THE *SUPPRESSOR-OF-FORKED* LOCUS OF *DROSOPHILA MELANOGASTER*

Ian Andrew Mark Mitchelson

Department of Biochemistry

Imperial College of Science and Technology

London SW7 2AZ

Submitted for the Degree of PhD

July 1988

Il est démontré, disait-il, que les choses ne peuvent être autrement; car tout étant fait pour une fin, tout est nécessairement pour la meilleure fin.

Voltaire, *Candide*.

ABSTRACT

The *suppressor-of-forked* locus [*su(f)*; 1:65.9] of *D.melanogaster* is a cell-autonomous lethal gene and an allele-specific, trans-acting modifier. Several unlinked mutant alleles whose phenotypes are suppressed or enhanced by mutations at *su(f)* are associated with insertions of retroviral-like transposable elements. The *su(f)* gene product may therefore play a role in regulating transposable element activity, in addition to, or as part of, its role as a vital cellular gene. In addition, *su(f)* is the most proximal known "euchromatic" gene on the *D.melanogaster* X chromosome, lying at the euchromatin/heterochromatin boundary. Using P element transposon tagging, a total of 32kb of wild type DNA from the *su(f)* region was isolated using a P element insertion allele of *su(f)*. Approximately two thirds of this interval consists of repetitive DNA sequences, interspersed with regions of unique DNA. The site of insertion of the P element allele lies in a ~7kb unique DNA segment flanked by direct copies of a ~1.5kb middle-repetitive DNA sequence. DNA blotting analysis of 18 mutant *su(f)* alleles showed that 9 lethal *su(f)* alleles differed from wild type in the 14kb region examined. 8 of these 9 alleles had lesions clustering in a 4.3kb interval, including the site of insertion of the P element used to clone *su(f)*. DNA from the *su(f)* region of two other lethal *su(f)* alleles was cloned for further analysis: a second P element insertion 75bp from the site of the first, and a Doc transposable element insertion. Several poly-A^+ RNA transcripts were found in the region where DNA lesions of *su(f)* alleles are clustered; the most abundant being a 2.6kb species, together with a 2.9kb and a 1.3kb RNA. All three RNAs are transcribed throughout development, but their levels are highest in embryos, early pupae and adult females. The three RNAs are qualitatively altered to give slightly larger transcripts in the viable allele $su(f)^1$; other qualitative changes in their transcription pattern were seen in the P element insertion allele $su(f)^{1-MS252}$. Several cDNA clones were isolated from libraries of poly-A^+ RNA from first/second instar poly-A^+ RNA and from adult females. Structural and sequence analysis of these gave evidence for two alternative 3' polyadenylation sites in the *su(f)* transcription unit, as well as a divergent transcription unit ~900bp 5' to the *su(f)* transcription unit. The largest cDNA corresponds to the 2.6kb RNA; this transcript has nine exons and a long open reading frame encoding a putative 733 amino acid, 84kD protein. The carboxy-terminal 20% of the putative protein is proline rich, suggesting a C-terminal portion with little or no alpha-helix or beta-sheet tertiary structure. As one of the few allele-specific modifier genes cloned, further analysis of *su(f)* should provide a better understanding of the function of genes which, when mutated, modify the mutagenic effects of transposable element insertions.

ACKNOWLEDGEMENTS

I am indebted to Kevin O'Hare for his excellent supervision, a talent developed without the benefit of a previous graduate student to practise on. His good advice and continued interest in my work combined to make the time I spent at Imperial College extremely enjoyable. Alan Driver and Jamie Paterson provided a constant source of good humour and good company in the workplace. Many other inhabitants of the 6th floor helped me by giving their advice and opinions, including Roger Karess, David Glover, Steve Delaney, Matt Freeman, Claudio Sunkel and Debbie Smith. Thanks are also due to the labours of Alan Cheshire, Julie Groves and Brenda Barlow for fly media and technical support.

Many thanks to Mike Simmons, who provided the basic resource without which my thesis work would have been on something else. Thanks also to Margaret Kidwell for teaching me the basics of *Drosophila* genetics. I am indebted to several workers for freely sharing their unpublished results and observations: Abraham Schalet, Susan Parkhurst, Victor Corces, Jack Girton, George Miklos, and Danielle Thierry-Mieg. Thanks are also due to those who sent me fly stocks; including George Miklos, Elizabeth Gateff, Tom Wilson, and Abraham Schalet. The last named was also very generous in explaining the varied (and in some cases confusing) provenances of *su(f)* alleles, for which I am very grateful.

The work contained in this thesis was carried out whilst I was in receipt of a Training Award from the Medical Research Council. I am also very grateful for financial support from the Imperial Chemical Industries Educational Trust, who awarded me a Postgraduate Research Studentship.

Finally, thanks are due to my family for their support, and to Claire and Simon Oldfield and Robin Earle for helping me to stay cheerful.

CONTENTS

LIST OF FIGURES

LIST OF FIGURES (cont.)

LIST OF TABLES

LIST OF ABBREVIATIONS

bp	nucleotide base or base pair
cDNA	complementary DNA
Ci	Curie
cpm	counts per minute
DNase	deoxyribonuclease
DTT	dithiothreitol
EDTA	ethylenediamine tetra-acetic acid
F_1	first filial generation
F_2	second filial generation
Hepes	N-2-hydroxyethylpiperazine N'-2-ethanesulphonic acid
IPTG	isopropyl β-D-thiogalactopyranoside
kb	kilobase
kD	kiloDalton
LTR	long terminal repeat
PEG	polyethylene glycol
pers.comm.	personal communication
poly-A^+ RNA	polyadenylated RNA
rDNA	ribosomal DNA
RNase	ribonuclease
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
SSC	standard saline citrate
STE	saline tris EDTA
T.E. (10,1)	Tris.HCl pH7.5 (10mM), EDTA (1mM)
TM10	Tris.HCl pH7.5 (10mM), $MgCl_2$ (10mM)
Tris	tris (hydroxymethyl) methylamine
ts	temperature sensitive
XGal	5-bromo-4-chloro-3-indolyl β-D-galactopyranoside

CHAPTER 1

INTRODUCTION

1.1 PRELIMINARY REMARKS

A major goal of biology is the understanding of processes by which genes are regulated. What factors are responsible for controlling the activity of genes, their tissue specificity, temporal regulation and their response to extracellular stimuli? In a few cases, the approach to answering such questions has involved studying mutations whose phenotypes identify them as potential regulatory genes. Unsurprisingly, mutations of this type are usually to be found in organisms where genetic analysis has been possible at advanced levels; such as bacteria, yeast and the fruit fly *Drosophila melanogaster*. The molecular cloning and further analysis of one such gene in *D.melanogaster* is the subject of this thesis: *suppressor-of-forked*. It is one of a small group of genes in *D.melanogaster* which genetic studies have implicated as being trans-acting factors with a role in regulating gene activity. This conclusion was reached after examining their interactions with mutations at unlinked, mutant loci associated with transposable element insertions.

Transposable elements are but one type of repetitive DNA sequence found in *D.melanogaster*. A striking feature of the structural organisation of eukaryotic genomes is that protein coding sequences and their associated controlling regions form only a small proportion of the total genome content. The bulk of genomic DNA consists of various types of repetitive sequence (Appels and Peacock, 1978). Much of this is in the form of very highly repetitive tandem arrays of short (12bp or less) sequences known as satellite DNA. The other types of repetitive DNA are classed together as "moderately repetitive sequences", where each repeat unit is longer (300 to 14,000bp) but occurs less frequently. Such sequences can either occur in tandem arrays, or be dispersed throughout the genome.

Drosophila melanogaster is the best analysed eukaryote in terms of genome structure (reviewed by Spradling and Rubin, 1981). It has a haploid genome size of approximately 165,000kb, of which about 16% consists of four major classes of

satellite DNA. The satellite sequences occur almost entirely within heterochromatin: that cytologically distinguishable portion of the genome surrounding the centromeres of each chromosome and filling the Y chromosome (Heitz, 1934). In addition, about 15% of the *D.melanogaster* genome consists of moderately repetitive DNA. Of this, approximately one quarter occurs as tandem arrays of rDNA genes and histone genes; the remainder consists of a wide variety of different repetitive sequence families dispersed throughout the genome. These latter vary in size from about 0.5 to 14kb.

A great deal of intellectual and experimental effort has been spent in trying to discover how such sequences arose, and how they came to form such a high proportion of eukaryotic genomes when any "classical" genetic functions they might have are not immediately apparent (for example: Doolittle and Sapienza, 1980; Orgel and Crick, 1980; Brutlag, 1980). This has led to a greater awareness of the dynamic nature of eukaryotic genomes, with saltatory replication, gene converion, unequal crossing-over and transposition all contributing to a "fluid" genome. The behaviour and properties of repetitive sequences, though the sequences themselves may be phenotypically silent, may play a hitherto underestimated role in evolutionary processes. They may promote the rearrangement of genes and their controlling sequences, allowing more efficient combinations of genes (Temin and Engels, 1984); and possibly speciation through the fixation of sequence variations in controlling regions which interact with trans-acting regulatory factors (Dover, 1982).

Apart from possible evolutionary roles, can repetitive DNA sequences directly affect the phenotype of the organism in which they are found? Repetitive sequences such as mobile genetic elements can disrupt gene function if they are inserted within or near coding sequences. Indeed, in *D.melanogaster* a high proportion of spontaneously occurring mutations are due to transposable element insertions of

this nature (see for example Modolell *et al.*, 1983). Such mutations form the most readily detectable manifestation of transposable element activity in *D.melanogaster*. Yet interactions between transposable elements and genes resulting from their physical juxtaposition are non-specific; transposable elements may in some cases have a preference for particular DNA sequences at their insertion sites (Finnegan and Fawcett, 1986), but they are blind to the identity of the gene being disrupted.

In several organisms there are examples of what appears to be direct and specific interactions between cellular genes and various transposable elements, and as such they are unique. These genes are collectively known as second site modifiers, and their curious genetic properties identify them as genes which may have a role as regulators of gene activity. The *suppressor-of-forked* gene of *D.melanogaster* is one such gene.

1.2 TRANSPOSABLE ELEMENTS IN *DROSOPHILA MELANOGASTER*

About 10% of the *D.melanogaster* genome consists of dispersed repetitive DNA sequences (Rubin and Spradling, 1981). This fraction itself comprises a number of discrete families of transposable elements. Our knowledge of *D.melanogaster* transposable elements has been recently reviewed (Finnegan and Fawcett, 1986); this section will only describe those properties of transposable elements relevant to this thesis. Transposable elements of *D.melanogaster* fall into four classes on the basis of their sequence organisation: copia-like elements, retroposons, elements with short inverted terminal repeats and foldback elements.

1.2.1 COPIA-LIKE ELEMENTS

There are some 30-50 transposable element families of this class, each of which may be reiterated within the genome 5-50 times. Of these families, 17 have been characterised by cloning (reviewed in Finnegan and Fawcett, 1986). Elements of

this class create a short target site duplication upon insertion, whose length varies but is usually characteristic for a particular family. Copia-like elements are characterised by having long direct terminal repeats (LTRs) of 250-500bp which flank the central region of the element, giving an overall size of between 4.4-8.8kb. With one or two exceptions, the LTRs of different families of copia-like elements do not share sequence homology.

The structure of copia-like elements is very similar to that of integrated proviruses of vertebrate retroviruses (Varmus, 1983), a similarity maintained in several other respects. Like retroviral LTRs, the LTRs of copia-like elements contain sequences which could regulate transcriptional activity: a TATA box promoter sequence and an AATAAA polyadenylation signal. Nucleotide sequence analysis of the elements copia (Emori *et al.*, 1985), 17.6 (Saigo *et al.*, 1984) and gypsy (Marlor *et al.*, 1986) reveals in each of them open reading frames with amino-acid homology to several retroviral *gag* and *pol* proteins. Extrachromosomal circular molecules bearing copia-like element sequences have been detected in tissue culture cells for various elements, including copia (Flavell and Ish Horowicz, 1981), mdg1 and mdg3 (Ilyin *et al.*, 1984). These and other discoveries suggest that copia-like transposable elements transpose via a mechanism similar to that of proviral integration (Flavell, 1984; Arkhipova *et al.*, 1986).

Copia-like elements, especially copia itself, are similar in structure to Ty transposable elements found in yeast (Roeder and Fink, 1983), and Ty elements are known to transpose via an RNA intermediate in a retroviral fashion (Boeke *et al.*, 1985). Most intriguingly, structures resembling retroviral core particles and which contain copia-complementary RNA have been described in *D.melanogaster* tissue culture cells (Shiba and Saigo, 1983). Such observations have led to the notion that retroviruses may be evolutionarily descended from cellular transposable elements of a copia-like form (Temin, 1980; Temin and Engels, 1984).

Most copia-like elements are transcribed to yield poly-A^+ RNA messages of various sizes; both copia and B104 produce transcripts which initiate in the 5' LTR and terminate within the 3' LTR, as well as shorter transcripts (Flavell *et al.*, 1980; Scherer *et al.*, 1982). The amount of transcripts from copia-like elements varies considerably during *D.melanogaster* development, and various elements display different transcriptional profiles (Schwartz *et al.*, 1982; Parkhurst and Corces, 1987). Parkhurst and Corces (1987) have made an attempt at classifying copia-like elements on the basis of their transcriptional profiles. Those in the first class (412, mdg1, 17.6 and 3S18) are transcribed mainly in the early larval and pupal stages; HMS Beagle and Springer form another class where transcripts are most abundant in adults, with the amounts in the larval stages higher than in pupae, and lastly copia, 297, and B104 each show an individual pattern differing from the two previous classes. For all elements where RNA was detectable, they found a transcript whose size corresponded to transcription extending from the 5' LTR to the 3' LTR, with the exception of 297 elements. In three different wild type strains, Schwartz *et al.* (1982) found a similar profile of copia transcription. As the chromosomal distribution of copia elements presumably differ widely in these strains, this suggests that the transcriptional activity of copia is largely unaffected by chromosomal location.

Thus, for copia-like elements there is variation in transcriptional activity during development; and variation between different elements in their transcriptional profiles. This has been taken to suggest that the transcriptional activity of copia-like elements may be affected by a number of cellular genes.

1.2.2 ELEMENTS WITHOUT TERMINAL REPEATS : RETROPOSONS

Element families of this class include F, G, I, Doc and Jockey. These are dispersed repetitive elements which are bounded by short direct repeats of target DNA sequences, as for copia-like elements. The elements themselves have no

terminal repeats, direct or otherwise. Instead, their most striking feature is a tract of A residues at what is conventionally referred to as their 3' or right hand termini; preceeding the poly-A tract is an AATAAA polyadenylation signal. Elements of this class have been termed retroposons, a class of transposable element including human Alu sequences, mammalian LINEs and several rat and human pseudogenes (reviewed by Rogers, 1985).

Retroposons are believed to transpose by integration of double-stranded molecules formed by reverse transcription of poly-A^+ RNA species. Once inserted, retroposons may or may not be able to autonomously catalyse their own transposition once inserted, depending on their sequence. To do so, they must possess the transcriptional signals needed to produce an RNA transcript suitable for a further round of reverse translation and integration; this is not true at least for the F element at 101F (Dawid *et al.*, 1981). Secondly, they must also possess an internally-encoded reverse transcriptase gene suitable for catalysing further transposition. At least two different retroposon-type transposable elements appear to encode a reverse transcriptase-like protein: the I factor and G elements. I factors differ from other elements without terminal repeats such as F or Doc in that they do not have a poly-A tract at the 3' end; instead they have a variable number of repeats of the triplet TAA (Fawcett *et al.*, 1986). The nucleotide sequence of an I factor has two large open reading frames, the larger of which has amino acid homology to several retroviral reverse transcriptases (Fawcett *et al.*, 1986). Similarly, the nucleotide sequence of a G element shows that it too has coding capacity for two polypeptides, one of which has homology to reverse transcriptases (Di Nocera, 1988).

Doc elements are members of the class of *D.melanogaster* transposable elements without a terminal repeat. The first example, approximately 4.3kb long, was discovered in DNA from the *bithorax* region in bx^3 flies, but is not responsible for

this mutation (Bender *et al.*, 1983). Schneuwly *et al.* (1987) have analysed two Doc elements lying at each end of the *Antennapedia*^{73b} inversion. Another Doc element is associated with the mutation *white*^{1}, and the nucleotide sequnce of the termini of both this copy and of several other Doc elements cloned by homology to it have been analysed (K.O'Hare and S.Lacey, unpublished). All have a poly-A tract of variable length at the 3' end; some appear to be slightly truncated at the 5' end; there is also some sequence variation between the 5' ends. The size of the target site duplication caused by Doc insertions varies between 2 and 14bp.

Data on transcriptional activity of *D.melanogaster* retroposons is scarce. No poly-A^+ RNA from F elements was detectable in embryos, larvae, pupae or adults (Dawid *et al.*, 1981). Whether or not G, Doc and Jockey elements are transcribed is not known.

1.2.3 P ELEMENTS AND HYBRID DYSGENESIS

Hybrid dysgenesis is the name given to a collection of traits obtained after certain interstrain crosses of *D.melanogaster*. These include sterility, increased rates of spontaneous mutation, altered transmission ratios, chromosomal aberrations and male recombination. Two independent systems of hybrid dysgenesis have been studied in detail in *D.melanogaster*: I-R dysgenesis and P-M dysgenesis (reviewed by Bregliano and Kidwell, 1983). I-R dysgenesis appears to be due to the activity of I factors, and P-M dygenesis is associated with the activity of P factors. Evidence has also been found for a third system, H-E dysgenesis, involving a class of transposable elements named hobo (Blackman *et al.*, 1987; Yannopoulos *et al.*, 1987).

In the P-M system, the dysgenic traits are obtained when P strain males are mated to M strain females. In the reciprocal cross, the progeny remain unaffected. P-M dysgenic traits are controlled by P elements: sequences which fall into a third

structural class of *D.melanogaster* transposable elements characterised by having short inverted terminal repeats. P strains contain multiple dispersed copies of a 2.9kb P element, the P factor, along with an array of smaller elements which are apparently derived from internal deletions within the 2.9kb factor (O'Hare and Rubin, 1983). When P strain males are mated to females lacking P elements (M strain females), P elements transpose at high frequencies in the germline of the F_1. This leads in the F_2 to flies with mutations resulting from P element insertions (Rubin *et al.*, 1982), along with other events such as deletions and chromosomal rearrangements catalysed by P elements (Engels and Preston, 1981; Engels and Preston, 1984).

There are 30-50 P elements per haploid genome in P strains; in the P strain π_2 approximately one third of these are 2.9kb P factors (Bingham *et al.*, 1982; O'Hare and Rubin, 1983). P elements have perfect 31bp inverted terminal repeats, and their insertion into the genome results in an 8bp duplication of the cellular target sequence (O'Hare and Rubin, 1983). Because of the high rates of P element transposition during hybrid dysgenesis, P elements have been widely used in a technique of gene cloning known as transposon tagging (Bingham *et al.*, 1981). This involves the isolation of a P element insertion allele of the gene of interest; DNA sequences from the gene may then be cloned using P element sequences as hybridisation probes to screen genomic libraries of DNA from the insertion allele (for example, Searles *et al.*, 1983).

When microinjected into embryos, cloned DNA carrying a 2.9kb P factor can autonomously promote P element transposition (Spradling and Rubin, 1982). This capability forms the basis of another major practical application of P elements; that of P element-mediated germ line transformation (Spradling and Rubin, 1982; Rubin and Spradling, 1982). Cloned genes may now be inserted within the termini of a non-autonomous P element and coinjected with a plasmid carrying an autonomously

functioning "helper" P element which has been altered to prevent its own transposition. Transposition of the vector transposon and its insertion into a chromosome of a germ line cell results in transformed progeny carrying a defined segment of DNA. Transformed progeny are identified through the use of a marker gene also carried by the P element transformation vector. P element-mediated germ line transformation provides an immensely powerful technique for analysing the function and expression of genes contained in cloned fragments of DNA.

