
critical value of Student's two tailed \( t \)-distribution at the 5% level is 2.26. Therefore
for each of the comparisons, the null hypothesis that the two population means are
equal is not rejected at this level. The critical value at the 10% level is 1.83 so for the
comparisons of the sense and antisense experiments with the control the null
hypothesis is not rejected at this less stringent level. The comparison of the sense
experiments with the antisense experiments gives a much lower \( t \) statistic and there is
clearly no evidence for a difference between these results. The probability of
obtaining the experimental results if the sense and antisense constructs have no effect
on \( I\text{-CAT} \) expression is between 0.05 and 0.1.

<table>
<thead>
<tr>
<th></th>
<th>Sum of squares</th>
<th>Degrees of freedom</th>
<th>Mean square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within groups</td>
<td>0.247</td>
<td>9</td>
<td>0.0275</td>
</tr>
<tr>
<td>Between groups</td>
<td>0.0137</td>
<td>2</td>
<td>0.00686</td>
</tr>
</tbody>
</table>

Table 6.2 Analysis of variance table for the results shown in Table 6.1.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Difference of means</th>
<th>Estimated Standard error</th>
<th>( t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sense vs control</td>
<td>0.373</td>
<td>0.179</td>
<td>2.08</td>
</tr>
<tr>
<td>Antisense vs control</td>
<td>0.402</td>
<td>0.182</td>
<td>2.22</td>
</tr>
<tr>
<td>Sense vs antisense</td>
<td>0.0291</td>
<td>0.1003</td>
<td>0.290</td>
</tr>
</tbody>
</table>

Table 6.3 Calculation of \( t \) statistics for comparisons of sense, antisense and control
measurements.

### 6.3 Discussion

The effect of sense and antisense constructs, designed to express AEF-1 RNA
in the female germ line, on expression from the \( I\text{-CAT} \) reporter has been investigated.
In both cases the levels of CAT were lower than in the control experiment with no AEF-1 transgene. The differences were statistically significant at the 10% level but not at the 5% level. However the $t$ statistics for comparisons between the control and the two experimental groups were not much lower than the 5% critical value and were considerably greater than that for the comparison of the sense and antisense results with each other. It is possible that there is an effect of the transgenes on expression but that there is insufficient data here to give a result that is significant at the 5% level.

It should be noted that the power of the statistical test is limited by the fact that the control result represents only a single strain. One of the sources of variation present in the sense and antisense experiments, reflecting position effects on expression of the transgenes, is not present in the control data. This limits the degrees of freedom of the test. With more degrees of freedom the critical values of the $t$ distribution are lower and the test would have more power to detect significant differences. To improve the chance of obtaining significant results if the transgenes do have an effect, it would be necessary to test more strains carrying the sense and antisense constructs, which would increase the degrees of freedom.

Before a positive interpretation is put on the results, it would necessary to demonstrate that RNA is expressed from the sense and antisense transgenes. It would be possible to detect the presence the antisense RNA by northern blotting using a sense AEF-1 probe. For the sense RNA a probe could be designed based on the myc tag sequence at the 5' end of the transcript, which would avoid the problem of detecting RNA derived from the endogenous AEF-1 gene.

It was anticipated that expression of sense RNA from the transgene would lead to an increase in AEF-1 protein levels while expression of antisense RNA would result in a decrease. It is not clear what the actual effect of the sense and antisense constructs was on the AEF-1 concentration in the germ line. To resolve this question, western blotting experiments could be carried out to detect any differences in protein levels between JA and the transformed lines.

If there is any effect of the transgenes then both sense and antisense constructs would appear to cause a change in the same direction, reducing
expression. It is possible that the expression of sense AEF-1 RNA could have resulted in a decrease in AEF-1 protein levels. In plant systems, expression of a transgene that is homologous to an endogenous gene frequently results in post-transcriptional silencing of all copies by a mechanism involving RNA degradation (see Section 1.3.1.3). This is thought to be related to the process of double-stranded RNA-mediated gene silencing which is known to occur in *Drosophila* (Kennerdell and Carthew, 1998). It is also possible that the AEF-1 concentration may be regulated at the protein level, perhaps by post-translational modification, so that sense RNA expression would not increase the amount of an active form of the protein.