1.2.4 TRANSPOSABLE ELEMENTS AS MUTAGENIC AGENTS

When analysed at the molecular level, many spontaneously occurring mutants of *D.melanogaster* turn out to be associated with transposable element insertions, as in the case of the *white* locus (Bingham *et al.*, 1981; Levis *et al.*, 1982; Zachar and Bingham, 1982; O'Hare *et al.*, 1984); the *Notch* locus (Kidd and Young, 1986) and the *bithorax* complex (Bender *et al.*, 1983). Indeed, the very first mutation isolated in *D.melanogaster*, $white^{1}$, is associated with the insertion of a Doc element in the 5' region of the *white* gene (S.Lacey and K.O'Hare, unpublished). As a result of their mobile nature, transposable elements can act as mutagenic agents by transposing into genes, causing physical disruption. It must be remembered that most transposable elements carry sequences for directing and regulating transcriptional activity, and that these may in some instances be the root cause of the mutant phenotype observed.

Mutations caused by transposable element insertions fall into several classes, according to the nature of the mutation and by the means by which the mutagenic effect is exerted. The first of these are mutations caused by the insertion of transposable element sequences within coding regions, and are usually null alleles. Several P element alleles of the white locus are of this type: $w^{\#6}$, $w^{\#12}$, and $w^{hd81b11}$ (O'Hare *et al.*, 1984).

Another type of effect may be exerted by an insertion which lies within the transcribed region, but within an intron. An example of a mutation like this is *white*^{apricot}, an insertion of a copia element within the 73bp second intron of *white* (Bingham *et al.*, 1981; O'Hare *et al.*, 1984). Figure 1.1a shows a diagram of the structure of the w^a insertion. In w^a there are novel transcripts which initiate at the *white* transcriptional startsite, but appear to terminate within one or other of the copia LTRs (O'Hare *et al.*, 1984; Pirrotta and Brockl, 1984; Zachar *et al.*, 1985). The copia insertion in w^a is such that transcription of the element is in the same direction as *white*, so the transcriptional termination signals within the LTRs are orientated in the appropriate manner. Also detectable were small amounts of an apparently wild-type sized *white* transcript, presumably formed by readthrough transcription and splicing out of copia sequences along with the second intron. Zachar *et al.* (1985) also present evidence for transcripts which appear to initiate within the 3' copia LTR, and continue through the rest of the *white* gene. Several different effects may therefore be responsible for the mutant phenotype of w^a. The simplest is premature termination of *white* transcription in the copia LTRs, and inefficient splicing of any readthrough transcript formed. In addition, any copia/*white* chimaeric transcripts initiating from the copia 3' LTR might possibly be able to act as a template for the production of a truncated *white* protein which might retain some biological function. The relative contributions of these effects are unknown, but it is relevant to note that w^{aR59K1}, a derivative of w^a which retains a single copia LTR at the site of insertion, has a phenotype intermediate between wild type and that of w^a (Carbonara and Gehring, 1985; Mount *et al.*, 1988). Premature *white* transcript termination in an appropriately orientated LTR also appears to be the case in *white*^{buff}, a B104 insertion in the fourth intron of *white* (O'Hare *et al.*, 1984). Bingham and Chapman (1986) have presented evidence that the insertion of a blood transposable element in *white*^{blood} exerts its mutagenic effects through temperature-dependent effects on transcript splicing efficiency.

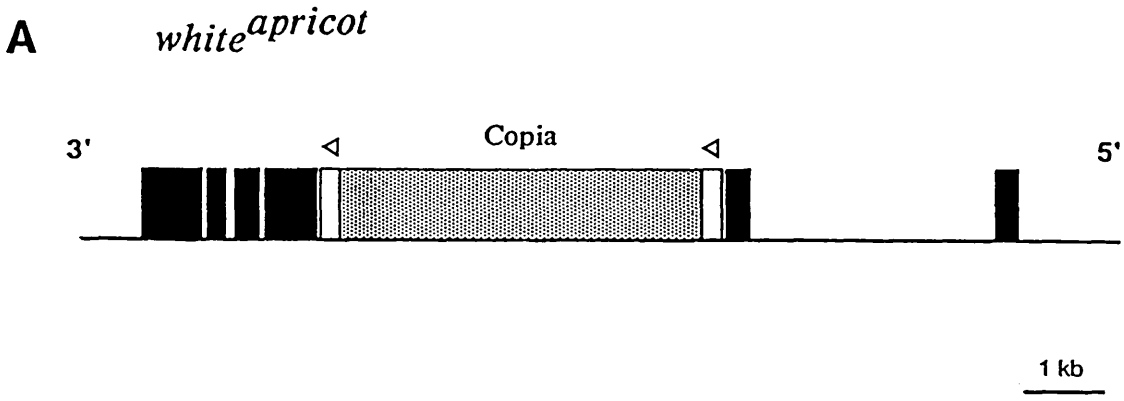
Figure 1.1 Structures of the transposable element insertion alleles *white*^{apricot}, *yellow*^{2} and *forked*^{1}

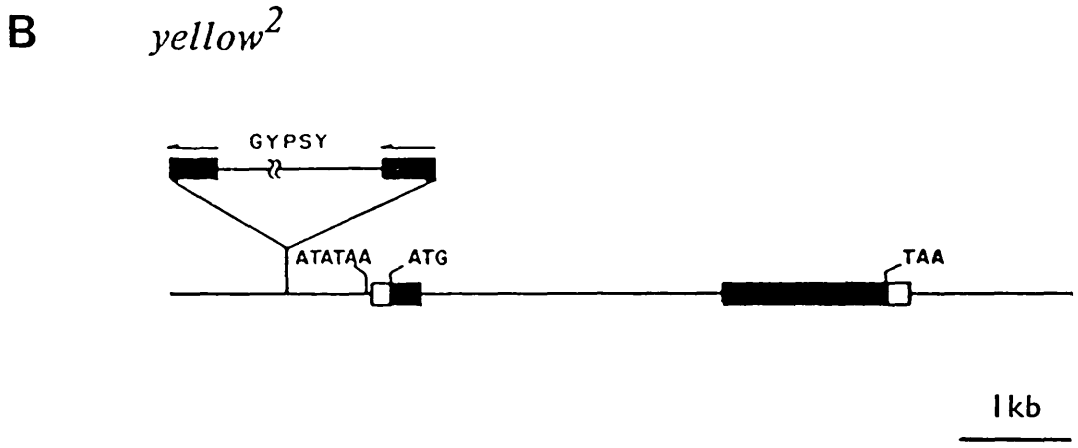
Part A shows the structure of the copia insertion in w^{a}. The filled boxes represent *white* exons, the stippled box is the internal region of copia, and the open boxes are copia LTRs. The direction of *w* transcription is from right to left, and the 5' to 3' transcriptional orientation of the copia element is shown by the arrows above the LTRs. (Taken from Mount *et al.*, 1988).

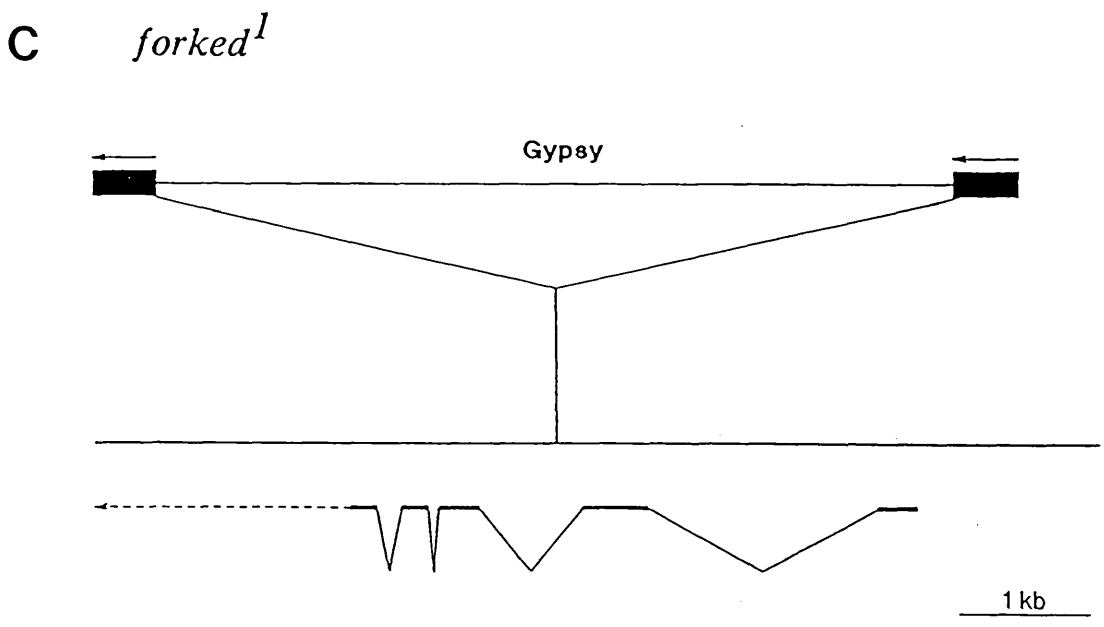
Part B shows the structure of the gypsy insertion in y^{2}. The boxed regions on the restriction map are the two exons of *y*; filled boxes represent coding regions, open boxes represent untranslated 5' and 3' regions. The direction of *y* transcription is from left to right; the 5' to 3' transcriptional orientation of the gypsy element is shown by the arrows above its LTRs. (Taken from Geyer *et al.*, 1986).

Part C shows the structure of the gypsy insertion in f^{1}. The intron/exon structure of the *forked* transcript is shown; the structure of the 3' end of the transcript is not known (dotted line). The direction of *f* transcription is from right to left; the 5' to 3' transcriptional orientation of the gypsy element is shown by the arrows above its LTRs. (S.Parkhurst and V.Corces, pers.comm.).

FIGURE 1.1


A
white^apricot
3'
Copia
5'
1 kb


B
yellow^2
GYPSY
ATATAA
ATG
TAA
1kb


C
forked^1
Gypsy
1 kb

The final type of mutagenic effect is shown by insertions into regulatory, non-transcribed sequences. These are often hypomorphs, or affect some more subtle manifestation of wild type gene function, as in the case of *white*^{spotted1} (O'Hare *et al.*, 1984). Insertions of this type may act by physically disrupting regulatory sequences, as appears to be the case for w^{sp1}, which interferes with the effects that the *zeste* gene product has on *white* expression. The product of the *zeste*$^{+}$ gene is a protein that binds specifically to sequences near the site of the B104 insertion in w^{sp1} (Benson and Pirrotta, 1987).

Other transposable element insertions into regulatory, non-transcribed sequences involve a mechanism termed "promoter interference". The concept of transcriptional interference derives from experimental observations that the activity of a vertebrate retroviral LTR may be decreased if another transcriptionally active promoter is present nearby (Cullen *et al.*, 1984; Emerman and Temin, 1984). In chicken lymphomas induced by avian leukosis virus (ALV), the ALV retroviral provirus inserts upstream of the c-myc gene; neoplastic transformation is thought to result from the 3' retroviral LTR promoting readthrough transcription of the adjacent c-myc gene, resulting in abnormal levels of c-myc transcription. Yet in all ALV induced chicken lymphomas examined, the ALV provirus upstream of the c-myc gene is deleted such that its 5' LTR is absent (Cullen *et al.*, 1984). This would suggest that the 3' LTR cannot promote transcription if the 5' LTR is itself transcriptionally active. Further evidence for such "promoter interference" comes from studies using in vitro constructed retrovirus vector containing two genes, each with its own promoter (Emerman and Temin, 1984). The products of each gene could be independently selected for. In cells containing this provirus, selecting for the expression of the 5' gene suppressed the 3' gene, and vice versa. This experimental system for studying promoter interference shows that the "suppressing" promoter may be either upstream or downstream of the cis-inactivated target gene.

The mutagenic effects of some transposable element insertions may therefore depend upon the juxtaposition of sequences carrying an active promoter with the cellular gene, resulting in promoter interference. An example of this is w^{DZL}, a 13kb insertion 5kb 5' to the *white* gene which consists of a 6.5kb unique DNA segment with an FB element flanking each end (Zachar and Bingham, 1982; Levis and Rubin, 1982). Bingham and Zachar (1985) have provided evidence for a transcript in adult heads which initiates within the central unique portion of the w^{DZL} transposon and terminates at the 3' end of the *white* transcriptional unit. In w^{DZL}, therefore, the mutant phenotype appears to arise from the presence of a promoter in the transposable element leading to transcription through the downstream cellular gene, thereby disrupting its activity.

Promoter interference does not necessarily involve transcription from the transposable element promoter. In the $yellow^2$ mutation a gypsy element is inserted 700bp 5' to the *yellow* transcription unit (see Figure 1.1b), (Parkhurst and Corces, 1986a; Chia *et al.*, 1986). It is not a simple amorph, but only affects *yellow* function in certain tissues at certain times (Nash and Yarkin, 1974). Simple disruption of regulatory sequences is not the cause of the mutant phenotype, as Geyer *et al.* (1986) have characterised a fully wild-type revertant of y^2 which retains a single gypsy LTR at the site of the original insertion. Furthermore, readthrough transcription is also discounted as the transcriptional orientation of the gypsy is opposite to that of *yellow*. Instead, Parkhurst and Corces (1986a) have noted that the peak of gypsy transcription during pupation correlates with the period when transcriptional activity of *yellow* in y^2 is lacking. They propose that the gypsy insertion causes developmentally-specific transcriptional interference of *yellow* activity, possibly by the suppression of activity of the nearby promoter by alterations in local chromatin structure.

1.3 TRANSPOSABLE ELEMENTS AND MODIFIER GENES

1.3.1 SUPPRESSORS AND ENHANCERS IN *D.MELANOGASTER*

Modifier genes are genetically defined by mutations which worsen (*enhancers*) or partially or completely reverse (*suppressors*) the phenotypes of unlinked mutant genes. A large number of genes of this type, mostly suppressors, have been found in *D.melanogaster* (Lindsley and Grell, 1968; Suzuki *et al.*, 1976). At least for some *D.melanogaster* modifier genes, the mutants whose phenotypes are modulated are associated with transposable element insertions. In a classic study, Lewis (1949) examined 207 mutant alleles at 123 loci for their ability to be suppressed by a mutation of *suppressor-of-Hairy-wing* [*su(Hw)*, 3-54.8]. Of these, 15 alleles at 11 different loci were suppressed. Approximately half of the 207 mutations were X-ray induced, yet all the suppressible alleles (except one) arose spontaneously. Modolell *et al.* (1983) then showed that most mutant alleles suppressible by mutations of *su(Hw)* were insertions of the same transposable element: gypsy. Therefore the genetic interaction is not due to a specific interaction between *su(Hw)* and each of the genes where affected loci occur. Instead, it appears to result from the alleviation of the mutagenic effects of gypsy insertions in the absence of wild type *su(Hw)* gene product.

Other allele-specific, recessive modifier genes in *Drosophila melanogaster* include *suppressor-of-forked* [*su(f)*, 1-65.9], *suppressor-of-white-apricot* [$su(w^a)$, 1-0.1], *suppressor-of-sable* [*su(s)*, 1-0.0], *suppressor-of-purple* [*su(pr)*, 3-95.5] and *enhancer-of-white-eosin* [$e(w^e)$, 1-32]. A number of alleles at 17 different loci whose phenotypes are known to be affected by mutations at one or more of the six modifier loci described. A large number of observations of the effects of modifier loci on modifiable alleles has been collected in Table 1.1. At first sight, there appears to be no underlying pattern in the matrix of interactions, yet a number of useful conclusions can be made from the data in Table 1.1:

Table 1.1 Interactions between modifier genes and affected mutant alleles

Shown are the effects of six *D.melanogaster* modifier mutations (including *su(f)*) on a number of susceptible alleles. The table indicates whether the mutant allele is suppressed, enhanced or unaffected by mutations at each modifier locus; blank spaces indicate that the interaction has not been tested.

S - suppresses
E - enhances
O - no effect

Also shown for each affected allele is the identity of the transposable element insertion with which it is associated (where known). For a number of alleles, this has been established by cloning; for others the association has been inferred from in situ hybridisation of transposable element probes to polytene chromosomes : these cases are marked with an asterisk (*).

The data for this table comes largely from Rutledge *et al.* (1988), except for the following cases indicated by superscripts:

1. Lewis (1949).
2. D.Thierry-Mieg (pers.comm.).
3. M.Crosby, cited in Craymer (1980).
4. K.Louis (pers.comm.).
5. Holmgren (1984).
6. Lewis (1981).
7. D.Kuhn, cited in Modolell *et al.* (1983).
8. T.Cullingford, M.Goodell and K.O'Hare (unpublished).
9. Green (1959a).
10. A.Driver and K.O'Hare (unpublished).
11. Jacobson *et al.* (1982).
12. Lindsley and Grell (1968).

TABLE 1.1

	su(Hw)	*e(w^e)*	*su(f)*	*su(s)*	*su(w^a)*	*su(pr)*	
y^2	S	O	O	O	O	O	gypsy
Hw^1	S	O	O			E	gypsy
sc^1	S	O	O	O	O	E	gypsy
sc^{D1}	S^1					E^2	gypsy
dm^1	S^1		S^2			O^2	gypsy*
ct^6	S	O	O	O	O	O	gypsy
ct^K	S		S	E	E	E	gypsy
lz^1	S	S	S	E	E	S	gypsy*
lz^{34}	O	O	O	E	E	S	
lz^{37}	E	O	E	E	O	O	
lz^k	O	S	O	E	E	S	
r^{sP1}	S^3					E^2	
r^{sP2}	S^3						mdg3^4
f^1	S	S	S	E	E	S	gypsy
f^5	S	S	S			S	gypsy
Bx^2	S^1			O^2			gypsy
h^1	S^5						gypsy
bx^{34e}	S	S	S	E	E	E	gypsy
bx^3	S			E		E	gypsy
bxd^1	S^1						gypsy
bxd^{51j}	S^6						gypsy
bxd^{55i}	S^6						gypsy
bxd^K	S^7						gypsy
ci^1	S^1			O^2		E^2	
ci^w	O^2			E^2		O^2	
w^a	O	O	E	O	S	O	copia
w^e	O	E	O		O	O	Doc+pogo^8
w^h	O^1	E^9	O^9		O^9		Doc+B104^{10}
v^1	O		O	S	O	O	412
s^1	O	O	O	S	O	S	
pr^1	O			S		S	412*
pr^{bw}						S^{11}	412*
sp^1				S^{12}			412/B104*

- Mutations in two different modifier genes can affect some alleles in the same way, but others in different ways. For example, *su(Hw)* and *su(f)* both suppress f^1, f^5 and ct^K; but Hw^1 and y^2 are suppressed by *su(Hw)*, whereas they are unaffected by *su(f)*. This suggests that the interactions may involve similar mechanisms for different modifier mutations, but no two can operate in completely the same way.
- Mutations of different modifier genes may affect the same allele in different ways. For example, w^a is suppressed by *su(w^a)*, enhanced by *su(f)*, yet remains unaffected by *su(Hw)*, *su(s)* and *su(pr)*. Whilst different modifiers may vary in effect, this observation would indicate that certain alleles remain responsive to several different modifiers, perhaps suggesting a common fundamental basis through which their effects are made.
- Within one locus, different modifiable alleles may be affected in different ways by the same modifier mutation. For example, *su(f)* suppresses lz^1, has no effect on lz^{34}, and enhances lz^{37}. This tells us that the mechanism of modifying interactions is independent of the gene where modifiable alleles are found, but instead depends upon the precise nature of the modifiable mutation.

In a few cases, two different modifier mutations acting on an allele responsive to both have been combined to examine possible epistatic interactions. The allele w^a is enhanced by *su(f)* and suppressed by *su(w^a)*, but *su(w^a)* w^a *su(f)* flies have eyes only marginally darker than w^a itself (K.O'Hare, pers.comm.). Rutledge *et al.* (1988) have examined the effects of the combinations *su(Hw)/e(w^e)*, *su(Hw)/su(f)*, *su(Hw)/su(s)*, *su(Hw)/su(w^a)*, *su(f)/su(s)* and *su(f)/su(w^a)* on the alleles y^2, w^a, ct^6 and f^1. They found that the modifiers acted additively, and that modifiers which acted oppositely on a given allele gave an intermediate effect in when combined. Thus there are no obvious examples of epistasis between different modifier genes.