The results described in Chapter 5 suggested that there is no dramatic effect on expression when AEF-1 binding is reduced or eliminated by mutation of site 1. If reducing AEF-1 levels does have an effect on expression from the wild-type 5'UTR, this could indicate that the effect is indirect. AEF-1 may act to alter the level of another protein or proteins that interact with the enhancer. Alternatively DNA binding may not be required for a role of AEF-1 in regulation which could be mediated through other proteins by protein-protein interactions. Perhaps a more likely alternative would be that the effect of AEF-1 is subtle so that the experiments of Chapter 5, in which few transformant lines were tested for each mutation, were unable to detect a small effect on expression resulting from the prevention of AEF-1 binding.

Some of these questions could be resolved by testing the effect of the sense and antisense AEF-1 constructs on expression from the mutant versions of the *I-CAT* reporter. It would also be interesting to test the effect of having both the sense and antisense constructs in the same flies. If a double-stranded AEF-1 RNA is produced it would be predicted to reduce AEF-1 levels, perhaps more effectively than the antisense construct alone, and may result in a measurable significant effect on expression from the 5'UTR.
Chapter 7 - Reactivity and the *opa* motif of the AEF-1 gene
Reactivity and the *opa* motif of the AEF-1 gene

7.1 Introduction

The variability in the response of reactive strains to a dysgenic cross, characterised as 'reactivity', is affected by genes carried on all three major chromosomes (Bucheton and Bregliano, 1982). These genes may affect the transposition frequency of the *I* factor or the ability of the germ line cells to cope with the consequences of transposition. Although none have as yet been identified it is likely that they are diverse, having a variety of functions in transcription, recombination, DNA repair and chromatin structure. The discovery that AEF-1 binds in the *I* factor 5' UTR raises the possibility that polymorphisms in the AEF-1 gene may be one of the sources of variation that determine the reactivity of a strain.

The coding region of AEF-1 contains an *opa* motif (Wharton *et al*., 1985, Falb and Maniatis, 1992a), a repeat consisting of (CAN)$_{21}$ encoding the sequence Q$_{13}$H$_{2}$Q$_{6}$. In humans, several neurodegenerative diseases are associated with instability in the length of trinucleotide repeat tracts located in certain genes. The type I trinucleotide repeat disorders involve amplification in coding regions of CAG repeats which encode stretches of glutamine. These include Huntington's disease (The Huntington's Disease Collaborative Research Group, 1993), the spinocerebellar ataxias (Orr *et al*., 1993, Imbert *et al*., 1996, Pulst *et al*., 1996, Sanpei *et al*., 1996, Zhuchenko *et al*., 1997), dentatorubral-pallidoluysian atrophy (Koide *et al*., 1994, Nagafuchi *et al*., 1994), Machado-Joseph's disease (Kawaguchi *et al*., 1993) and spinobulbar muscular atrophy (La Spada *et al*., 1991). Repeat expansion results in the production of proteins with aberrant properties as a result of elongated glutamine stretches (Trottier *et al*., 1995a,b, Persichetti *et al*., 1995) which is believed to affect the proteins' cellular functions resulting in disease (Reddy and Houseman, 1997). Although there are no reports of expansion of *opa* repeats in *Drosophila* the possibility exists that there is variability in the AEF-1 repeat length and that this could partially explain the differences in reactivity.
between certain strains. This chapter describes an experiment designed to detect any variability in the AEF-1 *opa* repeat length between strong and weak reactive strains.

### 7.2 Results

Using a polymerase chain reaction assay, genomic DNA from five *Drosophila melanogaster* strains was tested for the length of the *opa* motif in the AEF-1 gene. The strains are listed in Table 6.1 along with an indication of their reactivity level based on previous studies (Marie-Christine Chaboissier, personal communication).

<table>
<thead>
<tr>
<th>Strain</th>
<th>Reactivity</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>vest</em></td>
<td>Low</td>
</tr>
<tr>
<td><em>Paris</em></td>
<td>Low</td>
</tr>
<tr>
<td><em>H530</em></td>
<td>Low</td>
</tr>
<tr>
<td><em>Charolle</em></td>
<td>High</td>
</tr>
<tr>
<td><em>JA</em></td>
<td>High</td>
</tr>
</tbody>
</table>

*Table 6.1* Reactive *Drosophila melanogaster* strains.