Of the six modifier loci described, four are X-linked. Assuming an even genomic distribution of such genes, this implies that other autosomal modifier loci remain to be discovered. The X-linked loci may represent the total complement for this chromosome. In a screen for new modifier loci on a $y^2\ w^a\ ct^6\ v^1\ f^1$ X chromosome, three mutations were recovered in 20,000 EMS mutagenised chromosomes, one each at *su(f)*, *su(s)* and *su(w^a)* (Rutledge *et al.*, 1988); although any new modifier gene would have had to have effects on one of the above responsive alelles.

1.3.2 THE MOLECULAR BASIS OF MODIFIABLE MUTATIONS

Of the 17 loci in where modifiable alleles occur (Table 1.1) 10 have been cloned, allowing an analysis of the molecular basis of the mutations. These loci include the *yellow - Hairy wing - scute* region of the X chromosome (Parkhurst and Corces, 1986a; Chia *et al.*, 1986; Biessman, 1986; Campuzano *et al.*, 1985, 1986), the *cut* locus (Jack, 1985), the *forked* locus (Parkhurst and Corces, 1985; McLachlan, 1986), *Beadex* (Mattox and Davidson, 1984), *hairy* (Ish-Horowicz *et al.*, 1985; Holmgren, 1984), the *bithorax* complex (Bender *et al.*, 1983), *white* (Bingham *et al.*, 1981) and *vermilion* (Searles and Voelker, 1986; Walker *et al.*, 1986). The data in most cases confirms the presence of a transposable element insertion inferred from in situ hybridisation data. The modifiable alleles at these loci are all associated with the insertion of a gypsy element, except for w^a, $w^e\ w^h$ and v^1. Using a gypsy probe for in situ hybridisation to polytene chromosomes, gypsy elements have been found associated with all alleles suppressible by *su(Hw)*, except for r^{sP1}, r^{sP2} and *Bar* (Modolell *et al.*, 1983). Other in situ data suggests that pr^1 and pr^{bw} may be 412 insertions; that s^1 is neither a 412 nor a B104 insertion; and that sp^1 may either be a 412 or a B104 insertion (Searles and Voelker, 1986). Rutledge *et al.* (1988) examined lz^1, lz^3, lz^{34}, lz^{46}, lz^{37} and lz^k for in situ hybridisation to gypsy, copia and 412 probes, but only found hybridisation at the *lz* locus for lz^1 using a gypsy probe.

In the case of the ten cloned loci, the structural details of the transposable element insertions in their modifiable alleles are as follows:

- y^2 has a gypsy insertion 700bp 5' to the *yellow* transcriptional startsite, with the transcriptional orientation of the gypsy opposite to that of *yellow* (Figure 1.1b),(Parkhurst and Corces, 1986a; Chia, 1986).

- Hw^1 has a gypsy insertion within the *achaete* transcription unit of the *achaete-scute* complex (AS-C) (transcript T5 of Campuzano *et al.*, 1986). The transcriptional orientations of this gypsy and T5 are the same; furthermore, T5 has no introns (Campuzano *et al.*, 1986). The *Hairy wing* gene does not exist as such; *Hw* mutants form a particular class of alleles of the AS-C (Garcia-Bellido, 1979). The phenotype of Hw^{BS}, a spontaneous derivative of Hw^1, is also suppressed by *su(Hw)*; in Hw^{BS} a second transposable element has inserted into the gypsy found in Hw^1 (Campuzano *et al.*, 1986). This element, 8kb in length and known as a BS element, is a transposable element with long inverted terminal repeats (Finnegan and Fawcett, 1986).

- sc^1, sc^{D1} and sc^{D2} all have gypsy insertions in the same orientation and apparently in the same site in the AS-C (Campuzano *et al.*, 1985). In addition, sc^1 has a second insertion 2kb from the gypsy which is a Sancho 1 element (Finnegan and Fawcett, 1986). Five sc^1 revertants were associated with partial deletions of the gypsy, indicating that it is the gypsy which is responsible for the mutant phenotype (Campuzano *et al.*, 1985). The site of the gypsy insertions in these *scute* alleles lies between the T2 and T3 transcripts of the AS-C, approximately 3kb 3' to the T2 transcription unit. As they lie outside any transcribed region, it is unclear which transcription unit is being affected by the insertions; a situation complicated by the uncertainty as to which transcript or transcripts contributes to AS-C function (Campuzano *et al.*, 1985).

- ct^6 and ct^K are both associated with gypsy insertions, but the sites of insertions lie 75kb apart; DNA lesions associated with *cut* alleles cover an interval of approximately 150kb (Jack, 1985). In addition to ct^6 and ct^K, another twelve *ct* mutants associated with gypsy insertions are at least partially suppressible by *su(Hw)* (Jack, 1985). The size of the *cut* locus and its molecular complexity preclude useful conclusions as to the mutagenic effects of the suppressible gypsy insertions. Any explanation of the basis of the suppression of *ct* mutant phenotypes must be able to account for the large distances between the transcription unit described by Blochlinger *et al.* (1988) and the insertion sites of several of the suppressible gypsy insertions.

- f^1 and f^5 both have gypsy insertions of the same orientation and same position; in addition, f^5 has a second gypsy is inserted ~1kb from the first (Parkhurst and Corces, 1985; McLachlan, 1986). (See Figure 1.1c for structure of f^1). The gypsys of f^1 and f^5 lying in apparently the same position are inserted in an intron of the 2.4kb *forked*$^+$ mRNA; both they and the *f* gene are in the same transcriptional orientation (V.Corces, pers.comm.). The second gypsy in f^5 lies within the same transcription unit, and may or may not be inserted in an intron.

- Bx^2 is associated with a gypsy insertion (Mattox and Davidson, 1984). There may be no *Beadex* gene as such; the *Beadex* phenotype is thought to result from over-expression of the *heldup-a* gene (Lifschytz and Green, 1979; Mattox and Davidson, 1984). The position of the gypsy insertion relative to the *heldup-a* transcription unit is unknown.

- h^1 has a gypsy inserted in unknown orientation ~4.5kb 5' to the start of the *hairy* transcription unit (Holmgren, 1984). This gypsy also contains some non-gypsy middle-repetitive DNA of unknown identity (Holmgren, 1984).

- *bx*. Several spontaneous alleles of the *bithorax* complex (BX-C) are modifiable alleles, and all are associated with gypsy insertions (Bender *et al.*, 1983; Peifer and Bender, 1986). The alleles bx^{34e} and bx^{3} have gypsys lying 6.5kb apart within an intron of the 75kb *Ubx* transcription unit, both in the opposite transcriptional orientation. The gypsys associated with bxd^{1}, bxd^{51j}, bxd^{55i} and bxd^{K} lie within or just 3' to the *bxd* transcription unit, two each in either orientation. Their insertion sites cover a 21kb interval. The two separate insertion groups are reflected in the phenotypic differences of *bx* and *bxd* mutations (Bender *et al.*, 1983). The complex pattern of various transcripts from the *Ubx* and *bxd* regions makes it difficult to determine how these gypsy insertions upset BX-C function; especially since there is evidence to suggest that the *bxd* region acts as a cis-regulator of *Ubx* transcription (Duncan, 1987).

- w^{a} is due to a copia insertion in the 74bp second intron of the *white* gene, in the same transcriptional orientation (Figure 1.1a), (O'Hare *et al.*, 1984). The w^{e} mutation is associated with a complex insertion close to the 5' end of the transcribed region. The insertion consists of a 200bp pogo element inserted near one end of a Doc element (T.Cullingford, M.Goodell and K.O'Hare, unpublished), which is itself orientated such that its poly-A tract is at the 3' end with respect to *white* transcription. The w^{h} mutation (another derivative of w^{1}, like w^{e}) consists of a B104 element inserted into a Doc element, which is itself inserted in the same orientation and position as in w^{e} (A.Driver and K.O'Hare, unpublished).

- v^{1} has a 412 insertion very close to or within the *vermilion* transcription unit (Searles and Voelker, 1986; Walker *et al.*, 1986). The exact structural relationship between the insertion and the transcript is unknown. The alleles v^{2} and v^{k} are also suppressible by *su(s)*, and have structures apparently identical to that of v^{1} (Searles and Voelker, 1986; Walker *et al.*, 1986).

The unifying feature of modifiable alleles is that almost all are associated with insertions of only one class of *D.melanogaster* transposable element: the copia-like or retroviral-like class. The only exceptions are the mutations w^e and w^h, and the *rudimentary* allele r^{SP1}. Each of the *w* alleles is associated with a composite insertion of two transposable elements. The w^e allele is associated with pogo, a family of elements with short inverted repeats, and Doc, a retroposon; the w^h allele is associated with a B104 element (retroviral-like) inserted into a Doc element. Although suppressible by *su(Hw)*, r^{SP1} is not associated with a gypsy insertion. Instead, it appears to be a point mutation (K.Louis, pers.comm.). Also, the transposon insertion of a modifiable allele can either lie within the transcribed region or 5' to it, and apparently in either orientation (cf. suppressible *bxd* alleles).

1.3.3 MODIFIER GENES IN OTHER EUKARYOTES

In mice, the coat-colour mutation *dilute* is suppressed in trans by the recessive mutation *dilute suppressor* (Sweet, 1983). Jenkins *et al.* (1981) have shown that the *dilute* mutation is associated with the insertion of the provirus of an endogenous leukaemia retrovirus; a genetic element structurally related to *Drosophila melanogaster* copia-like elements. The relationship of this insertion to the *dilute* transcript is not known, but revertants which leave behind a single retroviral LTR have been described (Copeland *et al.*, 1983).

Similar suppressors have also been described in the yeast *S.cerevisiae*, which act on mutations caused by the insertion of Ty elements. Winston *et al.* (1984a) isolated mutations in seven unlinked genes (the *SPT* genes) which suppress the phenotypes of Ty and Ty soloLTR insertions in the 5' noncoding region of the *HIS4* gene. Further characterisation of the mechanism of suppression has shown that several *SPT* genes are regulators of Ty element activity. For example, Winston *et al.* (1984b) have shown that the *SPT3* gene is required for normal LTR to LTR

transcription of Ty elements; in its absence transcription starts at a point 800bp into the central region of the element. In the *his4-912delta* mutant, transcription from the solo Ty LTR disrupts transcription from the downstream *HIS4* gene, producing the mutant phenotype. In an *SPT3* mutant background, the aberrant transcript is not produced, thus alleviating the *HIS4* mutation. Other *SPT* genes are also regulators of Ty expression, including *SPT2* (Roeder *et al.*, 1985) and *SPT6* (Clark-Adams and Winston, 1987). *SPT4*, *SPT5* and *SPT6* mutations are able to suppress soloLTR insertions, but have little effect on Ty insertion mutations and do not affect Ty transcription (Clark-Adams and Winston, 1987). Both *SPT2* and *SPT3* have now been cloned (Roeder *et al.*, 1985; Winston and Minehart, 1986), permitting a closer analysis of the functions of their products.

Most recently, Fassler and Winston (1988) have identified another six new *SPT* genes. Two of these, *SPT13* and *SPT14*, form another class of suppressor genes; mutants at these loci suppress complete Ty insertions but not those of soloLTRs (Fassler and Winston, 1988). Thus in the absence of wild type *SPT* gene products, Ty transcription is reduced or otherwise altered, providing an explanation for the suppression of Ty induced mutations by these suppressors. Further analysis of *SPT* loci raises the possibility of vastly improved knowledge of the manner by which cellular genes affect Ty activity.

1.3.4 THE MECHANISM OF SUPPRESSION AND ENHANCEMENT

It is unlikely that modifier mutations act through qualitative changes in the translational machinery, as do classical tRNA suppressors in prokaryotes (Murgola, 1985). Like tRNA suppressors, *D.melanogaster* modifier mutants are allele-specific in their effects, but this specificity is correlated with the presence of a transposable element insertion in the affected mutant allele (Modolell *et al.*, 1983; Kubli, 1986). Mutations at the six modifier loci under discussion are all recessive; one would expect partial or complete dominance if a tRNA-suppressor type

model of translational modification were invoked. The same applies to models involving post-translational, indirect biochemical effects of mutant products of modifier genes (Suzuki *et al.*, 1976). Jacobson *et al.* (1982) have proposed a model of this type for suppression by *su(s)* mutants; prompted by the prevalence of (easily identifiable) eye colour mutants amongst modifiable alleles; such genes often encode pigment biosynthetic enzymes.

Instead, the available evidence suggests that modifier mutations act pre-translationally. The most fruitful avenue of investigation has been an examination of the transcriptional activity of modifiable alleles both with and without the presence of a modifier mutation. The gypsy in Hw^1 is inserted in the middle of an AS-C transcription unit which produces a 1.6kb wild type mRNA; in the mutant there is a truncated message of 0.9kb concomitant with termination in the upstream gypsy LTR (Campuzano *et al.*, 1986). The 0.9kb RNA is greatly more abundant than its wild type counterpart, and in Hw^1; *su(Hw)* flies the level of the 0.9kb mutant transcript is reduced. These observations suggest that the excess function Hw^1 phenotype is linked to high levels of an aberrant transcript which retains some biological activity, and that *su(Hw)* mutants suppress by reducing the level of the aberrant transcript.

In the case of w^a, most *white* transcripts terminate in one or the other LTR of the inserted copia, but there are low levels of wild type 2.6kb w^+ mRNA. In keeping with their phenotypic effects, mutants of *su*(w^a) increase the levels of this RNA. Mutants at *su(f)*, which enhances w^a, decrease its abundance (Levis *et al.*, 1984). In this allele, the modifiers thus appear to act either by affecting the degree of premature termination in the copia LTRs, or by affecting the efficiency of splicing out the copia insert along with the second intron. Mount *et al.* (1988) have studied the responsiveness of partial revertants of w^a to *su*(w^a) and *su(f)*. Partial revertants where the copia has departed leaving behind a solo LTR (w^{aRsLTR}

alleles) are no longer responsive to either modifier, a result confirmed by germ line transformation experiments using an in vitro constructed w^{aRsLTR} derivative. This shows that sequences of the internal portion of copia are necessary both for the severity of the w^a phenotype and for its response to the two modifiers. The derivative w^{aR84h} results from the insertion of the 83 most 3' nucleotides of the I factor very close to the 3' end of the copia element. The phenotype of w^{aR84h} is enhanced by *su(f)*, but remains unaffected by *su(w^a)*, suggesting that these modifiers have different sites of action within the copia.

The *yellow* locus encodes a 2.1kb mRNA present during the early larval and pupal periods of development; pupal levels of this message are greatly reduced in y^2 flies - in agreement with the observation that y^2 affects adult, but not larval pigmentation (Parkhurst and Corces, 1986a). Mutations at *su(Hw)*, which suppresses y^2, causes pupal amounts of the *yellow* mRNA to return to wild-type levels. They suggest that the reason why y^2 only affects the adult phenotype is linked to the transcriptional activity of the inserted gypsy: gypsy transcripts are present at low levels in larval and adult stages, but are greatly more abundant in pupae, when the expression of genes for pigmentation of adult structures occurs. If promoter interference forms the basis of the y^2 phenotype, then suppression would result from a reduction of the transcriptional activity of the gypsy element.

The *forked* locus encodes several transcripts which are only present in pupal stages (Parkhurst and Corces, 1985); the period of bristle formation (Lees and Waddington, 1942). In f^1 flies, the putative 2.4kb f^+ mRNA is greatly reduced in abundance; but in *su(Hw)* f^1 and *su(f)* f^1 flies it is restored to wild type levels (Parkhurst and Corces, 1985). Both of these modifiers suppress f^1. There are apparently no novel transcripts in f^1 flies deriving from premature termination within either gypsy LTR, so the gypsy therefore acts either by affecting the activity of the *forked* promoter, or by affecting the efficiency of splicing to produce an mRNA of

wild type structure.

Modifiable alleles are thus associated with transposable element insertions that act by altering the levels of gene transcription, or production of wild type message. To the above list may be added the 412 insertion in v^1: the 1.5kb *vermilion*$^+$ mRNA is not detectable in v^1 flies (Walker *et al.*, 1986). Parkhurst and Corces (1985, 1986a, 1986b, 1986c; Kubli, 1986) have proposed that the transcriptional activation of a gypsy insertion in a modifiable allele is the key to the molecular basis of the modifying effects in these cases. Gypsy transcripts are most abundant during the pupal period, when *forked* and *yellow* are active, suggesting that the mutagenic effects of the y^2 and f^1 insertions involves promoter interference: the active gypsy promoter repressing the activity of the nearby cellular gene promoter. If the wild type product of a modifier gene acted to stimulate gypsy expression, then in the absence of that product gypsy expression would be reduced, alleviating the effects of the promoter interference.

Evidence that modifier gene products do regulate gypsy expression comes from studies of the relative abundance of gypsy transcripts in modifier$^+$ and modifier$^-$ genetic backgrounds. In *su(Hw)* flies, gypsy RNAs are apparently 5-fold less abundant than in *su(Hw)*$^+$ flies (Parkhurst and Corces, 1986a). Such experiments are complicated by the need to express gypsy activity in terms of transcripts per gypsy copy as the two stocks may have differing numbers of gypsy elements. In addition, different gypsy copies may not have equivalent transcriptional activities, depending on their chromosomal location. In a *su(f)* mutant background, gypsy transcripts appear to become more abundant. The disparate effects of *su(Hw)* and *su(f)* mutants on gypsy transcription are not what one might expect, given that *su(Hw)* and *su(f)* suppress the same *forked* alleles. Yet Table 1.1 shows that these modifiers cannot act in entirely the same manner, as *su(Hw)* suppresses several mutations that *su(f)* does not. The conclusion that *su(Hw)* and *su(f)* act as

regulators of gypsy activity is strengthened by observations of the effects of *SPT* mutations upon Ty transcription (see section 1.3.3).

Another approach in the investigation of the molecular basis of modifying interactions is to discover the function of the modifier gene products themselves. The genes *su(s)* (Chang *et al.*, 1986), *su(w^a)* (Zachar *et al.*, 1987a) and *su(Hw)* (S.Parkhurst and V.Corces, pers.comm.) have been cloned. The best characterised of these is *su(w^a)*, whose wild type product appears to play a role in regulating splicing. Of the three transcripts produced by the *su(w^a)* gene, only the smallest, fully-spliced RNA has the structure of an mRNA (Chou *et al.*, 1987). Reverse genetic analysis shows that only the fully-spliced transcript is able to supply *su(w^a)$^+$* function (Zachar *et al.*, 1987b). Furthermore, the product of the fully-spliced RNA appears to be responsible for the repression of splicing at *su(w^a)* itself; point mutant alleles of *su(w^a)* accumulate the fully-spliced RNA (Zachar *et al.*, 1987b). If the *su(w^a)$^+$* product plays a role in regulating the splicing of other genes, then the suppression of w^a by mutants of *su(w^a)* may be due to effects on splicing. If the absence of *su(w^a)* product can lead to more efficient splicing at *su(w^a)* itself, then it might also lead to more efficient splicing of the enlarged *white* second intron in w^a, thus producing more of a functional *w* message. However, some results of Mount *et al.* (1988) are difficult to explain in this context. The w^a partial revertant w^{aR84h} differs from w^a by the insertion of 83bp of I factor DNA into the copia, yet it is not responsive to *su(w^a)*. This suggests that the nature of the inserted sequences must also be taken into account. Other effects of the copia insertion in w^a probably need to be considered to fully understand the phenotypic suppression by *su(w^a)*.

Investigation of *su(s)* and *su(Hw)* is continuing; the most interesting result in this respect to date is that the putative protein product of the *su(Hw)* gene encodes multiple copies of the Zinc finger protein sequence motif (S.Parkhurst and

V.Corces, pers.comm.); this sequence motif has been found in a variety of DNA binding proteins (Klug and Rhodes, 1987). Originally characterised in *Xenopus* TFIIIA (Miller *et al.*, 1985), such motifs have been found for example in the mammalian Sp1 protein (Kadonaga *et al.*, 1987) and the *D.melanogaster* genes *serendipity* (Vincent *et al.*, 1985), *Kruppel* (Rosenberg *et al.*, 1986) and *hunchback* (Tautz *et al.*, 1987). Thus the *su(Hw)* protein product might regulate gypsy transcriptional activity in some way following direct binding to gypsy DNA.

Interactions with transposable element associated mutations are not the only feature of modifier gene mutations, several show phenotypes suggesting that they have other cellular functions. Some alleles of *su(Hw)* are recessive female steriles (Lindsley and Grell, 1968). The allele $su(pr)^{B}$ is male sterile, female partially sterile and shows a range of other mutant phenes including irregular eyes and wing venation, low viability, late hatching and a short life span (Lindsley and Grell, 1968). The original $e(w^{e})$ allele isolated by Green was female sterile and had a *plexate*-like wing venation phenotype (Lindsley and Grell, 1968), although the mutation $e(w^{e})^{S}$, whose effects are described in Table 1.1 and is almost certainly an allele of the former, shows neither phenotype (Rutledge *et al.*, 1988). All available alleles of $su(w^{a})$ are viable and show no other phenotypes, but Zachar *et al.* (1987a) claim that they are leaky alleles; they found several presumptive P element or EMS induced $su(w^{a})$ alleles which showed reduced fertility and life span in heterozygous females, and are attempting to isolate possible lethal $su(w^{a})$ alleles. Alleles of *su(s)* have no other apparent phenotype than their modifying effects (R.A.Voelker, pers.comm.; Lindsley and Grell, 1968). The *su(f)* gene itself is a lethal locus and is the most studied modifier in terms of its other mutant phenotypes; these studies and their implications for the cellular function(s) of the *su(f)* product are discussed in the next section.

1.4 THE *SUPPRESSOR OF FORKED* LOCUS

The first allele of *su(f)*, $su(f)^1$, was isolated by M.Whittinghill in an X-ray mutagenesis experiment (Whittinghill, 1937, 1938). Flies of the genotype f^1 $su(f)^1$ showed virtually complete suppression of the *forked* bristle phenotype. Since then, allele-specific enhancing and modifying effects have been found on unlinked mutations at *forked*, *lozenge*, *white*, *diminutive*, *cut* and and *bithorax* (Table 1.1). The mutation $su(f)^1$ suppresses f^4 and f^5 (Green, 1955); suppresses lz^1 and enhances lz^{37} (Snyder and Smith, 1976); enhances w^a (Green, 1959a); suppresses dm^1 (D.Thierry-Mieg, pers.comm.); and suppresses ct^6 and bx^{34e} (Rutledge *et al.*, 1988). In addition to these modifying effects, the *su(f)* alleles available show a range of other mutant phenotypes.

1.4.1 THE LETHAL FUNCTION OF *su(f)*

Several recessive lethal *su(f)* alleles exist, such as $su(f)^{l\text{-}3DES}$ and $su(f)^{l\text{-}D13}$ (Schalet and Lefevre, 1973), thus identifying *su(f)* as a vital locus. There are also four temperature-sensitive lethal alleles: $su(f)^{l\text{-}ts67g}$ (Dudick *et al.*, 1974), $su(f)^{l\text{-}ts726}$ (Russell, 1974), $su(f)^{l\text{-}ts76a}$ (Wilson, 1980) and $su(f)^{l\text{-}mad.ts}$ (Jurgens and Gateff, 1979). When grown at 25°C, all four ts alleles are viable but suppress f^1. In addition, $su(f)^{l\text{-}3DES}$ fails to complement $su(f)^1$ in terms of suppression: f^1 $su(f)^1/f^1$ $su(f)^{l\text{-}3DES}$ females are viable, but are $forked^+$ in appearance (Schalet and Lefevre, 1973). Temperature shift experiment showed that all four ts alleles have approximately the same temperature sensitive period for lethality from late 1st/early 2nd instar to just after puparium formation (Dudick *et al.*, 1974; Russell, 1974; Wilson, 1980; Jurgens and Gateff, 1979). This differs from the temperature sensitive period of $su(f)^{l\text{-}ts67g}$ for suppression of forked bristles, which lasts for 24 hours, starting just prior to the initiation of bristle formation, 30 hours after puparium formation (Dudick *et al.*, 1974).

For the alleles $su(f)^{l\text{-}ts726}$ and $su(f)^{l\text{-}mad.ts}$, analysis of mutant clones derived

from X-ray induced somatic recombination showed that the lethal phenotype is cell-autonomous (Russell, 1974; Jurgens and Gateff, 1979). If subjected to brief pulses of the restrictive temperature in late larval life, both alleles give adults with abnormal, partially or wholly duplicated cuticular structures such as legs and head parts (Russell *et al.*, 1977; Jurgens and Gateff, 1979). In $su(f)^{l\text{-}ts726}$, Russell (1974) found histological evidence of cell death following heat treatment; this was confined mostly to the few larval structures which retain mitotic activity, such as the imaginal discs, ganglion cells and abdominal histoblasts. He proposed that the mutation was a cell autonomous lethal specific for mitotically active cells, and that the partial or complete pattern duplications resulted from imaginal discs compensating for the missing cells. In gynandromorphs, pattern duplications were found in wild type tissue if it were beside mutant tissue in the same disc, suggesting that even though the mutation is cell autonomous, it could affect pattern formation in a non-autonomous way. In contrast, Jurgens and Gateff (1979) arrived at a different conclusion after studying $su(f)^{l\text{-}mad.ts}$. Using a different histochemical technique from Russells, they could find no evidence of cell death in imaginal discs of heat treated larvae. They concluded instead that the mutation affected the development of cells, rather than killing them rapidly, based on observations on the appearance of somatic $su(f)^{l\text{-}mad.ts}$ clones induced in a *Minute* background. The $su(f)^{l\text{-}mad.ts}$ clones survived longer and were larger in a $Minute^{-}$ background than in a $Minute^{+}$ background. Their data suggest that retarded cell growth in the mutant leads to aberrant positional specification of imaginal disc cells, ultimately yielding abnormal or duplicated adult structures.

1.4.2 *su(f)* FUNCTION AND FEMALE FERTILITY

In addition to their f^{1}-suppressing and lethal phenotypes, ts alleles of *su(f)* also exhibit ts female sterility. Wilson (1980) has examined the female-sterile phenotype of $su(f)^{l\text{-}ts76a}$. When placed at the restrictive temperature, $su(f)^{l\text{-}ts76a}$ females stop laying eggs after 4 to 5 days. During this period, an

increasing proportion of the eggs layed are abnormal: they lack a chorion and are small and miss-shapen. Some 4 days after being returned to the permissive temperature, the laying of morphologically normal cells resumed. On examining the ovaries of mutant females which had spent 5 days at the restrictive temperature, he found no distinct egg chambers, an apparent lack of of follicle cells and debris suggestive of dead follicle cells. Ovarian follicle cells are mitotically very active, thus providing further evidence that actively dividing cells are most affected by lack of *su(f)* function. Transplanting $su(f)^{l\text{-}ts76a}$ ovaries into wild type females showed that the defect was ovary autonomous. Cell death at the restrictive temperature could account for the presence of dead follicle cells; yet the reversibility of the female sterility phenotype shows that not all the follicle cells are killed, or that they are only temporarily afflicted. It is unclear whether $su(f)^{l\text{-}ts76a}$ affects gametic development in both sexes. Male mutant flies placed at the restrictive temperature also showed a decrease in fertility, but this is complicated by the fact that Oregon R males also show a similar susceptibility to increased temperatures. With $su(f)^{l\text{-}ts726}$, Russell (1974) showed that the restrictive temperature had no effect on male fertility.

Dudick *et al.* (1974) found a similar phenotype of reversible ts female sterility in $su(f)^{l\text{-}ts67g}$, along with abnormal ovaries containing degenerating oocytes in sterile females. Lineruth and Lambertsson (1986) have analysed the conditional female sterility of $su(f)^{l\text{-}ts67g}$ in more detail, following the morphological degeneration of egg chambers after shifting females to the restrictive temperature. Histological examination gave no clear-cut answer as to whether cell death is occurring in the egg chambers. Dead cells were found, but not in association with any particular cell type or oocyte stage. Only after 5 days at the restrictive temperature did the number of dead cells increase dramatically. This time postdates the onset of sterility, so the abnormal function and morphology does not appear to be linked to cell death in the first instance.

Is *su(f)* required for the production of viable oocytes? This question may be addressed by an experimental technique using the dominant female sterile mutation *Fs(1)K1237* (Perrimon and Gans, 1983). X-ray induction of somatic clones in the progeny of *Balancer / l(1)* females crossed to *Fs(1)K1237* males will give *l(1) / l(1)* clones in a *Fs(1)K1237* background. As this dominant female sterile is blocked in oogenesis, any eggs which are produced therefore result from homozygous *l(1)* germ cells and one may conclude that the lethal has no function in oocyte production. Using $su(f)^{l\text{-}S2}$, $su(f)^{l\text{-}3DES}$ and $su(f)^{l\text{-}X2}$, such germ line clones are lethal (N.Perrimon, pers.comm.), therefore *su(f)* is required for the production of viable oocytes. It remains unclear whether *su(f)* mutations have a maternal effect.

1.4.3 THE *su(f)* "DEFICIENCY PHENOTYPE"

Several trans-heterozygous combinations of *su(f)* alleles exhibit a characteristic collection of mutant phenes known as the "deficiency phenotype" (Schalet, 1968; Schalet and Lefevre, 1973). It is usually characterised by absent or reduced bristles; in some cases the flies also have rough eyes with an anterior indentation, a rough abdomen and wings which are blistered and misshapen. This phenotype has similarities to that of *Minute* mutations (Lindsley and Grell, 1968), and is sometimes known as *Minute*-like (Schalet, 1972). The bristle phenotype is not related to suppression of *forked* bristles; the deficiency phenotype is seen in flies with a $forked^{+}$ genotype. In general, $su(f)^{1}$ / $Df(1)su(f)^{1}$ heterozygotes show this phenotype at 18-25°C; they are lethal at 29-30°C (Schalet and Lefevre, 1976). The phenotype is also found in the following combinations :

$su(f)^{l\text{-}D13}$ / $su(f)^{1}$	at 25°C	(Schalet and Lefevre, 1973)
$su(f)^{l\text{-}X2}$ / $su(f)^{1}$	at 25°C	"
$su(f)^{l\text{-}3DES}$ / $su(f)^{1}$	at 29-30°C	"
$su(f)^{1}$ homozygotes	at 29-30°C	"
$su(f)^{l\text{-}ts726}$ homozygotes	at 25°C	(Russell, 1974)
$su(f)^{l\text{-}ts67g}$ / *Df(1)su(f)*	at 25°C	(Dudick *et al.*, 1974)
$su(f)^{l\text{-}ts67g}$ / $su(f)^{1}$	at 29°C	"

The basis for the phenotype presumably results from abnormalities in or death amongst the larval progenitor cells of these adult structures. The induction of *su(f)*$^{l\text{-}mad.ts}$ clones in gynandromorphs or by somatic recombination gave adult cuticle whose bristles ranged from wild type, through "*Minute*-like" to completely absent, so the deficiency phenotype may be a less extreme manifestation of the *su(f)* lethal phenotype. If this is correct, the fact that *su(f)*^{1} over a deficiency can yield the deficiency phenotype suggests that this otherwise viable allele is hypomorphic for the vital function of *su(f)*.

1.4.4 *su(f)* AND ECDYSONE - AN ASIDE

When shifted to the restrictive temperature before the end of the second instar, *su(f)*$^{l\text{-}ts67g}$ and *su(f)*$^{l\text{-}mad.ts}$ larvae fail to pupate and eventually die after 1 to 3 weeks (Lambertsson and Fekete, 1981; Klose *et al.*, 1980). This phenotype is similar to that observed in another temperature sensitive recessive lethal, *ecdysone-1*^{ts}, characterised by Garen *et al.* (1977). Analysis of this mutant has revealed the major role that the insect steroid hormone ecdysone plays throughout *D.melanogaster* development. In *ecd-1*^{ts}, the 12-fold rise in 20-OH-ecdysone content at pupariation seen in wild type flies does not occur. Klose *et al.* (1980) found that *su(f)*$^{l\text{-}mad.ts}$ larvae at the restrictive temperature have a much reduced ecdysteroid content, with no true titre peaks as observed in wild type. Furthermore, developmentally arrested *su(f)*$^{l\text{-}mad.ts}$ larvae could be induced to pupate after feeding with exogenous 20-OH-ecdysone, although they died soon after. Thus corrct ecdysone metabolism and/or function is dependent upon *su(f)*$^{+}$ function. Experiments with *ecd-1*^{ts} have shown that some ecdysone synthesis in *D.melanogaster* occurs in the ovary (Garen *et al.*, 1977), as well as in the ring-gland (King, 1970). It has been suggested that within the ovary, it is the follicle cells which make ecdysone, as in other insects (reviewed by Hoffman *et al.*, 1980). That adult female ovarian follicle cells are affected by *su(f)* has been described (section 1.4.2), so if their larval progenitors are similarly affected, this may provide an

explanation for the "*ecdysone-1*^{ts}-like" phenes seen in *su(f)*$^{l\text{-}mad.ts}$.

Hansson and Lambertsson (1984) have shown that arrested *su(f)*$^{l\text{-}ts67g}$ third instar larvae can also form pseudopupae if fed ecdysterone. It has also been shown that at the restrictive temperature, such larvae are deficient for salivary glue protein synthesis (Hansson *et al.*, 1981). The *sgs* family of genes produce the protein glue which, when excreted by the pupariating larva, will eventually fix the puparium to a solid support (Beckendorf and Kafatos, 1976; Korge, 1977). Cloning has shown that the chromosomal locations of the *sgs* genes are associated with puff sites in the polytene chromosomes of the salivary gland; e.g. *sgs4* (Muskavitch and Hogness, 1980) and the *sgs3/sgs7/sgs8* complex (Meyerowitz and Hogness, 1982). These third instar puff sites regress just prior to pupariation in response to a sudden rise in ecdysterone levels, with a concomitant shut-down in production of *sgs* proteins (reviewed in Ashburner and Berendes, 1978). Hansson and Lambertsson (1983) found that *sgs* transcripts were absent in third instar *su(f)*$^{l\text{-}ts67g}$ larvae at the restrictive temperature. Feeding ecdysterone to such larvae induced *sgs* transcription, thus ecdysterone can also act as a positive regulator of *sgs* activity. The control of *sgs* expression by ecdysone therefore appears to depend upon a correct developmental timetable, which is itself at least partly determined by ecdysone action. The expression of *sgs* genes is therefore not directly under the control of *su(f)* function, but it does provide a good demonstration of how mutations at *su(f)* may indirectly affect a wide range of cellular functions. The primary defect in *su(f)* mutants responsible for the abnormal ecdysone titres remains unknown.

1.4.5 INTRA-ALLELIC COMPLEMENTATION AT *su(f)*

Another interesting allele of *su(f)* is *su(f)*$^{R\text{-}9\text{-}18}$ (also known as *su(f)*^{pbl} - pale bristles), characterised by Schalet (1972). At 25^{o}C *su(f)*$^{R\text{-}9\text{-}18}$ homozygotes die at pupariation, but at 17-18^{o}C they eclose showing the following phenoytpe: thread-

like, pale yellow bristles and hairs, dark pigment on the dorsal anterior thorax and wrinkled or curled wings. The males are sterile. It suppresses f^1 at 17-18°C and in $su(f)^{R-9-18}$ / $su(f)^1$ homozygotes f^1 is suppressed at all temperatures from 17°C to 29°C. Although it is allelic to *su(f)* with respect to *forked* suppression, females of the genotype $su(f)^{R-9-18}$ / $su(f)^{l-3DES}$ are fully viable at all temperatures and show no signs of the pale bristle phenotype. At 18°C, $su(f)^{R-9-18}$ / $su(f)^{l-X2}$ flies are fully viable and appear normal; at 25°C viability is decreased, with signs of the pale bristles phenotype; and at 29°C they die before pupariation. At 17-18°C, $su(f)^{R-9-18}$ / $su(f)^{l-D13}$ flies are fully viable, although some bristle and hairs of all individuals show the pale bristles phenotype; at 25°C they die at pupariation. Though $su(f)^{R-9-18}$ differs from other *su(f)* alleles in these respects, it is not unique: the allele $su(f)^{pb2}$ (pale bristles 2) has a phenotype similar to that of $su(f)^{R-9-18}$ (A.Schalet, pers.comm.).

In combination with $su(f)^{l-3DES}$, $su(f)^{R-9-18}$ thus shows complete intra-allelic complementation for *su(f)* lethality. For $su(f)^{l-X2}$ and $su(f)^{l-D13}$, the complementation is partial. Schalet has studied the complementation pattern for viability and bristle phenotype at 25°C in a large number of combinations of 11 alleles of *su(f)* (A.Schalet, unpublished data); the results are shown in Table 1.2. They show that $su(f)^{R-9-18}$ only complements some lethal *su(f)* alleles for lethality and not others. All lethal *su(f)* alleles are viable when heterozygous with $su(f)^1$, although several combinations show the deficiency phenotype, as described above. Except for $su(f)^{R-9-18}$, the lethal alleles almost always fail to complement one another. The two exceptions to this rule are $su(f)^{l-X3}$ / $su(f)^{l-X1}$ and $su(f)^{l-X3}$ / $su(f)^{l-3DES}$; however, there is some doubt as to whether $su(f)^{l-X3}$ is truly a *su(f)* lethal. When first isolated, it was not completely lethal (A.Schalet, pers.comm.). It is a *su(f)* allele of some form though, as f^1 $su(f)^{l-X3}$ / f^1 $su(f)^{R-9-18}$ flies have a bristle phenotype which is $forked^+$ (A.Schalet, pers.comm.).

Table 1.2 Complementation pattern for viability between various *su(f)* alleles

Shown is a matrix of heterozygous combinations between various mutant alleles of *su(f)*, scored for viability and bristle phenotype. All crosses were performed at 25°C. For clarity, each allele is represented by its superscript: i.e. $su(f)^{1}$ is *1*, $su(f)^{l\text{-}X1}$ is *X1*, etc. All the lethal alleles are viable in combination with $su(f)^{1}$; and all the lethal alleles are lethal in combination with each other except for $su(f)^{l\text{-}X3}$ / $su(f)^{l\text{-}X1}$ and $su(f)^{l\text{-}X3}$ / $su(f)^{l\text{-}3DES}$ (see text). (Data from A.Schalet, pers.comm.).

TABLE 1.2

	1	*X1*	*X2*	*X3*	*L26*	*16-1-85*	*3DES*	*D13*	*S1*	*S2*	*R-9-18*
1	+	+	M	+	M	M	+	M	M	+	+
X1		L	L	+[1]	L	nt	L	L	nt	nt	+
X2			L	L	L	nt	L	L	L	L	+
X3				L	nt	nt	+	L	L	nt	+[2]
L26					L	nt	L	L	L	nt	L
16-1-85						L	nt	L	nt	L	L
3DES							L	L	L	L	+
D13								L	L	nt	L
S1									L	nt	L
S2										L	+
R-9-18											L

+ = viable, bristles normal
M = viable, bristles *Minute*-like (deficiency phenotype)
L = lethal
nt = not tested

1 - Only three $su(f)^{l\text{-}X3}$ / $su(f)^{l\text{-}X1}$ females observed; one female had one bristle that was *pale bristle* in appearance

2 - $su(f)^{l\text{-}X3}$ / $su(f)^{R\text{-}9\text{-}18}$ females have some bristles which are *pale bristle* in appearance

Does the *su(f)* locus consist of one gene or several? At first sight, the pleiotropic effects of mutaions at *su(f)* might suggest a locus consisting of more than one gene. Yet the existence of partial intra-allelic complementation does not necessarily mean more than one cistron. In the monocistronic *white* locus there is intra-allelic complementation between $w^{spotted1}$ and certain other *white* alleles (Green, 1956; O'Hare *et al.*, 1984), although the comparison is not strictly fair as *white* is not a lethal gene. The fact that <u>all</u> types of *su(f)* alleles fall into a single complementation group in regard to their ability to suppress f^1 means that the various phenotypes seen in *su(f)* mutants might result from lesions in a single gene product with several functions, or a single gene producing several related proteins with different functions.