The primers 5’*opa* and 3’*opa* amplify across the AEF-1 *opa* motif to produce a fragment of 128bp assuming that there are 21 CAN repeats as in the published AEF-1 sequence (Falb and Maniatis, 1992a). PCR products were run in an agarose gel (Figure 6.1) to detect any differences in the size of the products amplified from different strains. In each case a band is amplified that migrates at the same mobility as the 128bp band amplified from the cloned AEF-1 cDNA (lane 2).
Figure 6.1 PCR fragments containing the AEF-1 opa motif amplified from genomic DNA from strong and weak reactive strains of *Drosophila melanogaster*. Lane 1 - Size markers (φX174 DNA digested with *HaeIII*), lane 2 - PCR using plasmid pBS35 containing the cloned AEF-1 cDNA (Falb and Maniatis, 1992a), lane 3 - Control PCR with no template DNA, lanes 4-8 - PCR using *Drosophila* genomic DNA from different strains as the template: lane 4 - vest, lane 5 - Paris, lane 6 - H530, lane 7 - Charolle, lane 8 - JA
7.3 Discussion

The experiment described in this chapter provides no evidence that there is variation in the length of the AEF-1 *opa* motif. The gel used would have detected a difference in fragment length of 10bp or just over 3 repeat copies. Clearly variability in repeat length is not responsible for the variation in reactivity between the strains tested here although it remains a possibility that expansion at the AEF-1 gene plays a role in some strains.
Chapter 8 – General discussion
General Discussion

8.1 Conclusions

The I factor 5' untranslated region contains an internal promoter and enhancer elements that direct high levels of transcription in the female germ line. An earlier study established that there is a protein present in Drosophila ovaries which binds specifically to site 1, an element within the enhancer. Experiments using a reporter gene, in which expression is under the control of the I factor regulatory sequences, suggested that site 1 is essential for the function of the enhancer. A construct in which site 1 was deleted gave a substantially reduced level of expression and expression in ovaries was not greater than in other tissues. This suggested that the site 1 binding protein might be a strong activator of transcription in the female germ line.

The data presented in this thesis strongly suggest that the site 1 binding protein is AEF-1, a known repressor of the transcription of several genes including yolk proteins 1 and 2 and alcohol dehydrogenase. This raised the question of whether AEF-1 is an activator or a repressor of the I factor enhancer. Two experimental approaches were employed to investigate this question; mutagenesis of site 1 within the context of the I factor regulatory region, and varying the level of AEF-1 in the female germ line. In each case the results provided no evidence that AEF-1 affects expression directed by the I factor 5'UTR. The evidence that site 1 is important for the function of the enhancer was investigated. It was found that the reporter gene with site 1 deleted used in the original experiment had a different structure from that of the wild type 5'UTR construct which was used for comparison. In conclusion, it is doubtful whether site 1 and AEF-1 are involved in the function of the I factor enhancer.

8.2 Implications for theories of I factor regulation

The possibility that site 1 may not be required for high levels of germ line-specific I factor transcription has implications for the interpretation of certain earlier experiments. Udomkit et al. (1996) showed that deleting the region from base pairs
101 to 186, which includes site 1, resulted in a reduction of expression in ovaries of around 5 fold. However deletion of site 1 gave a much greater reduction of 100 fold or more. The difficulty of explaining this result is resolved if the reduction observed with the site 1 deleted construct can be attributed to a feature of the structure of that transgene other than the deletion of site 1. It seems clear that the region from 101 to 186 contains sequences that are required for the full level of enhancer activity observed with the complete 5'UTR, but are not crucial for germ line specific transcription.

A revised interpretation of the effect of deleting site 1 also has implications for theories relating to the mechanism by which I factor copy number is controlled in inducer strains. Chaboissier et al. (1998) showed that increasing the number of copies of the 5'UTR alone is sufficient to reduce expression directed by the I factor regulatory sequences and to protect the flies from the sterility associated with I-R hybrid dysgenesis. Similar results were obtained by Jensen et al. (1999) who found a reduction in sterility associated with increasing the dosage of transgenes carrying different fragments of the I factor. For one of these transgenes it was shown that the presence of a promoter was required for this effect, suggesting that transcription of the I factor RNA is necessary to trigger the repression mechanism. As discussed by Birchler et al. (1999), this apparently contrasts with the results of Chaboissier et al. who found that deletion of site 1 had little effect on the ability of multiple 5'UTR copies to repress expression and dysgenesis. If deleting site 1 causes a dramatic reduction in transcription this result is hard to reconcile with a model in which I factor RNA is involved in establishing repression. However if deleting site 1 does not substantially reduce transcription then this is not a problem.

The experiments of Chaboissier et al. were carried out in the genetic background of the JA strain. As discussed in chapter 5, differences in the strain background provide one possible explanation for the different results obtained with site 1 deleted I-CAT constructs between this study and Udomkit et al. (1996). In the experiments described here the JA strain was used, as in Chaboissier et al., suggesting that, at least in that genetic background, deletion of site 1 does not have a substantial effect on expression.
The results in this thesis therefore weaken the hypothesis that transcription of \( I \) factor sequences is not required for copy number control. However another aspect of the methodology of Chaboissier et al. should perhaps be considered in this context. To increase 5'UTR copy number they constructed two \( P \) elements, one containing two tandem copies of the 5'UTR and the other containing three, and combined chromosomes carrying these elements in various ways. In each of the constructs the 5'UTR copies are not separated by 3'RNA processing signals, so that effectively there is only one transcription unit in each case, albeit containing more than one internal promoter. It is not clear that the amount of RNA transcribed from these constructs would be proportional to the copy number of the 5'UTR. It could be less, or perhaps even more because it is possible that placing more than one copy of the \( I \) factor enhancer in close proximity would lead to a co-operative effect in activation. It is therefore not simple to interpret the results of these experiments according to an RNA based model.

To resolve the question of whether transcription of \( I \) factor sequences is required for repression mediated by the 5'UTR, the experiments of Chaboissier et al. could be repeated, but with a mutation in the \( I \) factor promoter in the multicopy transgene, instead of the site 1 deletion. For example, mutating the conserved CA(G/T)T motif at the 5' end of the element to ACCG abolishes transcription from the \( I \) factor promoter in cultured cells (McLean et al., 1993).

The question as to whether silencing operates at the level of transcription or that of RNA stability is not resolved by demonstrating a requirement for \( I \) factor RNA in the process. It is possible that the RNA may be involved in altering the chromatin structure and therefore repressing transcription at the \( I \) factor copies rather than triggering RNA degradation. This question could potentially be resolved by introducing a transgene in which the \( I \) factor enhancer acts on a downstream promoter into an inducer background. For example a construct in which the region from bp 41 to 186 is placed upstream of the \( hsp70 \) promoter gave germ line specific expression of a \( \beta \)-galactosidase reporter in a reactive background (Udomkit et al., 1996). The RNA produced from this transgene does not contain \( I \) factor sequences, so any effect of the inducer state on expression could only be attributed to regulation at the level of transcription.
8.3 Future work

The differences between the results presented here and those of Udomkit et al. (1996) in experiments using site 1 deleted I-CAT reporters have not been unequivocally explained. The most likely reasons relate to the differences between the transgenes used and the different strain backgrounds of the two experiments. The transgenes used by Udomkit et al. should be cloned and sequenced to check the interpretation that the PstI fragment of p1-186Δ.T1 was inserted in the opposite orientation from that of the other constructs. To test the effect of the strain background, the plasmid pΔ.W8 that was used in the transformation of JA could be introduced into W^k^ and the level of CAT expression measured in homozygotes.

If it is found that deleting site 1 does not affect expression in the W^k^ background then it will be necessary to restart the search for proteins that interact with the I factor enhancer. Ovarian expression libraries could be screened, either by the λ method employed in chapter 3 using a probe consisting of the entire 5’UTR, or by a yeast one hybrid approach.
Appendix – Characterisation of transgenic *Drosophila* lines
Appendix - Characterisation of transgenic *Drosophila* lines