1.4.6 *su(f)* AND THE PROXIMAL X CHROMOSOME

The proximal region of the X chromosome of *D.melanogaster* has been extensively mutagenised in order to study the genetic loci in polytene chromosome sections 19EF and 20 (reviewed by Schalet and Lefevre, 1976). Figure 1.2 shows a map of the known genetic loci in polytene chromosome section 20, including *su(f)*. The most proximal known "euchromatic" complementation group is *su(f)*, placed in section 20DE, and lying at the boundary of euchromatic and heterochromatic sequences. The terms heterochromatin and euchromatin were first applied to cytologically distinguishable regions of eukaryotic chromosomes (Heitz, 1928, 1929). In *D.melanogaster*, the heterochromatic regions of late prophase chromosomes have a very condensed appearance; these regions form the proximal half of the X chromosome, the proximal quarters of chromosomes 2 and 3 and the entire Y chromosome. In diploid cells, heterochromatin is delayed in replication by comparison to euchromatin (Barigozzi *et al.*, 1966). Euchromatin and heterochromatin also have different replicative properties during the formation of polytene chromosomes (reviewed by Gall, 1973). Euchromatic regions replicate to give a characteristic banding pattern of chromomeres (reviewed by Lefevre, 1974).

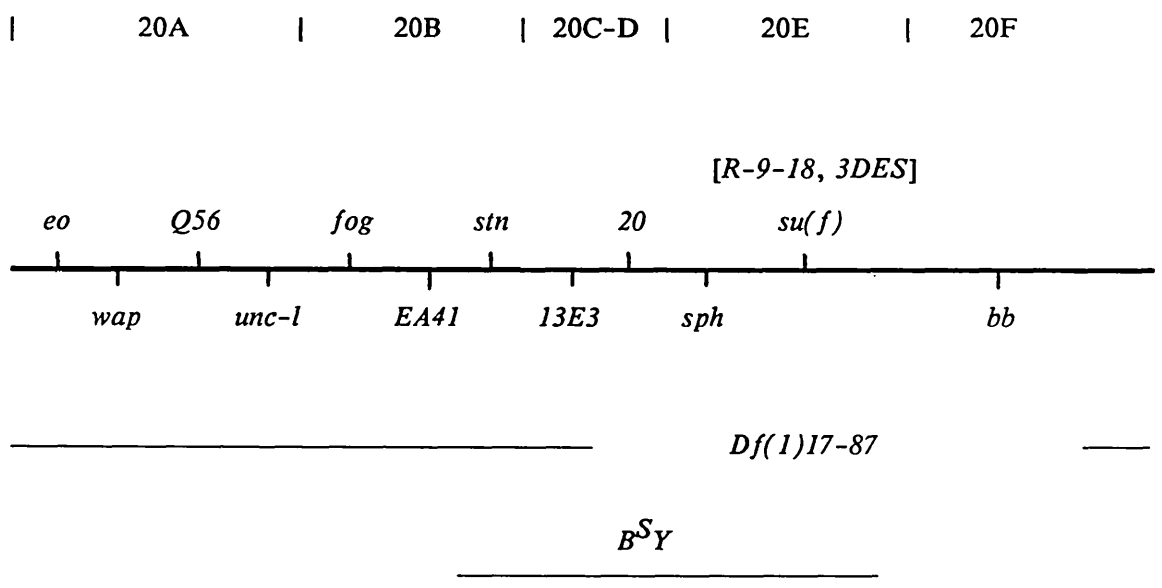
Figure 1.2 The proximal X chromosome : genetic loci in polytene section 20

Shown are the known genetic loci in polytene section 20 of the X chromosome of *D.melanogaster*. Above the genetic map is given the extents of the subdivisions of section 20. Below the map are shown the extent of the deletion *Df(1)17-87* and the region covered by the $B^{S}Y$ translocation. (After Schalet and Lefevre, 1976; Lindsley and Zimm, 1986).

The names of the loci shown are as follows:

eo : *extra organs*
wap : *wings apart*
uncl : *uncoordinated-like*
fog : *folded gastrula*
stn : *stoned*
sph : *sparse hairs*
su(f): *suppressor-of-forked*

FIGURE 1.2


20A
20B
20C-D
20E
20F
[R-9-18, 3DES]
eo
Q56
fog
stn
20
su(f)
wap
unc-l
EA41
13E3
sph
bb
Df(1)17-87
$B^{S}Y$

The bulk of heterochromatin, given the name α-heterochromatin by Heitz (1934), does not replicate (Rudkin, 1969), and forms the chromocentre where the euchromatic chromosome arms fuse. Where the euchromatic arms meet the chromocentre are poorly banded regions, named β-heterochromatin (Heitz, 1934), which do replicate (Gall, 1973).

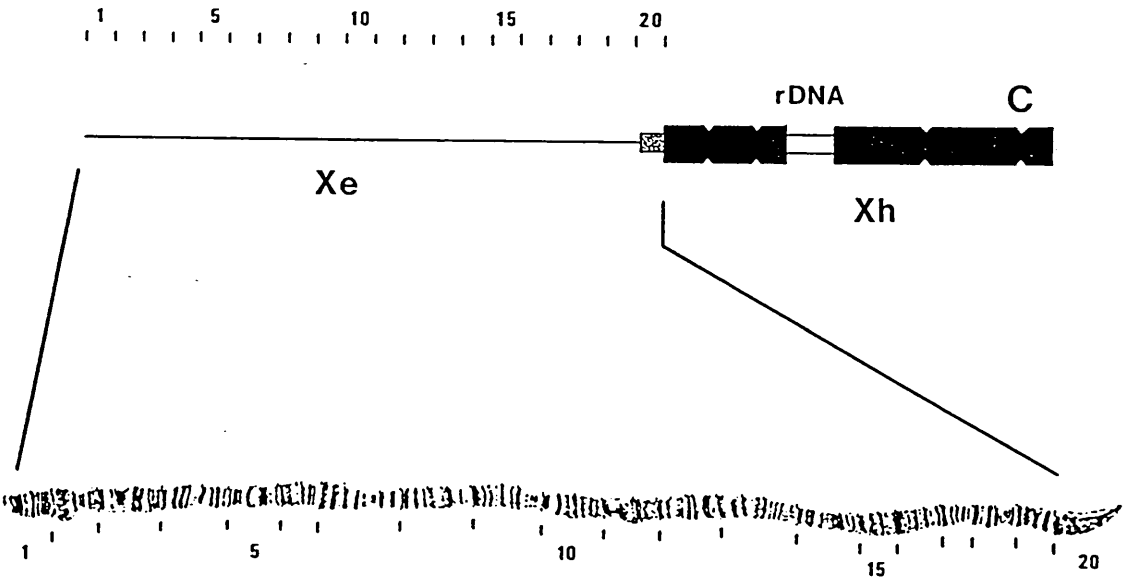
Figure 1.3 is a diagram of the euchromatic/heterochromatic structure of the X chromosome, showing the position of the β-heterochromatin. X chromosome β-heterochromatin corresponds to polytene section 20B-F, which contains some 10 known genetic loci, including *su(f)* (Schalet and Lefevre, 1976). The heterochromatic proximal half of the *D.melanogaster* X chromosome is almost completely devoid of genetic loci, and consists mostly of repetitive DNA sequences. Much of this comes in the form of satellite sequences, multiple tandem repeats of short (4-10bp) sequences in large arrays, which are restricted to heterochromatin (Gall *et al.*, 1971). Proximal to *su(f)*, the only well characterised genetic locus is *bobbed*: 200-250 tandem copies of the rDNA repeat (Figure 1.3), (Tartof, 1973), which lies approximately in the middle of the proximal heterochromatin (Pimpinelli *et al.*, 1986). The rDNA repeats are not unique to the X chromosome, but are also found on the Y chromosome. Another genetic "locus" in the heterochromatin of the X is the *ABO* element, which interacts with the maternal effect mutation *abnormal oocyte*, which lies between *su(f)* and *bb* (reviewed by Pimpinelli *et al.*, 1986). The probability of survival of abo^+ progeny of *abo/abo* mothers is affected by the presence or absence of the *ABO* elements in the X heterochromatin and the heterochromatin of the right arm of chromosome 2; such progeny are more likely to die if one or the other is deleted. In addition, if one of the *ABO* elements is absent, only about 50% of the progeny of abo^+ mothers survive. The function of the *ABO* elements and the basis of their interaction with *abo* remain mysterious.

Figure 1.3 The structure of the X chromosome

Shown is a diagram of the X chromosome of *D.melanogaster*, with euchromatin (Xe) represented by the thin line, and heterochromatin (Xh) by the filled boxes. The open box in the middle of Xh is the rDNA repeat cluster (the *bobbed* locus), and C represents the centromere. The stippled box at the Xe/Xh junction (polytene section 20B-F) represents β-heterochromatin. The positions of the cytologically visible restrictions in diploid metaphase chromosomes are shown.(Adapted from Pimpinelli *et al.*, 1986).

Beneath the diagram is a photographic representation of a polytene X chromosome from a third instar larval salivary gland, showing the characteristic chromomeral banding pattern divided into 20 sections. Only the euchromatic portion of the X is multiply replicated and therefore visible. (Taken from Lefevre, 1976).

FIGURE 1.3



Little is known of the sequence organisation of β-heterochromatin. It has been suggested that a high proportion of β-heterochromatin may consist of mobile element sequences (Ananiev *et al.*, 1984; Young, 1979). There is almost no meiotic recombination in the heterochromatic portion of the X chromosome (Schalet and Lefevre, 1976), which has led Charlesworth *et al.* (1986) to suggest that repetitive DNA sequences may accumulate in chromosomal regions which are undisturbed by crossing over. If such regions are recombinationally undisturbed, repetitive sequences might escape removal from the genome by processes of stochastic loss.

In a pioneering study, Miklos *et al.* (1988) microdissected and microcloned the β-heterochromatin transition region of the X chromosome. The clones obtained contained a high proportion of many different repetitive sequences, including sequences homologous to mobile elements, interspersed with single copy sequences. In situ hybridisation to polytene chromosomes showed that some of the repetitive sequences are characteristic for β-heterochromatin. Some of the repetitive DNA sequences which were not recognised may be examples of previously uncharacterised mobile elements. In addition, some sequences may not be recoverable in a microclone library due to the nature of the microcloning technique. For example, no satellite DNA sequences were found in the microclone collection, when by in situ hybridisation to polytene chromosomes, Singh *et al.* (1981) found the GATA minor satellite in polytene sections 19F and 20AB. Nevertheless, the analysis of the microcloned DNA sequences shows that β-heterochromatin is clearly different in sequence composition and organisation from euchromatin. As the most proximal "classical" genetic locus on the X chromosome, an analysis of the DNA sequence organisation around the *su(f)* locus may illuminate the peculiar structure of the euchromatin-heterochromatin transition region.

1.5 AIMS OF THE PROJECT

As the preceding has shown, the cloning and molecular analysis of the *su(f)* locus is attractive for several reasons. Firstly, *su(f)* is of interest as a member of a genetically curious class of trans-acting modifier genes. A study of the molecular structure of *su(f)* therefore forms an approach to understanding how one gene affects the expression of a second, physically unlinked gene. Secondly, the *su(f)* locus is interesting in its own right as a cell-autonomous lethal gene, especially with regard to its apparent specificity for mitotically active cells, and to its effects on female fertility and ecdysone metabolism. Thirdly, *su(f)* is located in a region of the X cromosome which is structurally intriguing, yet poorly characterised. Cloning a locus located within β-heterochromatin, at the junction of euchromatic and heterochromatic DNA, would provide an opportunity for examining the local DNA sequence organisation, and as a defined starting point for possible chromosomal walking, it would allow for further study of this chromosomal region.

The primary aim of this work was to obtain molecular clones of DNA from the *su(f)* locus. The strategy used to clone *su(f)* was that of P element transposon tagging (Bingham *et al.*, 1982). Once obtained, flanking DNA sequences from the P element insertion allele would allow the isolation of wild type genomic clones of *su(f)*, thus permitting analysis of the wild type locus with respect to sequence organisation in this region of β-heterochromatin. Cloned DNA probes would also facilitate the examination of the structure of mutant alleles of *su(f)* and the pattern of transcription of the locus in both wild type and mutant alleles. Cloned DNA probes from the *su(f)* locus could also be used to isolate cDNA clones of potential $su(f)^{+}$ transcripts; further analysis of such clones might help to shed some light on the role of any protein products of *su(f)*.

CHAPTER 2

MATERIALS AND METHODS

2.1 BACTERIAL STRAINS

Plasmids and lambda bacteriophages were propagated in *E.coli* strain C600 : F^-, *thi-1, thr-1, leu-B6, lac-Y1, ton-A21, supE44,* λ^- (Appleyard, 1954), or K802 : $hsdR^-$, $hsdM^+$, gal^-, met^-, *supE* (Wood, 1966). For *spi* selection of λred^-gam^- recombinants the strain Q359 was used: $hsdR^-$, $hsdM^+$, *supE, Φ80, P2* (Karn *et al.*, 1980). To grow recombinant λgt10 bacteriophage, the strain C600 *hflA* was used (Hoyt *et al.*, 1982). M13 bacteriophage and plasmids requiring blue-white selection for recombinants were propagated in *E.coli* JM101 : *delta(lac, pro), supE,* thi^-, *F'traD36, proAB,* $lacI^q$, *Z delta M15* (Messing *et al.*, 1981).

2.2 BACTERIAL MEDIA

Liquid cultures of *E.coli* were grown in Luria broth (L-broth) which contains per litre: 5g NaCl, 5g yeast extract (Difco) and 10g Bacto-tryptone (Difco). L-plates were made by adding 15g per litre of Bacto-agar (Difco) to L-broth and autoclaving. M13 and λ bacteriophage were plated onto L-plates in an overlay of top agar containing the following per litre: 10g BBL trypticase (BBL Microbiology Systems), 5g NaCl and 6g Bacto-agar. Blue-white selection of M13 recombinants due to the insertional inactivation of the *lacZ* gene was performed by adding 10ul 0.1M IPTG and 25ul 0.2% X-GAL in dimethylformamide to the top agar of each plate. For blue-white selection when cloning using the plasmid vector pBluescribe M13-, the same amounts of IPTG and X-GAL along with 4ul of 60mg.ml^{-1} ampicillin were added to a pre-poured layer of L-agar on L-plates containing ampicillin. Genomic libraries in λ vectors were plated in an overlay of L-agarose; L-broth containing 0.7% agarose (FMC BioProducts) and 10mM $MgCl_2$, onto a support of λ agar containing 5g NaCl, 10g Bacto-tryptone, 5g yeast extract, 12g Bacto-agar and 120mg $MgSO_4$ per litre. For dilution and storage of bacteriophage λ, phage buffer was used: 10mM tris.HCl pH7.5, 10mM $MgSO_4$, 50mM NaCl, 0.001% gelatin. Antibiotics were prepared and used as described by Maniatis *et al.* (1982).

2.3 PREPARATION OF DNA

2.3.1 SMALL SCALE PREPARATION OF PLASMID DNA

Bacterial colonies were inoculated into 2ml L-Broth containing 50-70ug.ml^{-1} ampicillin and grown overnight with shaking at 37^{o}C. Small scale preparations of plasmid DNA were carried out using 1ml of the overnight cultures according to the method of Birnboim and Doly (1979), as adapted by Burke and Ish-Horowicz (1981).

2.3.2 LARGE SCALE PREPARATION OF PLASMID DNA

Large scale plasmid DNA preparations were carried out by the above method as modified by D.Ish-Horowicz. Plasmid cultures (200-500ml) were grown overnight with shaking at 37^{o}C in L-broth containing 50-70ug.ml^{-1} ampicillin. Cells were harvested by centrifugation in a Sorvall GS-3 rotor at 7,000rpm for 5 minutes. The pellet was resuspended in 38ml of 0.9% glucose, 25mM tris.HCl pH8, 10mM EDTA and then 40ml of 0.2M NaOH, 1% SDS added. After 5 minutes, 20ml of ice-cold 29.5% potassium acetate, 11.4% glacial acetic acid was added and the mixture stored on ice for 10 minutes. After centrifugation at 8,000rpm for 5 minutes the supernatant was filtered through muslin and 0.6 volumes of isopropanol added and then centrifuged again at 8,000rpm for 5 minutes. After draining, the pellet was resuspended in 8.1ml T.E.(10,1), 0.9ml of 3M Na acetate added and transferred to a 30ml Corex tube where 15ml of ethanol was added. The precipitate was pelleted at 5,000rpm for 5 minutes in a Sorvall SS-34 rotor and resuspended in 9ml T.E.(10,1); 12g CsCl and 1ml of 10mg.ml^{-1} ethidium bromide were then added and the sample centrifuged in Oakridge tubes at 10,000rpm for 30 minutes to pellet excess RNA. The supernatant was transferred to an ultracentrifuge tube, made up with 1.55g.ml^{-1} CsCl if necessary and then centrifuged in a Beckman Vti.50 or 50.Ti rotor at 45,000rpm for 24-48 hours. The plasmid band was visualised under UV illumination and extracted using a needle and syringe. The ethidium bromide was removed by repeated extractions with isopropanol saturated with CsCl-saturated T.E.(10,1) and then the plasmid sample was dialysed exhaustively in T.E.(10,1). Plasmid DNA

concentrations were determined spectrophotometrically (Maniatis *et al.*, 1982) and adjusted to 200ug.ml^{-1}. Yields were typically 1-4mg plasmid DNA.

2.3.3 LARGE SCALE PREPARATION OF LAMBDA DNA

An aliquot (2-3ml) of C600 grown overnight in L-broth containing 10mM $MgCl_2$ and 0.2% maltose was infected with λ bacteriophage at a multiplicity of 0.01-0.1 by incubating at 37^{o}C for 15 minutes. The infected cells were added to 250-400ml of pre-warmed L-broth containing 10mM $MgCl_2$ and shaken at 37^{o}C for 4-7 hours until lysis occurred. After the addition of 3ml chloroform and 30 minutes further shaking, the mixture was centrifuged in a Sorvall GS-3 rotor at 5,000rpm for 10 minutes. To the supernatant was added NaCl to 6% and PEG 6000 (BDH) to 7%, and the mixture left overnight at 4^{o}C. The PEG precipitate was pelleted by centrifugation at 5,000rpm for 30 minutes at 4^{o}C, resuspended in 3ml of phage buffer and transferred to a 15ml Corex tube (Du Pont) to which was added 30ul of 1mg.ml^{-1} DNase I (Sigma) and 30ul of 1mg.ml^{-1} RNase (Sigma). After incubating at 37^{o}C for 1 hour the sample was extracted with 3ml chloroform and then loaded on the top of a CsCl density gradient consisting of 2ml layers of 1.6, 1.5, 1.4 and 1.3g.ml^{-1} CsCl made up in 10mM tris.HCl pH7.5, 10mM $MgCl_2$ (TM10). The sample was then centrifuged in a Beckman SW-41Ti rotor at 25,000rpm for 2.5-3 hours at 4^{o}C. The phage band was extracted using a needle and syringe and dialysed extensively against TM10. The phage sample was made 20mM with EDTA and 0.1% with SDS and then extracted once with phenol, once with 1:1 phenol/chloroform and once with chloroform. After ethanol precipitation, washing with 70% ethanol and drying, the λ DNA pellet was resuspended in T.E.(10,1). Total yields were typically 100-500 ug λ DNA.

2.3.4 PREPARATION OF *DROSOPHILA* DNA

Fresh or frozen adult flies (100-500mg) were weighed and added to 5ml of homogenisation buffer: 0.1M NaCl, 30mM tris.HCl pH8, 10mM EDTA, 7.7mM 2-

mercaptoethanol and homogenised in an electrically driven 30ml Dounce homogeniser. The homogenate was filtered through muslin into a 30ml Corex tube. The homogeniser was rinsed with a further 2ml of homogenisation buffer and this was also filtered into the Corex tube. Nuclei were pelleted in a Sorvall HB-4 rotor at 4,000rpm for 10 minutes at $4^{o}C$ and the pellet resuspended in 1ml homogenisation buffer and transferred to an Eppendorf tube. After centrifugation for 1 minute in a microcentrifuge the pellet was resuspended in 0.3ml of lysis buffer: 30mM tris.HCl pH8, 100mM EDTA and 30ul of $1mg.ml^{-1}$ proteinase-K (Sigma) and 30ul of 10% N-lauroyl sarcosinate added. After incubation at $45^{o}C$ for 2-3 hours, the sample was extracted once with 0.3ml 1:1 phenol/chloroform, once with 0.3ml chloroform and the DNA precipitated by the addition of 15ul 5M NaCl and 0.8ml ethanol. The pellet was washed twice with 70% ethanol, air-dried and resuspended in 0.8ml T.E.(10,1) per 1g of original weight of flies, giving a DNA concentration of $\sim 200ug.ml^{-1}$.