Transgenic *Drosophila melanogaster* lines established following *P* element mediated transformation were characterised by Southern blotting. For lines containing *I-CAT* reporter genes (chapter 5), genomic DNA was digested with *EcoRI* and the blots were probed with a 250bp *EcoRI/BamHI* fragment of pCAT.1 containing a segment at the 5' end of the *CAT* gene (see figure 5.1). DNA from lines containing BBS.AEF-1 or BBS.αAEF-1 transgenes (chapter 6) was digested with *XbaI* and the blots probed with a 350bp *EcoRI/XbaI* fragment of pBBS containing a sequence from the *hsp26* gene. This probe hybridises to an *XbaI* fragment of around 4.8kb, presumably derived from the endogenous *hsp26* gene, in addition to the transgene fragment. In each case the enzyme cuts once within the inserted *P* element and the probe hybridizes to only one of the two products of this cleavage. The mobility of the hybridising band depends on the distance from the end of the *P* element to the nearest recognition site for the restriction enzyme in the flanking genomic DNA. Where more than one line was derived from a single injected egg the insertions were considered to be independent if the mobility of hybridising bands in the southern blot was not identical. Southern blots are shown in figures A.1-A.10 and the data summarised in tables A.1 and A.2. For some lines the chromosome on which the transgene was inserted was mapped by the method described in Section 2.2.5.5, and this information is included in the tables.
<table>
<thead>
<tr>
<th>Strain</th>
<th>Plasmid injected</th>
<th>Chromosome</th>
<th>Southern blot figure</th>
<th>Single insertion</th>
</tr>
</thead>
<tbody>
<tr>
<td>466*</td>
<td>p186.W8</td>
<td>-</td>
<td>A.1</td>
<td>✓</td>
</tr>
<tr>
<td>467*</td>
<td>p186.W8</td>
<td>-</td>
<td>A.2</td>
<td>✓</td>
</tr>
<tr>
<td>448*</td>
<td>p186.W8</td>
<td>III</td>
<td>A.2</td>
<td>✓</td>
</tr>
<tr>
<td>469*</td>
<td>pΔ.W8</td>
<td>-</td>
<td>A.2</td>
<td>✓</td>
</tr>
<tr>
<td>470</td>
<td>pΔ.W8</td>
<td>-</td>
<td>A.2</td>
<td>✓</td>
</tr>
<tr>
<td>471*</td>
<td>pΔ.W8</td>
<td>II</td>
<td>A.1</td>
<td>✓</td>
</tr>
<tr>
<td>472*</td>
<td>pΔ.W8</td>
<td>II</td>
<td>A.3</td>
<td>✓</td>
</tr>
<tr>
<td>473</td>
<td>p137G138G.W8</td>
<td>III</td>
<td>A.4</td>
<td>✓</td>
</tr>
<tr>
<td>474*</td>
<td>p137G138G.W8</td>
<td>II</td>
<td>A.4</td>
<td>✓</td>
</tr>
<tr>
<td>475*</td>
<td>p137G138G.W8</td>
<td>-</td>
<td>A.4</td>
<td>✓</td>
</tr>
<tr>
<td>476</td>
<td>p143C144G.W8</td>
<td>-</td>
<td>A.4</td>
<td>✓</td>
</tr>
<tr>
<td>477</td>
<td>p143C144G.W8</td>
<td>-</td>
<td>A.4</td>
<td>✓</td>
</tr>
<tr>
<td>478*</td>
<td>p143C144G.W8</td>
<td>-</td>
<td>A.4</td>
<td>✓</td>
</tr>
<tr>
<td>479*</td>
<td>p143C144G.W8</td>
<td>II</td>
<td>A.3, A.4</td>
<td>✓</td>
</tr>
<tr>
<td>480*</td>
<td>p143C144G.W8</td>
<td>-</td>
<td>A.3, A.4</td>
<td>✓</td>
</tr>
<tr>
<td>481*</td>
<td>p143C144G.W8</td>
<td>-</td>
<td>A.4</td>
<td>✓</td>
</tr>
<tr>
<td>482*</td>
<td>p150G144C.W8</td>
<td>-</td>
<td>A.1</td>
<td>✓</td>
</tr>
<tr>
<td>494*</td>
<td>p150G144C.W8</td>
<td>-</td>
<td>A.1</td>
<td>✓</td>
</tr>
<tr>
<td>483*</td>
<td>p150G144C.W8</td>
<td>-</td>
<td>A.2</td>
<td>✓</td>
</tr>
<tr>
<td>484*</td>
<td>p150G144C.W8</td>
<td>-</td>
<td>A.2</td>
<td>✓</td>
</tr>
<tr>
<td>485*</td>
<td>p150G144C.W8</td>
<td>X</td>
<td>A.2</td>
<td>✓</td>
</tr>
<tr>
<td>486*</td>
<td>p135T136C.W8</td>
<td>-</td>
<td>A.5</td>
<td>✓</td>
</tr>
<tr>
<td>487</td>
<td>p135T136C.W8</td>
<td>-</td>
<td>A.5</td>
<td>✓</td>
</tr>
<tr>
<td>488</td>
<td>p135T136C.W8</td>
<td>-</td>
<td>A.6</td>
<td>x</td>
</tr>
<tr>
<td>489*</td>
<td>p135T136C.W8</td>
<td>-</td>
<td>A.6</td>
<td>✓</td>
</tr>
<tr>
<td>490</td>
<td>p135T136C.W8</td>
<td>-</td>
<td>A.6</td>
<td>✓</td>
</tr>
<tr>
<td>491</td>
<td>p135T136C.W8</td>
<td>-</td>
<td>A.6</td>
<td>✓</td>
</tr>
<tr>
<td>492*</td>
<td>p137G138G143C144G.W8</td>
<td>-</td>
<td>A.1</td>
<td>✓</td>
</tr>
<tr>
<td>493*</td>
<td>p137G138G143C144G.W8</td>
<td>II</td>
<td>A.1</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table A.1 Transgenic lines containing I-CAT reporter genes. The strains that were used in determinations of CAT protein (chapter 5) are marked with asterix. Strains containing independent P element insertions derived from a single injected egg are bracketed.
<table>
<thead>
<tr>
<th>Strain</th>
<th>Plasmid injected</th>
<th>Chromosome</th>
<th>Southern blot</th>
<th>Single insertion</th>
</tr>
</thead>
<tbody>
<tr>
<td>442</td>
<td>pW8.BBS.AEF1.SV2</td>
<td>III</td>
<td>A.7</td>
<td>✓</td>
</tr>
<tr>
<td>443</td>
<td>pW8.BBS.AEF1.SV2</td>
<td>II</td>
<td>A.8, A.9</td>
<td>✓</td>
</tr>
<tr>
<td>444</td>
<td>pW8.BBS.AEF1.SV2</td>
<td>-</td>
<td>A.7</td>
<td>✓</td>
</tr>
<tr>
<td>445</td>
<td>pW8.BBS.AEF1.SV2</td>
<td>X</td>
<td>A.7, A.8</td>
<td>✓</td>
</tr>
<tr>
<td>446</td>
<td>pW8.BBS.AEF1.SV2</td>
<td>-</td>
<td>A.7</td>
<td>×</td>
</tr>
<tr>
<td>447</td>
<td>pW8.BBS.AEF1.SV2</td>
<td>-</td>
<td>A.8</td>
<td>✓</td>
</tr>
<tr>
<td>448</td>
<td>pW8.BBS.AEF1.SV2</td>
<td>-</td>
<td>A.7, A.8</td>
<td>×</td>
</tr>
<tr>
<td>449</td>
<td>pW8.BBS.AEF1.SV2</td>
<td>II</td>
<td>A.7</td>
<td>✓</td>
</tr>
<tr>
<td>450</td>
<td>pW8.BBS.AEF1.SV2</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>451</td>
<td>pW8.BBS.AEF1.SV2</td>
<td>II</td>
<td>A.9</td>
<td>-</td>
</tr>
<tr>
<td>452</td>
<td>pW8.BBS.AEF1.SV2</td>
<td>III</td>
<td>A.7</td>
<td>✓</td>
</tr>
<tr>
<td>453</td>
<td>pW8.BBS.AEF1.SV2</td>
<td>III</td>
<td>A.7, A.8</td>
<td>✓</td>
</tr>
<tr>
<td>454</td>
<td>pW8.BBS.AEF1.SV2</td>
<td>III</td>
<td>A.7</td>
<td>✓</td>
</tr>
<tr>
<td>455</td>
<td>pW8.BBS.αAEF1.SV2</td>
<td>-</td>
<td>A.10</td>
<td>×</td>
</tr>
<tr>
<td>456</td>
<td>pW8.BBS.αAEF1.SV2</td>
<td>II</td>
<td>A.10</td>
<td>✓</td>
</tr>
<tr>
<td>457</td>
<td>pW8.BBS.αAEF1.SV2</td>
<td>III</td>
<td>A.10</td>
<td>✓</td>
</tr>
<tr>
<td>458</td>
<td>pW8.BBS.αAEF1.SV2</td>
<td>II</td>
<td>A.10</td>
<td>✓</td>
</tr>
<tr>
<td>459</td>
<td>pW8.BBS.αAEF1.SV2</td>
<td>III</td>
<td>A.10</td>
<td>x</td>
</tr>
<tr>
<td>460</td>
<td>pW8.BBS.αAEF1.SV2</td>
<td>-</td>
<td>A.10</td>
<td>✓</td>
</tr>
<tr>
<td>461</td>
<td>pW8.BBS.αAEF1.SV2</td>
<td>II</td>
<td>A.10</td>
<td>✓</td>
</tr>
<tr>
<td>462</td>
<td>pW8.BBS.αAEF1.SV2</td>
<td>II</td>
<td>A.10</td>
<td>✓</td>
</tr>
<tr>
<td>463</td>
<td>pW8.BBS.αAEF1.SV2</td>
<td>X</td>
<td>A.9</td>
<td>✓</td>
</tr>
<tr>
<td>464</td>
<td>pW8.BBS.αAEF1.SV2</td>
<td>X</td>
<td>A.10</td>
<td>✓</td>
</tr>
<tr>
<td>465</td>
<td>pW8.BBS.αAEF1.SV2</td>
<td>-</td>
<td>A.9</td>
<td>×</td>
</tr>
</tbody>
</table>