2.4 PREPARATION OF *DROSOPHILA* POLY-A+ RNA

All aqueous solutions and automatic disposable pipette tips used were autoclaved before use. Glass pipettes and Corex tubes were baked overnight at $250^{o}C$. Adult flies, pupae, larvae or embryos (2-4g) were placed in a 50ml sterile plastic Falcon tube containing 5ml 100:1 chloroform:isoamyl alcohol, 5ml neutralised phenol and 10ml homogenisation buffer: 0.15M Na-acetate, 5mM EDTA, 1% SDS, 50mM tris.HCl pH9, $20ug.ml^{-1}$ polyvinyl sulphate (Sigma). The mix was homogenised using a tissue disrupter (Polytron Kinematica) for 2 bursts of 30 seconds each. After spinning at full speed in a bench-top centrifuge for 5 minutes, the lower phase was removed and 5ml phenol and 5ml 100:1 chloroform:isoamyl alcohol added. The mix was vortexed and then centrifuged and the lower phase removed as above. This extraction was repeated, after which the upper phase was transferred to a 30ml Corex tube and 20ml 96% ethanol added. After storage overnight at $4^{o}C$, the precipitate was pelleted in a Sorvall HB-4 rotor at 10,000rpm for 15 minutes at $4^{o}C$, then resuspended in 4ml SDS buffer: 10mM tris.HCl pH7.5, 100mM NaCl, 1mM EDTA, 0.5% SDS containing

0.3mg.ml^{-1} proteinase-K (Sigma) and incubated for 2 hours at 37^{o}C. The sample was extracted twice with 2ml phenol and 2ml 100:1 chloroform:isoamyl alcohol and then twice with 4ml 100:1 chloroform:isoamyl alcohol with vortexing and centrifugation at 5,000rpm for 10 minutes in a Sorvall HB-4 rotor. The aqueous phase was then transferred to a 15ml Corex tube and ethanol precipitated. The precipitate was pelleted as above and resuspended in 3ml T.E.(10,1) containing 0.5% SDS.

Poly-A^{+} RNA was selected by chromatography through oligo-dT cellulose (Collaborative Research) according to the method of Levis and Penman (1978). The poly-A^{+} RNA obtained was ethanol precipitated and resuspended in SDS buffer. Yields were between 50 and 100ug per gram of starting material.

2.5 ETHANOL PRECIPITATION OF NUCLEIC ACIDS

Nucleic acids were precipitated from solution by the addition of 0.1 volumes of 3M Na-acetate and either 2 volumes (DNA precipitation) or 2.5 volumes (RNA precipitation) of 96% ethanol. Precipitates were pelleted by centrifugation in a Sorvall HB-4 rotor or in a microcentrifuge, depending on volume. Pellets were washed with 70% ethanol and either air-dried or dried in a vacuum dessicator.

2.6 PHENOL EXTRACTION OF DNA

DNA solutions were phenol extracted by adding an equal volume of neutralised phenol, vortexing and separating the layers by centrifugation. The aqueous layer was then extracted with an equal volume of 1:1 phenol:chloroform and then with an equal volume of chloroform in the same manner. Redistilled phenol was neutralised by vortexing with an equal volume of 0.3M tris base and separating the layers by centrifugation, after which the phenol was stored under a layer of T.E.(10,1) pH7.5.

2.7 RESTRICTION ENZYME DIGESTION OF DNA

Restriction enzyme digestion of DNA samples was carried out according to Maniatis *et al.* (1982) in either the Low, Medium or High salt buffers described therein to facilitate simultaneous multiple digests. The exception was digestion of DNA with SmaI, where the buffer used was 20mM KCl, 10mM tris.HCl pH8, 6mM $MgCl_2$, 0.6mM DTT. Digestions of genomic DNA were routinely carried out overnight. DNA fragments for use as molecular weight standards were generated by digesting λ *CIts857* DNA with HindIII or ΦX174 DNA with HaeIII.

2.8 LIGATION OF DNA

DNA samples were ligated in 50mM tris.HCl pH7.5, 10mM $MgCl_2$, 10mM DTT, 1mM rATP containing 0.5-5 units of T4 DNA ligase. The ligation volumes were 15ul containing 0.1-0.2ug total DNA and the reactions were carried out at 14^{o}C for either 4 hours (cohesive end ligations) or for 20 hours (blunt end ligations). Vector to target ratios were in the range 1:2 to 1:4.

2.9 AGAROSE GEL ELECTROPHORESIS

2.9.1 NEUTRAL AGAROSE GELS

Agarose gel electrophoresis of DNA was carried out in slab gels of 0.6-1.5% (w/v) agarose (FMC BioProducts) in 89mM tris borate pH8, 89mM boric acid, 2mM EDTA (TBE) containing 0.5-1ug.ml^{-1} ethidium bromide. Before loading, DNA samples were combined with a 1/6th volume of sample buffer: 0.25% bromophenol blue, 30% glycerol in H_2O. Large gels (15x22cm) were run at 2-4V.cm^{-1}; minigels (10x10cm) were run at up to 6V.cm^{-1}. DNA fragments were visualised and photographed under short-wave UV illumination. The sizes of DNA fragments were determined from a calibration curve of the migration distances vs. the fragment sizes of DNA size standards (section 2.7).

2.9.2 FORMALDEHYDE AGAROSE GELS

Formaldehyde agarose gel electrophoresis of RNA was carried out in submerged slab gels of 1-1.5% agarose in 18mM Na_2HPO_4, 2mM NaH_2PO_4 (1x gel buffer) containing 6% formaldehyde. The gels were submerged in 1x gel buffer containing 3% formaldehyde and run at $1V.cm^{-1}$ for 16-18 hours. Poly-A^+ RNA samples were ethanol precipitated and resuspended in 20ul of sample buffer: 50% formamide, 6% formaldehyde, 1x gel buffer; then combined with 20ul dye buffer: 50% formamide, 50% glycerol, 1x gel buffer, 0.1% bromophenol blue, 0.1% xylene cyanol. The samples were heated at 65°C for 5 minutes and cooled on ice before loading. Size standards were single stranded DNA fragments obtained by denaturing λ *CIts857* DNA digested with HindIII or ΦX174 DNA digested with HaeIII.

2.10 PURIFICATION OF DNA FRAGMENTS FROM AGAROSE GELS

DNA restriction enzyme digests were loaded on gels made with 0.7-0.9% low-gelling-temperature agarose (FMC BioProducts). Under UV illumination, the fragments of interest were excised from the gel using a scalpel blade and the slab placed in an Eppendorf tube. An equal volume of T.E.(10,1) was added and the agarose melted in a 65°C waterbath for 5 minutes, following which the sample was phenol extracted, ethanol precipitated and resuspended in T.E.(10,1).

2.11 NUCLEIC ACID TRANSFER AND FILTER HYBRIDISATION

2.11.1 DNA TRANSFER AND FILTER HYBRIDISATION

Transfer of DNA fragments to filters was carried out following the method of Southern (1975), with the modification that directly following electrophoresis, the DNA in the gel was partially depurinated by shaking in 0.25M HCl for 15 minutes. Prior to transfer, nitrocellulose filters (Schleicher and Schuell) were prewetted in H_2O and then 2x SSC (1x SSC is 150mM NaCl, 15mM Na-citrate). After transfer, the filters were rinsed in 2x SSC and then baked at 80°C for 2 hours. On occasions, Hybond-N filters (Amersham) were used, in which case the DNA was fixed

to the filter by irradiation with UV light. The filters were pre-hybridised in a solution of 5x SSC, 0.1% SDS at 65°C for at least 2 hours and then hybridised in 5x SSC, 0.1% SDS, 2x Denhardt's solution (1x Denhardt's is 0.2% BSA, 0.2% Ficoll, 0.2% poyvinyl pyrrolidone, all Sigma), 100ug.ml^{-1} single-stranded salmon sperm DNA. The radioactive DNA probe was denatured in a boiling waterbath for 5 minutes before addition to the filter bag, to yield a probe DNA concentration of 10-50ng.ml^{-1}. After 16-48 hours hybridisation, the filters were washed by shaking in 2x SSC, 1x Denhardt's, 0.1% SDS for 2 x 30 minutes, followed by 2 x 30 minutes in 2x SSC, 0.1% SDS, finishing with 2 x 30 minutes in 3mM tris base, 0.1% SDS; all washes were at room temperature. Filters were autoradiographed using X-ray film (RX, Fuji or XAR-5, Kodak) either at room temperature or at -70°C with an intensifying screen (Dupont Cronex Lightning Plus) (Laskey and Mills,1977).

For re-probing Southern blot filters, hybridisation probes were removed from nitrocellulose filters by boiling in water for three minutes. On occasions, the length of time elapsed between hybridisations obviated the need to remove the previous hybridisation probe. Hybridisation probes were removed from Hybond-N filters according to the manufacturers instructions (Amersham).

2.11.2 RNA TRANSFER AND FILTER HYBRIDISATION

RNA gels were transferred to filters in the same way as DNA gels except that following electrophoresis, the gels were equilibrated in 10x SSC for 20 minutes prior to setting up the transfer; and that the transfer solution used was 10x SSC. Filters were prehybridised overnight at 42°C in hybridisation solution: 50% de-ionised formamide, 5x STE (1x STE is 150mM NaCl, 30mM tris.HCl pH8, 2mM EDTA), 5x Denhardt's, 400ug.ml^{-1} single-stranded salmon sperm DNA, 0.5% SDS, 2x RNA gel buffer (see section 2.11.2). Filters were hybridised in hybridisation solution at 42°C for two overnights using radioactive single-stranded M13 DNA probes. The filters were washed at room temperature for 2 hours in four changes of 5x

Denhardt's, 2x STE, 0.5% SDS and then for 30 minutes in two changes of 0.5x STE, 0.5% SDS. Autoradiography was performed as for DNA hybridisation filters. To control for the amont of poly-A^+ sample loaded in each lane, the filters were later hybridised to a single stranded DNA probe of *ras*64B DNA, whose transcript levels remain constant throughout development (Mozer *et al.*, 1984).

2.12 RADIOACTIVE LABELLING OF DNA

2.12.1 NICK TRANSLATION

Radioactive labelling of λ DNA, plasmid DNA or DNA fragments purified by agarose gel electrophoresis was carried out by the nick translation method of Rigby *et al.* (1977). Typically 0.05 to 0.5ug of DNA and 0.5 to 2ul of alpha-^{32}P-dCTP at ~3000Ci/mmol (Amersham) was used per reaction, and unincorporated nucleotides were removed by passage down a Sephadex G-50 column using the spun-column method (Maniatis *et al.*,1982). The specific activities obtained for the probes were typically 1-2x10^8dpm.ug^{-1} DNA.

2.12.2 M13 PROBES

M13 clones of genomic DNA were used to generate radioactive single-stranded DNA probes by using the Klenow fragment of *E.coli* DNA polymerase I (Amersham) to synthesise a strand complementary to the insert in the presence of alpha-^{32}P-dCTP (Amersham) (Burke, 1984). For most experiments, unincorporated nucleotides were removed from the reaction products and the probe boiled and added to the hybridisation as for nick translated radioactive probes. However, in poly-A^+ RNA filter hybridisation experiments, purifed single-stranded radioactive DNA probes were used. The products of the extension reaction were digested with a restriction enzyme that cleaved either at the distal end of the inserted DNA fragment or at a selected site within the insert. The sample was denatured in 0.1M NaOH and then loaded onto a denaturing, 1.5% low-gelling-temperature agarose gel containing 30mM NaOH, 10mM EDTA. After electrophoresis at 3V.cm^{-1} for 16 hours, the position of

the radioactive single-stranded DNA was determined by autoradiography and the gel slab excised. The slab was melted at 65°C for 5 minutes and added directly to the hybridisation solution.

2.13 CONSTRUCTION OF GENOMIC LAMBDA LIBRARIES

Sau3aI partial-digest genomic DNA libraries were made as described in Maniatis *et al.* (1982). A total of 6ug of *Drosophila* DNA was digested with Sau3aI for various lengths of time under conditions determined empirically in a pilot experiment. The digests were separately loaded onto a 0.6% low-gelling-temperature agarose gel and after electrophoresis, all DNA fragments in the size range 14-20kb were purified from the gel. These were combined with 1-2ug of λ EMBL4 (Frischauf *et al.*, 1983) digested with BamHI and SalI, and then ligated in a 30ul volume. Aliquots of the ligation were packaged into bacteriophage particles in vitro by the method of Hohn and Murray (1977).

Limit digest libraries were made by digesting 2ug of *Drosophila* DNA with the appropriate restriction enzyme, freezing to destroy the enzyme activity and then combining the digest with 1ug λ Charon28 (Rimm *et al.*, 1980) or λ EMBL4 digested with the appropriate enzyme. Ligation and subsequent in vitro packaging were carried out as described above. In some cases the products of the limit digest were size selected by purification from low-gelling-temperature agarose gels prior to ligation.

Libraries were plated out to a density of 15-30,000 plaques per 150mm petri dish using *E.coli* C600 or K802 (λ Charon28 libraries), Q359 (λ EMBL4 libraries) or C600 *hflA* (λ gt10 libraries). Filter lifts were made as described in section 2.14.1, and following hybridisation of the filters, positive plaques were removed as cores and placed in a small volume of phage buffer. Two to three rounds of plaque hybridisation were sufficient to purify the positive plaques, after which

they were amplified by the plate-lysate technique (Maniatis *et al.*,1982), followed by large scale preparation of λ DNA.

2.14 SUBCLONING OF DNA FRAGMENTS

Subcloning was generally carried out using gel-purified DNA fragments which were ligated to M13 or plasmid vectors as described above. Plasmid vectors used were pBR327 (Soberon *et al.*, 1980) and pBluescribe M13- : a derivative of pUC19 containing bacteriophage T3 and T7 promoters flanking the cloning site polylinker (Vector Cloning Systems). M13 vectors used were M13mp8, mp9 (Messing and Vieira, 1982), mp18, mp19 (Yanisch-Perron *et al.*, 1985), tg130 and tg131 (Kieny *et al.*, 1983). Approximately 100ng of ligated DNA was used to transform *E.coli* made competent by the method of Mandel and Higa (1970).

Plasmid recombinants were generally identified by small scale plasmid preparations followed by restriction enzyme digestion and agarose gel electrophoresis. The use of pBluescribe M13- greatly facilitated the identification of recombinants through the blue-white colour selection associated with the insertional inactivation of the *lacZ* gene on the plasmid. In the case of M13 phage, single ddTTP/dTTP dideoxy sequencing reactions of several clones (T-tracking) was sufficient to identify the clone of interest. On occasions where the desired plaques or colonies formed a very small proportion of the total, the following hybridisation procedures were used.

2.14.1 PLAQUE HYBRIDISATION

M13 or λ plaques to be screened were plated using BBL top agar to a density of 50-300 plaques per 90mm petri dish. Filter lifts using nitrocellulose filters (Schleicher and Schuell) were carried out and further processed by the method of Benton and Davis (1977). After 2 hours baking at 80^{o}C, the filters were hybridised with a radioactive probe, washed and autoradiographed as described previously.

2.14.2 COLONY HYBRIDISATION

Colonies for screening were picked in duplicate into ordered arrays and one set transferred to Whatman 541 paper (Gergen *et al.*, 1979). The filter-borne colonies were denatured and neutralised as if gels for Southern transfer and then taken directly to hybridisation without baking or prehybridisation.

2.15 S1 NUCLEASE ANALYSIS OF M13 CLONE HYBRIDS

For each annealing reaction, 2ul each of two M13 clones were combined in capped tubes in a final volume of 10ul of 1x sequencing buffer (10x sequencing buffer is: 0.1M $MgCl_2$, 0.1M tris.HCl pH8) and placed in boiling water for 3 minutes. Annealing was completed by incubating at 65°C for 30 minutes. After cooling, 10ul of S1 digestion mix (2ul 10x S1 buffer, 8ul dH_2O containing 0.1-1 Weiss units of S1 (BRL) was added and the reaction incubated at 37°C for 45-60 minutes. (10x S1 buffer is: 3M NaCl, 300mM Na acetate pH4.5, 45mM Zn acetate). The reactions were stopped by adding EDTA to 20mM, and the hybrid DNA products recovered by adding 50ul 96% ethanol and precipitating at -70°C for 30 minutes. The DNA was collected by microcentrifugation for 10 minutes, washed with 70% ethanol and resuspended in a small volume of T.E.(10,1). The reaction products were analysed by electrophoresis in a neutral 0.9% agarose gel; double-stranded DNA hybrid molecules were detected by UV illumination. The sizes of the double stranded products were determined from a calibration curve of the migration distances vs. the fragment sizes of DNA size standards (section 2.7).

2.16 M13 DIDEOXY SEQUENCING

DNA sequencing was performed by the dideoxy chain terminating method of Sanger and Coulson (1979). The preparation of single stranded M13 DNA and the sequencing reactions were carried out as described fully in Bankier and Barrell (1983). The radioactive label used in the sequencing reactions was ^{35}S-dATP-alpha S at ~400 $Ci.mmol^{-1}$ (Amersham); and the enzyme used was DNA polymerase Klenow fragment

(Amersham). On occasions, clones were sequenced with Sequenase (United States Biochemicals), using the kit and following the directions provided. Sequence reactions were fractionated on 6% polyacrylamide gels (57:3 mono:bis-acrylamide) in 1x TBE buffer.

2.17 COMPUTER ANALYSIS OF SEQUENCE DATA

Sequence data from the "shotgun" strategy of sequencing a clone was assembled and analysed using the programmes of Queen and Korn (1984) contained in the Microgenie software package (Beckman Instruments) on an IBM PC microcomputer. Further analysis of sequences was performed using the ANALYSEQ (Mclachlan *et al.*, 1984), ANALYSEP (Staden, 1984) and DIAGON (Staden, 1982) programmes of the Staden-Plus software package (Amersham). Searches of the Imperial Cancer Research Fund sequence databases were performed using LSEARCH (Soundy, unpublished) on the ICRF Dec20. Searches of the Edinburgh Protein Sequence Database were carried out using the program of Collins and Coulson (J.F.Collins and A.F.W.Coulson, unpublished).

2.18 *DROSOPHILA* MEDIA AND STOCKS

2.18.1 *DROSOPHILA* MEDIA

Drosophila stocks were maintained on cornmeal, yeast, treacle food: 40ml powdered brewer's yeast (ICN), 105ml cornmeal (ICN), 60ml black treacle (Tate and Lyle) and 9g agar per litre to which was added 15ml of 10% (w/v) methyl-4-hydroxybenzoate (Nipagin) and 2ml propionic acid. To prepare food containing hygromycin B for selecting germ-line transformed flies, a 100ug.ml^{-1} solution of hygromycin B (Boehringer Mannheim) in 2x Hepes-buffered saline (16.36g NaCl, 11.9g Hepes, 0.4g Na_2HPO_4 per litre, pH 7.1) was added to food cooled to 40^{o}C to yield a hygromycin B concentration of 90-100ug.g^{-1}. Plates for egg-collection contained grape juice agar medium: 62.5g glucose, 325g sucrose, 8.5g powdered brewer's yeast, 37.5ml grape juice concentrate (Waitrose own brand), 25g agar per litre also containing Nipagin and propionic acid as above. Stocks were maintained at 18^{o}C or 25^{o}C.

2.18.2 *DROSOPHILA* STOCKS

The following list describes all the stocks used in this work together with their origins. A description of most of the genetic markers and mutations listed can be found in Lindsley and Grell (1968) or in Lindsley and Zimm (1985, 1986, 1987); for others, the appropriate reference has been given.