Table A.2 Transgenic lines containing sense and antisense AEF-1 germline expression constructs. The strains that were used in experiments (chapter 6) are marked with asterisk. Strains containing independent \( P \) element insertions derived from a single injected egg are bracketed.
Figure A.1 Southern blot analysis of transgenic fly lines containing p186.W8, pΔ.W8, p150G151C.W8 and p137G138G143C144G.W8 P element insertions. Eco RI digests of genomic DNA were separated on a 1% agarose gel before blotting. The probe was the 250bp EcoRI/BamHI fragment of pCAT.1.
Figure A.2 Southern blot analysis of transgenic fly lines containing p186.W8, pΔ.W8 and p150G151C.W8 P element insertions. Eco RI digests of genomic DNA were separated on a 1% agarose gel before blotting. The probe was the 250bp EcoRI/BamHI fragment of pCAT.1.
Figure A.3 Southern blot analysis of transgenic fly lines containing pΔ.W8 and p143C144G.W8 P element insertions. Eco RI digests of genomic DNA were separated on a 1% agarose gel before blotting. The probe was the 250bp EcoRI/BamHI fragment of pCAT.1.
Figure A.4 Southern blot analysis of transgenic fly lines containing p137G138G.W8 and p143C144G P element insertions. Eco RI digests of genomic DNA were separated on a 1% agarose gel before blotting. The probe was the 250bp EcoRI/BamHI fragment of pCAT.1.
Figure A.5 Southern blot analysis of transgenic fly lines containing p135T136C.W8 P element insertions. *Eco* RI digests of genomic DNA were separated on a 1% agarose gel before blotting. The probe was the 250bp *EcoRI/BamHI* fragment of pCAT.1.
Figure A.6 Southern blot analysis of transgenic fly lines containing p135T136C.W8 P element insertions. *Eco RI* digests of genomic DNA were separated on a 1% agarose gel before blotting. The probe was the 250bp *EcoRI/BamHI* fragment of pCAT.1.
Figure A.7 Southern blot analysis of transgenic fly lines containing pW8.BBS.AEF1.SV2 P element insertions. Xbal digests of genomic DNA were separated on a 1% agarose gel before blotting. The probe was the 350bp EcoRI/Xbal fragment of pBBS.
Figure A.8 Southern blot analysis of transgenic fly lines containing pW8.BBS.AEF1.SV2 P element insertions. XbaI digests of genomic DNA were separated on a 1% agarose gel before blotting. The probe was the 350bp EcoRI/XbaI fragment of pBBS.
Figure A.9 Southern blot analysis of transgenic fly lines containing pW8.BBS.AEF1.SV2 and pW8.BBS.αAEF1.SV2 P element insertions. XbaI digests of genomic DNA were separated on a 1% agarose gel before blotting. The probe was the 350bp EcoRI/XbaI fragment of pBBS.
Figure A.10 Southern blot analysis of transgenic fly lines containing pW8.BBS.αAEF1.SV2 P element insertions. *Xbal* digests of genomic DNA were separated on a 1% agarose gel before blotting. The probe was the 350bp EcoRI/Xbal fragment of pBBS.
References