Stock	Obtained from	Reference
Canton-S	Bowling Green	
Oregon-R	D.Ish-Horowicz ICRF, Oxford	
π_2	W.Engels Wisconsin-Madison	Engels and Preston (1979)
In(1)N$^{264-84}$, y N$^{264-84}$ / FM6	Bowling Green	
w m f	W.Engels Wisconsin-Madison	
y cv v f car	D.Ish-Horowicz ICRF, Oxford	
H-41 (RM) *Basc, w^a B; In(2LR)bw^{V1}, ds^{33k} dp b bw^{V1} / SM1, al^2 Cy cn^2 sp^2; Sb / TM2, Ubx^{130} e^s; spa^{pol}*	M.G.Kidwell Tucson	Kidwell (1985)
bw;st	D.Ish-Horowicz ICRF, Oxford	
ry^{506}	S.Delaney Imperial College	
sc z^1 w^{op7red} ec	M.M.Green UC Davis	Lifschytz and Green (1984)
y z^a w^{ch}	M.M.Green UC Davis	
(1)8P1 / FM6; y$^+$Ymal^{106}	G.Miklos Canberra	Schalet and Singer (1971)
l(1)13E3 / FM6; y$^+$Ymal^{106}	G.Miklos Canberra	Schalet and Singer (1971)
l(1)sph^{S1} / FM6; B^{S}Yy$^+$	A.Schalet Leiden	Schalet (1986)
f^5 su(f)^1	Bowling Green	

Stock	Obtained from	Reference
f^1 $su(f)^1$	Bowling Green	
$C(1)DX$, y f / y cin w f^s $su(f)^{ts67g}$	Bowling Green	Dudick *et al.* (1974)
y v f mal $su(f)^{ts76a}$	T.G.Wilson Vermont	Wilson (1980)
y w f $su(f)^{mad-ts}$	E.Gateff Freiburg	Jurgens and Gateff (1979)
$C(1)DX$, y f / y ct^6 v f car $su(f)^{R-9-18}$; y^+Ymal^{106}	G.Miklos Canberra	Lifschytz and Falk (1969)
$su(f)^{l-L26}$ / $FM6$; B^SY	A.Schalet Leiden	isolated by G.Lefevre
y ct^6 v f $su(f)^{l-D13}$ / $FM7$, ct^S	A.Schalet Leiden	Lifschytz and Falk (1968)
$C(1)DX$, y f / y^2 v f $su(f)^{l-3DES}$; B^SY	A.Schalet Leiden	Schalet (1972)
$C(1)DX$, y f / $Df(1)GA130$; y^2Y67g 19.1	G.Miklos Canberra	isolated by G.Lefevre
$C(1)DX$, y f / $In(1)sc^{4L}sc^{8R}$, y sc^4 sc^8 w^a Tu $su(f)^{l-S1}$; B^SY	A.Schalet Leiden	Schalet (unpublished)
$su(f)^{l-S2}$ / $FM7$	A.Schalet Leiden	Schalet (1986)
$In(1)sc^{4L}sc^{8R}$, y sc^4 sc^8 w^a f^X $B(+)$ $su(f)^{l-X1}$ / $In(1)dl49$, y ac v f mal $su(f)^1$; B^SY	A.Schalet Leiden	Schalet (unpublished)
$C(1)DX$, y f / y ct^6 v f $su(f)^{l-X2}$; B^SY	A.Schalet Leiden	Lifschytz and Falk (1968)
$Dp(1:1)sc^{V1}$ $In(1)sc^{J1}$, $l(1)J1$ $l(1)J1^+$ y^S y^+ sc^{J1} sc^{V1} v f mal $su(f)^1$ bb / $In(1)sc^{S1L}sc^{8R}$ $In(1)dl49$, sc^8 sc^{S1} w^a v f $su(f)^{l-X3}$	A.Schalet Leiden	Schalet (unpublished)
$Binsn$ / $su(f)^{l-16.1.85}$; $v^+m^+Yy^+$	A.Schalet Leiden	Schalet and Lefevre (1973)
$Binsn$ / $su(f)^{l-16.3.162}$;$v^+m^+Yy^+$	A.Schalet Leiden	Schalet and Lefevre (1973)

2.19 GENETIC ASSAY FOR P FACTOR ACTIVITY

To determine the degree of P factor activity in a given interstrain cross, the gonadal dysgenesis (GD) sterility assay of Kidwell and Novy (1979) was used. Five virgin females were mated to five males and the vial placed at 29°C. Emerging F_1 females were aged for five days at 25°C before scoring for the presence or absence of fully formed ovaries. P factor activities were expressed as % GD sterility: the number of dysgenic ovaries / the total number of ovaries examined. For each cross, the ovaries of 25 F_1 females were scored. The reference M and P strains used were Canton S and π_2 respectively.

2.20 GERM-LINE TRANSFORMATION OF *DROSOPHILA*

Germ-line transformation of *Drosophila* was performed according to the principles outlined by Rubin and Spradling (1982). In one series of experiments, $rosy^{506}$ flies were injected with a recombinant clone using the *ry* transformation vector pDM23 (D.Mismer and G.M.Rubin, unpublished) derived from the vector Carnegie3ry2 (Rubin and Spradling, 1983). Recombinant plasmid was co-injected with pπ25.7wc helper plasmid (Karess and Rubin, 1984) following the protocol described by Spradling (1986). Embryos were dechorionated by immersion for 1-2 minutes in a 1:1 mixture of water and hypochlorite bleach (Waitrose plc).

In a second series of experiments, *bw;st* flies were used as the injection stock. The transformation vector used contained a selectible marker gene (*hph*) for resistance to the aminoglycoside antibiotic hygromycin B (Gritz and Davies, 1983). The *hph* gene encodes a hygromycin B phosphotransferase and is derived from an *E.coli* plasmid. The DNA fragment for transformation was cloned into the hygromycin resistance P element transformation vector hshyg-B (S.Parkhurst, unpublished): a derivative of Carnegie 4 (Rubin and Spradling, 1983) containing the bacterial *hph* gene under the control of the *Drosophila hsp70* promoter. As for the injection of *ry* embryos, the helper plasmid used was pπ25.7wc (Karess and Rubin, 1984).

Egg-collections (30-45 minutes) from the *bw;st* host strain were made using small grape juice agar plates, and the embryos transferred by paintbrush to a damp 1cmx2cm oblong of Whatman filter paper. The filter paper was then placed on a strip of Scotch double-sided sticky tape (3M) affixed to a microscope slide. A second slide bearing a strip of Scotch double-sided tape was placed upside down on the first, sandwiching the embryos. After applying gentle pressure, the embryos were dechorionated by pulling the two slides apart. The embryos were then transferred during a 7 minute interval to a strip of Sellotape double-sided sticky tape mounted on a coverslip. They were covered in Voltalef 15S oil (Produits Chimiques Ugine Kuhlmann) and injected with a mixture of recombinant plasmid and pπ25.7wc helper plasmid (Karess and Rubin, 1984) using needles pulled from microcapillary tubes with an internal ridge (GC100TF-10, Clark Electromedical Instruments).

Surviving larvae were transferred en masse to food bottles and those which eclosed (G0 adults) were aged for 3 days and then individually mated to several Canton-S or Oregon-R flies and placed on food containing hygromycin. Hygromycin B resistant G1 progeny were then back-crossed to *bw;st* flies, again with hygromycin selection. The G_2 hygromycin B resistant flies were used to set up putative transformed lines. DNA prepared from these lines was then analysed by Southern blotting; probing with a probe bearing *hph* DNA to determine whether transformation had occurred.

CHAPTER 3

THE IDENTIFICATION AND GENETIC CHARACTERISATION OF TWO PM HYBRID DYSGENESIS INDUCED ALLELES OF SUPPRESSOR-OF-FORKED

3.1 INTRODUCTION

Of the many strategies available for cloning genes in *Drosophila melanogaster*, one of the most fruitful has been that of "transposon tagging". The idea of using transposable element insertions to clone genes in *Drosophila* was first suggested by Bingham *et al.* (1981), who used a copia insertion linked to a *white* mutant phenotype to obtain cloned DNA from the *white* locus. The technique has general application, and can be applied to other types of transposable element insertions, including P elements. Since Searles *et al.* (1983) first used P element transposon tagging to clone a *Drosophila* RNA polymerase gene, an increasingly large number of genes have been cloned by this method.

In attempting to clone the *su(f)* locus, it was decided that P element transposon tagging would be the strategy of choice. Firstly, the technique is especially suited for loci for which a simple but effective genetic screen can be used to identify the P element insertion allele required. This is the case for *su(f)*, where a number of lethal alleles are available for use in a complementation screen. Secondly, and more prosaically, it might be difficult to clone *su(f)* by any other strategy. There exist no convenient chromosomal deletions or inversions which might be used to "jump" into or near *su(f)* using DNA cloned from a distant breakpoint, as employed in the cloning of the *Notch* locus (Artavanis-Tsakonas *et al.*, 1983). Microdissection and microcloning of chromosomal DNA (Scalenghe *et al.*, 1981) followed by chromosome walking would in the first instance be compromised by the poor polytene chromosome cytology of the proximal region of the X. Furthermore, any strategy calling for extended chromosome walking might prove to be difficult due to the presence of repetitive DNA, in view of the close proximity of *su(f)* to heterochromatin. Whether or not *su(f)* is closely linked to repetitive DNA is not known; though Miklos *et al.* (1988) found large amounts of repeat sequences in a series of clones from microdissections of the β-heterochromatic region of the X chromosome, where *su(f)* lies.

To be useful in cloning, an induced mutant allele must retain P element sequences at the gene of interest in order to provide a "tag" for cloning. Whilst most mutations induced by PM hybrid dygenesis are due to the insertion of P elements (Rubin *et al.* 1982; Engels, 1988), insertions of other transposable elements have occasionally been found (Rubin *et al.* 1982). P elements are also capable of inducing deletions by imprecise excision, or rearrangements following chromosome breakage at a P element site (Engels and Preston, 1981, 1984; for review see Engels, 1983, 1988). Therefore in some mutants induced by PM hybrid dysgenesis it is possible that there are no P element sequences are left at the original mutation site. As a result, any PM hybrid dysgenesis induced allele to be used in transposon tagging must first be examined to try and confirm that it is a P element insertion mutation. One property unique to P element insertion mutations is that they will yield revertants at a high rate when subjected to PM hybrid dysgenesis (Engels, 1983,1988), as a result of precise excision restoring the original structure of the previously mutant gene. This provides a test of the usefulness for transposon tagging of any PM hybrid dysgenesis induced mutation identified.

This chapter describes the screening of a collection of X-linked PM hybrid dysgenesis induced lethals and the isolation and subsequent genetic characterisation of two P element induced lethal alleles of *su(f)*.

3.2 SCREENING FOR PM HYBRID DYSGENESIS INDUCED *su(f)* ALLELES

A collection of P element hybrid dysgenesis induced X-linked lethals (M.J. Simmons, University of Minnesota) was screened for alleles of *su(f)*. The collection consists of lethals induced on an M strain X chromosome bearing f^1 using the P strain π_2 (M.J. Simmons, pers.comm.), and lethals induced on the X chromosome of π_2 itself (Simmons *et al.*, 1984). The lethals are maintained as balanced stocks either over *FM7* or *Df(1)Basc;y$^+$Y*. For each lethal stock, 3 to 4 *l(1)/Balancer* virgin females were mated to 2 to 3 *y cin w* f^s *su(f)*$^{l\text{-}ts67g}$ males. One brood of

each cross was grown at 18°C and another at 29°C; respectively the permissive and restrictive temperatures of the fully penetrant temperature sensitive lethal allele, *su(f)$^{l-ts67g}$* (Dudick *et al.*, 1974).

Of 110 lethal stocks examined, all gave *l(1)/y cin w f^{s} su(f)$^{l-ts67g}$* females at 18°C, but two failed to do so at 29°C. These lethal stocks, *l(1)MS97/Df(1)Basc;y^{+}Y* and *f^{1} l(1)MS252/FM7*, were then further examined by genetic means to determine whether or not they carried alleles of *su(f)*.

3.3 GENETIC CHARACTERISATION OF *l(1)MS97* AND *l(1)MS252*

3.3.1 RECOMBINATION MAPPING

The lethal mutations on the *l(1)MS97* and *f^{1} l(1)MS252* chromosomes were mapped by recombination using the multiply marked X chromosome *y cv v f car*. Both mutations were found to lie proximal to *car* (1:62.5). For *l(1)MS97*, 32 crossovers were found in the *car-l(1)MS97* interval out of a total of 869 chromosomes scored, placing *l(1)MS97* at 1:66.2. For *l(1)MS252*, 31 crossovers out of a total of 921 were found in the *car-l(1)MS252* interval, placing *l(1)MS252* at 1:65.9. These positions are in good agreement with the given map position of *su(f)* at 1:65.9 (Lindsley and Grell, 1968).

3.3.2 COMPLEMENTATION TESTING

At 25°C, both PM hybrid dysgenesis induced lethals failed to complement the following *su(f)* alleles with respect to *su(f)* lethality: *su(f)$^{l-3DES}$*, *su(f)$^{l-S1}$*, *su(f)$^{l-S2}$*, *su(f)$^{l-X2}$*, *su(f)$^{l-L26}$*, *su(f)$^{l-16-1-85}$* and *su(f)$^{R-9-18}$*. Furthermore, females of the genotype *f^{5} su(f)^{1}/f^{1} l(1)MS97* and *f^{5} su(f)^{1}/f^{1} l(1)MS252* showed almost complete suppression of forked bristles, although their bristles were thinner and reduced in length by comparison to *f^{+} su(f)$^{+}$* flies. This latter phenotype is similar to the "deficiency" or "Minute-like" phenotype described for certain trans-heterozygotes of *su(f)* alleles, notably *su(f)^{1}* over a *su(f)*

deficiency (Lindsley and Grell, 1968; Schalet, 1968). Interestingly, both lethals complemented $su(f)^{R-9-18}$ for lethality at 17°C but failed to do so at 25°C. Several other lethal alleles of *su(f)* are known to complement $su(f)^{R-9-18}$ for lethality (see Chapter 1, section 1.4.5).

Both lethals were rescued by the B^SY chromosome (Brosseau *et al.*, 1961), which carries X chromosome material covering the five most proximal known complementation groups, including *su(f)* (Schalet and Lefevre, 1973). This result implies that neither lethal is a large deletion extending beyond the distal breakpoint of B^SY. In addition, the two lethals fully complemented alleles of the three adjacent complementation groups distal to *su(f)*: $l(1)sph^{S1}$, *l(1)13E3* and *l(1)8P1*. Figure 1.2 shows a genetic map of the proximal X chromosome of *D.melanogaster*, including *su(f)*, its neighboring loci and the loci covered by B^SY. The two P element lethals are therefore lethal for *su(f)*, but not for any other locus tested.

The preceding genetic analyses indicate that *l(1)MS97* and *l(1)MS252* behave as lethal alleles of *su(f)* and also fail to complement whatever function is lacking in $su(f)^1$ flies which results in suppression of $forked^1$. At this point the two lethals were renamed $su(f)^{l-MS97}$ and $su(f)^{l-MS252}$.

3.4 REVERSION EXPERIMENTS

3.4.1 OUTCROSSING OF $su(f)^{l-MS97}$ AND $su(f)^{l-MS252}$

The following reversion experiments were designed to discover whether $su(f)^{l-MS97}$ and $su(f)^{l-MS252}$ would yield viable revertants following PM hybrid dysgenesis, and if so, to obtain revertant lines for further analysis. However, before the reversion experiments were carried out, the stocks carrying the two mutants alleles were subjected to a breeding scheme to remove all P elements not linked to the lethals. This was done for two reasons; firstly to obtain M cytotype stocks of each allele to facilitate the reversion experiment, and secondly to obtain stocks

with as few P elements as possible: stocks which could prove useful when trying to clone any P element associated with the *su(f)* alleles.

Crossing over between the f^1 $su(f)^{l\text{-}MS252}$ chromosome and an M strain *w m* f^1 chromosome allowed the isolation of two recombinant chromosomes: *w m* f^1 $su(f)^{l\text{-}MS252}$*-1* and *w m* f^1 $su(f)^{l\text{-}MS252}$*-2*, each maintained over an FM6 balancer from the M strain $In(1)N^{264\text{-}84}$*/FM6*. Autosomal P elements were removed by outcrossing twice against M strain H-41 males (*Basc/Y; SM1,Cy/Pm; TM2,Ubx/Sb;* spa^{pol}), selecting *Cy/+;Ubx/+* females in the F_1 and *Cy/Pm; Ubx/Sb* females in the F_2. The lines were then outcrossed twice against M strain FM6 males to remove all autosomal balancers and marked chromosomes, yielding the sublines *w m* f^1 $su(f)^{l\text{-}MS252}$*-1/FM6* and *w m* f^1 $su(f)^{l\text{-}MS252}$*-2/FM6*.

From the $su(f)^{l\text{-}MS97}$ stock, the same procedure was used to give the two sublines *w m* f^1 $su(f)^{l\text{-}MS97}$*-1/FM6* and *w m* f^1 $su(f)^{l\text{-}MS97}$*-2/FM6*. In the backcrossed sublines, the only P elements present should thus either be on the proximal portion of the X chromosome from *m* (1:36.1) or *f* (1:56.7) to the centromere; or on chromosome IV, which was not selected against. A DNA blotting experiment using P element sequences as hybridisation probe confirmed the absence of P element DNA sequences from the three strains described above as M strains (data not shown).

3.4.2 REVERSION OF $su(f)^{l\text{-}MS97}$ AND $su(f)^{l\text{-}MS252}$ FOLLOWING PM HYBRID DYSGENESIS

To generate P strain male flies with a first chromosome balancer for use in the reversion experiments, π_2 virgin females were crossed to M strain *FM6* males and F_1 virgin females crossed to π_2 males to give *FM6(* π_2*)* F_2 males. The mating scheme used for the reversion experiments is shown as follows:

$$\frac{w\ m\ f^1\ su(f)^{l\text{-}MS}}{FM6} \quad x \quad \frac{FM6(\pi_2)}{Y} \quad : \quad \text{dysgenic cross}$$

$$\text{or:} \quad x \quad \frac{FM6(M)}{Y} \quad : \quad \text{control}$$

$$\Big|$$

$$\Big|$$

$$\Big|$$

F_1 $\quad \frac{w\ m\ f^1\ su(f)^{l\text{-}MS}}{FM6} \quad x \quad \frac{FM6}{Y}$ brothers

$$\Big|$$

$$\Big|$$

$$\Big|$$

F_2 $\quad$ count $\quad \frac{w\ m\ f^1\ su(f)^{l\text{-}MS}}{FM6}$ "balancer" females

$\quad$ score $\quad \frac{w\ m\ f^1}{Y}$ non-balancer males

The F_1 cross was carried out in six different bottles. The reversion rates obtained for the two $su(f)^{l\text{-}MS252}$ sublines were as follows, along with the values for their respective control experiments:

$su(f)^{l\text{-}MS252\_1}$: dysgenic cross: 3.2×10^{-3}
7 (4) *w m* f^1 males, 2159 "balancer" females

: control cross : 9.7×10^{-4}
2 (2) *w m* f^1 males, 2062 "balancer" females

$su(f)^{l\text{-}MS252\_2}$: dysgenic cross: 3.25×10^{-3}
6 (2) *w m* f^1 males, 1846 "balancer" females

: control cross : 4.5×10^{-4}
1 *w m* f^1 male, 2230 "balancer" females

The numbers in brackets refer to the minimum number of independent reversion events in each case. The non-random distribution of revertant males in the F_2 suggests that the reversion events are occurring pre-meiotically in bursts, a characteristic of reversion of P element insertion alleles (Engels, 1983). A number of flies were set aside from each dysgenic parental mating to determine the extent of P factor activity in each case using the gonadal dysgenesis (GD) sterility A-cross assay (Kidwell and Novy, 1979). Also tested were the P factor activity strengths of the

FM6(π_2) males, with π_2 males as controls. The results were as follows:

$\frac{w\ m\ f^1\ su(f)^{l\text{-}MS252}\_1}{FM6}$	x	$\frac{FM6\ (\pi_2)}{Y}$	:	87% GD sterility
$\frac{w\ m\ f^1\ su(f)^{l\text{-}MS252}\_2}{FM6}$	x	$\frac{FM6\ (\pi_2)}{Y}$	:	98% GD sterility
Canton S females	x	$\frac{FM6\ (\pi_2)}{Y}$	:	70% GD sterility
Canton S females	x	π_2 males	:	100% GD sterility

The results indicate that the reversion crosses were indeed dysgenic; however, the *FM6(π_2)* males, whilst yielding a high degree of GD sterility when crossed to Canton S females, are not as strong a P strain as π_2. The revertants obtained in each control cross indicates that these were also somewhat dysgenic, perhaps because of the presence of one or more intact P elements remaining in the two backcrossed $su(f)^{l\text{-}MS252}$ sublines. Subsequent cloning experiments indicated that this was indeed the case (see Chapter 4).