Antequera, F., Boyes, J. and Bird, A. 1990. High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. Cell 62:503-514


Atwood, A., Lin, J. H. and Levin, H. L. 1996. The retrotransposon Tf1 assembles virus-like particles that contain excess Gag relative to integrase because of a regulated degradation process. Molecular and Cellular Biology 16:338-346


**Brutnell, T. P. and Dellaporta, S. L.** 1994. Somatic inactivation and reactivation of *Ac* associated with changes in cytosine methylation and transposase expression. *Genetics* 138:231-225


Burgers, P. M. 1993. *Saccharomyces cerevisiae* replication factor C. II. Formation and activity of complexes with the proliferating cell nuclear antigen and with DNA polymerases delta and epsilon. *Journal of Biological Chemistry* 266:22698-22706

Busseau, I., Malinsky, S., Balakireva, Chaboissier, M-C., Teninges, D and Bucheton, A. 1998. A genetically marked *I* element in *Drosophila melanogaster* can be mobilised when ORF2 is provided in trans. *Genetics* 148:267-275


Faugeron, G., Rhounim, L. and Rossignol, J. L. 1990. How does the cell count the number of ectopic copies of a gene in the premeiotic inactivation process acting in *Ascobolus immersus?* *Genetics* 124:585-591


Fedoroff, N. V. and Banks, J. A. 1988. Is the Suppressor-mutator element controlled by a basic developmental regulatory mechanism? Genetics 120:559-577


Feng, Q., Schumann, G. and Boeke, J. D. 1998. Retrotranposon R1Bm endonuclease cleaves the target sequence. Proceedings of the National Academy of Sciences USA 95:2083-2088


Hanahan, D. 1990. Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* 166:577-580


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*. *Journal of Molecular Biology* 212:37-52


Linn, F., Heidmann, I., Saedler, H. and Meyer, P. 1990. Epigenetic changes in the expression of the maize A1 gene in Petunia hybrida: role of numbers of
integrated gene copies and state of methylation. *Molecular and General Genetics* **222**:329-336


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


Pelissier, T., Thalmeir, S., Kempe, D., Sanger, H. L. and Wassenegger, M. 1999. Heavy de novo methylation at symmetrical and non-symmetrical sites is a hallmark of RNA-directed DNA methylation. *Nucleic Acids Research* 27:1625-1634


Péligson, A. and Bregliano, J. C. 1987. Evidence for a rapid limitation of I element copy number in a genome submitted to several generations of I-R hybrid
dysgenesis in Drosophila melanogaster. Molecular and General Genetics 207:306-313


of in situ hybridisation to show the association of I factor DNA with induced sex-linked recessive lethals. *Mutation Research* 268:265-285


**Rabinow, L., Chiang S. L. and Birchler, J. A.** 1993. Mutations at the *Darkener of apricot* locus modulate transcript levels of *copia* and *copia*-induced mutations in *Drosophila melanogaster*. *Genetics* 134:1175-1185


**Rasmusson, K. E., Raymond, J. D. and Simmons, M. J.** 1993. Repression of hybrid dysgenesis in *Drosophila melanogaster* by individual naturally occurring P elements. *Genetics* 133:605-622


Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique. *Nature Genetics* 14:277-284


170


Tsurimoto T and Stillman, B. 1989. Multiple replication factors augment DNA synthesis by the two eukaryotic DNA polymerases, alpha and delta. *EMBO Journal* 8:3883-3889


Yang, J. and Eickbush, T. H. 1998. RNA-induced changes in the activity of the endonuclease encoded by the R2 retrotransposable element. Molecular and Cellular Biology 18:3455-3465