At a later date the *w m* f^1 $su(f)^{l\text{-}MS97}\_1$ subline was also tested in a reversion experiment of the same design, but using in the dysgenic cross *FM7(π_2)* males from a strong P strain (Engels, 1985). The reversion rate obtained is given below, along with the results of GD sterility assays performed concurrently:

$su(f)^{l\text{-}MS97}\_1$: dysgenic cross: 6.9×10^{-3}
15 (7) *w m* f^1 males, 2176 "balancer" females

: control cross : 0
no *w m* f^1 males in 1749 "balancer" females

$\frac{w\ m\ f^1\ su(f)^{l\text{-}MS97}\_1}{FM6}$	x	$\frac{FM7\ (\pi_2)}{y^+Y}$	:	100% GD sterility
$\frac{w\ m\ f^1\ su(f)^{l\text{-}MS97}\_1}{FM6}$	x	$\frac{FM6\ (M)}{Y}$	:	4% GD sterility
Canton S females	x	$\frac{FM7\ (\pi_2)}{y^+Y}$	:	100% GD sterility
Canton S females	x	π_2	:	100% GD sterility

In this experiment, a high reversion rate is found only in the dysgenic cross: the control cross gave a low figure for GD sterility, and no revertants. A number of revertant males obtained from each of the experiments described were used to set up revertant lines propagated using *C(1)DX, y f* females.

3.5 INTRA-ALLELIC COMPLEMENTATION AT *su(f)* AND THE *zeste* GENE

The lethal allele $su(f)^{l\text{-}3DES}$ and the extremely sub-viable allele $su(f)^{R\text{-}9\text{-}18}$ give perfectly viable female flies when heterozygous with each other (Chapter 1, section 1.4.5). To discover whether intra-allelic complementation at the *su(f)* locus can be affected by the allelic state of the *zeste* locus, trans-heterozygous female flies carrying these two alleles were examined in both a $zeste^{1}$ and a $zeste^{a}$ genetic background.

The expression of a range of mutant alleles of several loci are affected in a synapsis-dependent fashion in trans by mutations of the *zeste* locus (*z*; 1:1.0). This curious set of interactions has been named transvection, and the loci where transvection effects have been demonstrated include *white*, *Ultrabithorax* and *decapentaplegic* (Gans, 1953; Babu and Bhat, 1980; Kaufman *et al.*, 1973; Gelbart and Wu, 1982). The suspicion remains that transvection effects may involve a much wider range of genes, and that these might be identified if only the special genetic circumstances existed in each case to demonstrate *zeste* interactions. One such "special circumstance" might be that of intra-allelic complementation. In the case of *white*, intra-allelic complementation occurs between the $w^{spotted1}$ allele, which is caused by the insertion of a B104 element into the regulatory region of *w* (O'Hare *et al.*, 1984), and mutations affecting the structural region of *w* (Lewis, 1956; Green, 1959b). The w^{sp1} intra-allelic partial complementation is abolished in z^{a} flies (Babu and Bhat, 1980). The z^{a} class of alleles have been characterised as hypomorphic or nullomorphic alleles, whereas z^{1} appears to be a neomorphic allele (Jack and Judd, 1979). Therefore in the following experiments, the effects

of both a z^1 and a z^a genetic background have been examined.

The *zeste* alleles from the strains *sc* z^1 w^{op7red} *ec* and *y* z^a w^{ch} were placed by recombination onto the y^2 *v* f^1 *su(f)*$^{l\text{-}3DES}$ and the *y* ct^6 *v* f^1 *car su(f)*$^{R\text{-}9\text{-}18}$ chromosomes to give the following four recombinant chromosomes, each balanced over *FM6*:

sc z^1 w^{op7red} *ec* f^1 *su(f)*$^{l\text{-}3DES}$
sc z^1 w^{op7red} *ec* f^1 *su(f)*$^{R\text{-}9\text{-}18}$
y z^a w^{ch} f^1 *su(f)*$^{l\text{-}3DES}$
y z^a w^{ch} f^1 *su(f)*$^{R\text{-}9\text{-}18}$

The lethal or viable phenotypes of test females resulting from a series of crosses using these four recombinant stocks and their parental chromosomes were as follows. Genetic markers other than *z* and *su(f)* alleles have been omitted for clarity. All crosses were performed at 25°C using 4-5 virgin females and 3-4 males; and at least 100 progeny examined in each case.

Cross

A.	z^1 *su(f)*$^{l\text{-}3DES}$ / z^1 *su(f)*$^{l\text{-}3DES}$	Lethal
B.	z^a *su(f)*$^{l\text{-}3DES}$ / z^a *su(f)*$^{l\text{-}3DES}$	Lethal
C.	z^1 *su(f)*$^{R\text{-}9\text{-}18}$ / z^1 *su(f)*$^{R\text{-}9\text{-}18}$	Lethal
D.	z^a *su(f)*$^{R\text{-}9\text{-}18}$ / z^a *su(f)*$^{R\text{-}9\text{-}18}$	Lethal
E.	z^1 *su(f)*$^{l\text{-}3DES}$ / z^1	Viable
F.	z^a *su(f)*$^{l\text{-}3DES}$ / z^a	Viable
G.	z^1 *su(f)*$^{R\text{-}9\text{-}18}$ / z^1	Viable
H.	z^a *su(f)*$^{R\text{-}9\text{-}18}$ / z^a	Viable
I.	*su(f)*$^{l\text{-}3DES}$ / *su(f)*$^{R\text{-}9\text{-}18}$	Viable

J.	z^1 $su(f)^{l\text{-}3DES}$ / $su(f)^{R\text{-}9\text{-}18}$	Viable
K.	z^a $su(f)^{l\text{-}3DES}$ / $su(f)^{R\text{-}9\text{-}18}$	Viable
L.	z^1 $su(f)^{R\text{-}9\text{-}18}$ / $su(f)^{l\text{-}3DES}$	Viable
M.	z^a $su(f)^{R\text{-}9\text{-}18}$ / $su(f)^{l\text{-}3DES}$	Viable
N.	z^1 $su(f)^{l\text{-}3DES}$ / z^1 $su(f)^{R\text{-}9\text{-}18}$	Viable
O.	z^a $su(f)^{l\text{-}3DES}$ / z^a $su(f)^{R\text{-}9\text{-}18}$	Viable

The results show that $su(f)^{l\text{-}3DES}/su(f)^{R\text{-}9\text{-}18}$ females remain viable in both a z^1 and a z^a background. All the females in the above list are homozygous for f^1, and the the bristles of cross N and O females were identical to those of cross I females; i.e. f^1 is suppressed. Thus the allelic state of the *zeste* locus affects neither the viability of trans-heterozygotes of these two *su(f)* alleles nor their failure to complement each other in suppression of *forked* bristles.

3.6 CONCLUDING REMARKS

The screen of X-linked PM hybrid dysgenesis induced lethal mutations identified two mutations which on further analysis have been shown to be lethal alleles of *su(f)*: $su(f)^{l\text{-}MS97}$ and $su(f)^{l\text{-}MS252}$. Complementation testing with other *su(f)* alleles and adjacent lethals showed that both appear to be deficient only for *su(f)*, and are unlikely to be large deletions. However, much the best evidence regarding the nature of the causative lesions involved in each case comes from the reversion experiments. These show that if subjected to further PM hybrid dysgenesis, both $su(f)^{l\text{-}MS97}$ and $su(f)^{l\text{-}MS252}$ give fully viable revertants at a high rate. This strongly suggests that both alleles are indeed caused by the insertion of P elements, and therefore that they can provide the means by which cloned sequences from the *su(f)* locus can be obtained by transposon tagging. The sublines of each allele generated by recombination and outcrossing against a multiply balanced M strain should now prove useful in attempting to isolate a clone bearing the

putative P element insertions responsible for $su(f)^{l\text{-}MS97}$ and $su(f)^{l\text{-}MS252}$.

The possibility remains that both mutations are in fact identical, and have been graced with different stock numbers through mislabelling. Indeed, $su(f)^{l\text{-}MS97}$ and $su(f)^{l\text{-}MS252}$ consistently show exactly or essentially the same results in each of the analyses described. However, the one difference between them implies that this is not the case: the fact that there is a f^1 allele on the original chromosome carrying $su(f)^{l\text{-}MS252}$ but not on that carrying $su(f)^{l\text{-}MS97}$. This suggests that $su(f)^{l\text{-}MS97}$ has a π_2 derived X chromosome, but not $su(f)^{l\text{-}MS252}$.

Lastly, a series of genetic experiments has demonstrated that neither a z^1 nor a z^a background affects the intra-allelic complementation shown by $su(f)^{l\text{-}3DES}$ and $su(f)^{R\text{-}9\text{-}18}$; i.e. the results showed no detectable interactions between *zeste* and *su(f)*.

CHAPTER 4

MOLECULAR CLONING AND ANALYSIS OF DNA FROM THE SUPPRESSOR-OF-FORKED LOCUS

4.1 INTRODUCTION

The previous chapter described the identification and genetic characterisation of two new alleles of *su(f)*. The genetic evidence suggests that they may be P element insertion mutations. This chapter describes the subsequent use of one of these alleles to obtain cloned DNA sequences from the *su(f)* locus, and the use of the clones obtained in the structural analysis of both wild type and mutant alleles of the locus.

A commonly used procedure for establishing whether P element sequences reside at the correct chromosomal location in a putative P element insertion allele is in situ hybridisation to polytene chromosomes with a P element probe. However, this experiment was not attempted in view of the extremely poor cytology of polytene chromosomes in section 20 (Schalet and Lefevre, 1976). Instead, the decision was made to clone all the P elements in one of the PM hybrid dysgenesis induced *su(f)* alleles and determine which, if any, is responsible for the mutant phenotype.

4.2 CLONING OF DNA SEQUENCES FROM *su(f)*

4.2.1 CONSTRUCTION OF GENOMIC LIBRARIES FROM $su(f)^{l\text{-}MS252}\text{-}1$

Figure 4.1 shows the results of a DNA blotting experiment designed to discover the number of P elements present in the two $su(f)^{l\text{-}MS252}$ sublines. DNA from these two lines was digested with BamHI, a restriction enzyme whose recognition site is not present within P elements (O'Hare and Rubin, 1983). The filter was hybridised to pπ25.1, a clone bearing a complete P element together with flanking DNA from 17C (O'Hare and Rubin, 1983). The results reveal the presence of at least three P elements in the *w m* f^1 $su(f)^{l\text{-}MS252}\text{-}1/FM6$ subline and at least seven in the *w m* f^1 $su(f)^{l\text{-}MS252}\text{-}2/FM6$ subline. Figure 4.1 also shows the same DNA samples digested with XhoI and SalI. Intact P elements give a characteristic 1.7kb XhoI-SalI fragment, as shown by the strong band of this size in π_2 DNA. The results show that both $su(f)^{l\text{-}MS252}$ sublines appear to have at least one intact P element.

Figure 4.1 The number of P elements in the two *su(f)*$^{l\text{-}MS252}$ sublines

Shown is a DNA blot of samples from the two *su(f)*$^{l\text{-}MS252}$ sublines probed with pπ25.1. Lanes 1 and 2 : DNA digested with BamHI. Lanes 3, 4 and 5 : DNA digested with SalI and EcoRI.

Lanes 1 and 3 : *w m f*^{1} *su(f)*$^{l\text{-}MS252}$*-1 / FM6*

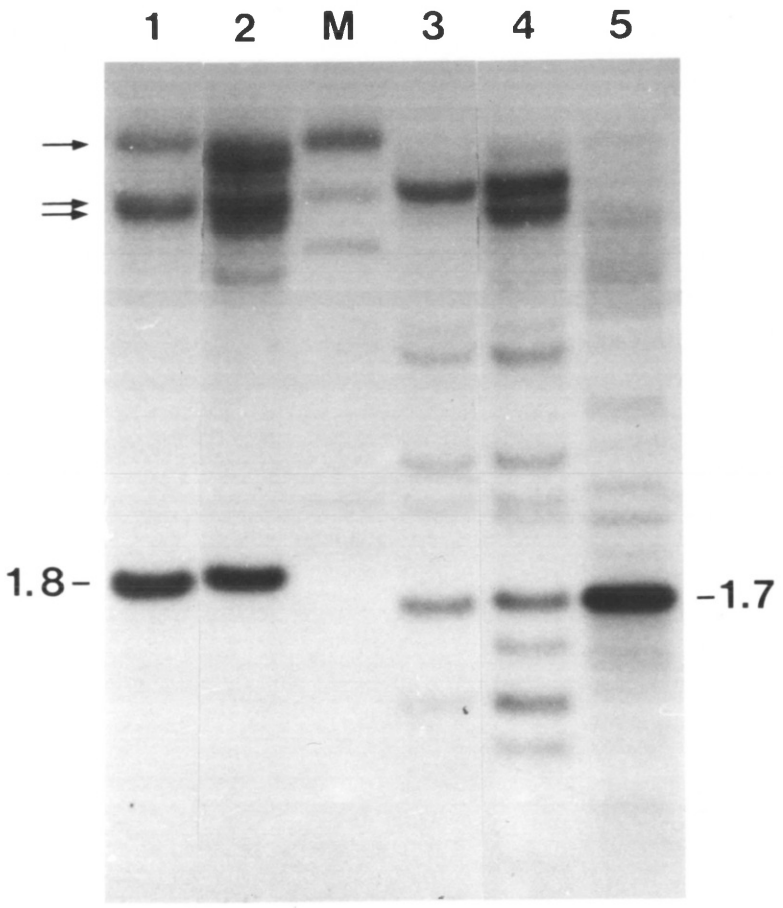
Lanes 2 and 4 : *w m f*^{1} *su(f)*$^{l\text{-}MS252}$*-2 / FM6*

Lane 5 : π_2

Lane M : λ digested with HindIII

The 1.8kb band marked in lanes 1 and 2 is the BamHI fragment which hybridises to DNA from 17C flanking the P element in the probe. The prominent 1.8kb band marked in lane 5 is the characteristic SalI-XhoI fragment from intact P elements; this band is also present in the samples of DNA from the two *su(f)*$^{l\text{-}MS252}$ sublines.

FIGURE 4.1



Using DNA from the *su(f)$^{l-MS252}$-1* subline, two genomic libraries were constructed in λ vectors. For the first, 2ug of *w m f^{l} su(f)$^{l-MS252}$-1/FM6* DNA was digested to completion with BamHI and ligated to 1ug of λCharon 28 vector arms prepared by digestion with BamHI followed by purification from an agarose gel. For the second, 6ug of the same DNA was partially digested with Sau3aI and fragments of 15-20kb purified from a low-gelling-temperature agarose gel. After recovery from the gel, these fragments were ligated to 2ug of λ EMBL4 DNA digested with both BamHI and SalI. Following ligation, the libraries were packaged in vitro and plated out using *E.coli* strains C600 (λCharon 28 library) or Q359 (λ EMBL4 library).

Filter lifts from each library were hybridised to nick-translated pπ25.7BWC (K.O'Hare, unpublished), which carries a complete P element less 38bp at its left hand end and 21bp at its right hand end (O'Hare and Rubin, 1983). One positively hybridising phage was obtained from the λCharon 28 library (λ252.B1) and sixteen from the λEMBL4 library (λ252.S1-S16). Except for λ252.S15, which failed to rescreen, the phage were purified and DNA prepared from large scale liquid lysate cultures of eight of the sixteen remaining. The inserts of these eight were characterised by mapping restriction enzyme sites and DNA blotting experiments using P element probes. Restriction fragments flanking the P elements in a number of the inserts were subcloned into plasmid vectors and used as hybridisation probes against all 16 P element bearing λ clones (data not shown). The results of these analyses are summarised in Figure 4.2, which shows the restriction enzyme site maps obtained and the plasmid subclones used. Fourteen of the sixteen λ clones were arranged in four groups (the B1, S5, S7 and S10 groups), each representing a single P element insertion site in the *w m f^{l} su(f)$^{l-MS252}$-1/FM6* genome:

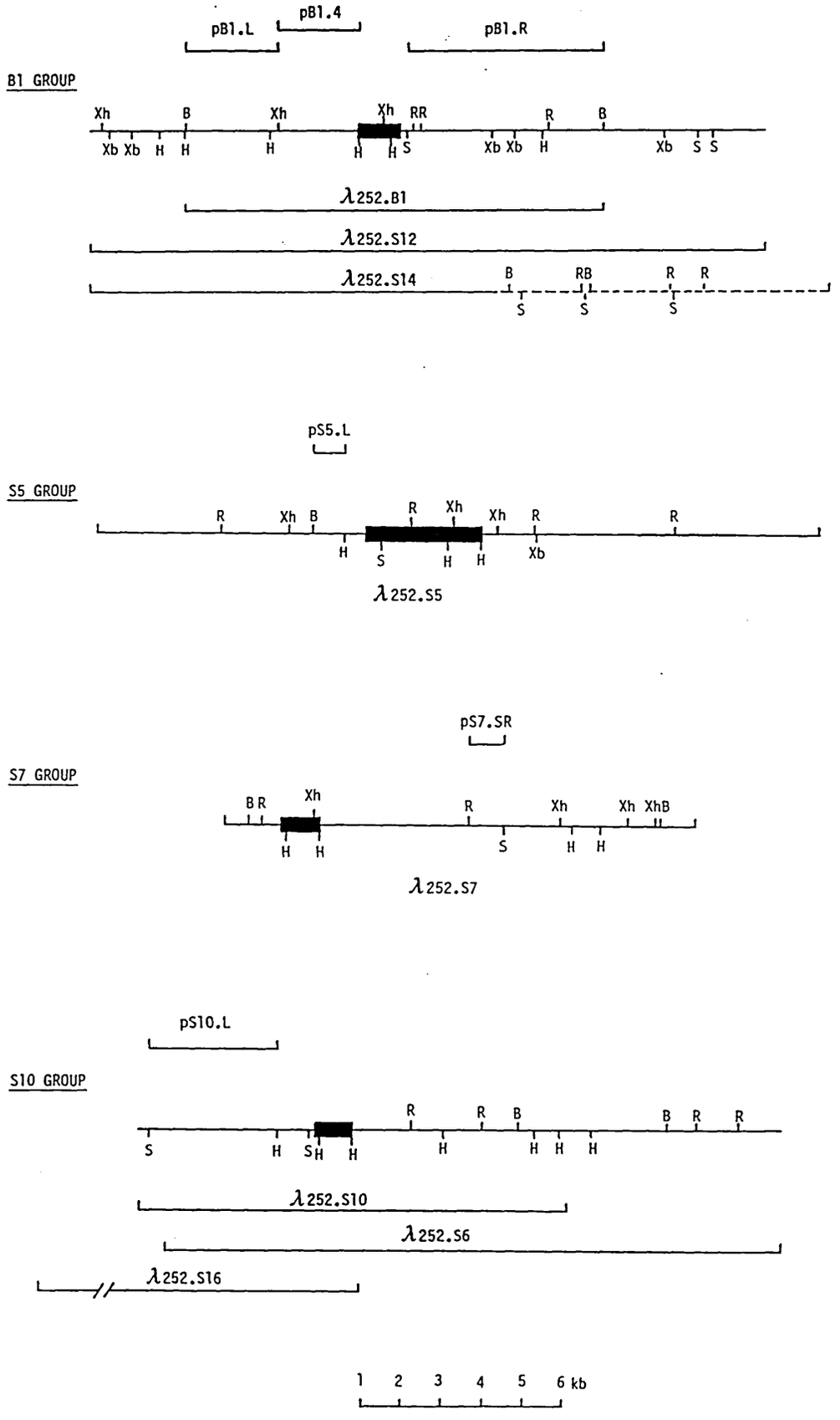
B1 group	S5 group	S7 group	S10 group
252.B1	252.S3	252.S1	252.S2
252.S12	252.S5	252.S7	252.S6
252.S14	252.S8	252.S9	252.S10
	252.S13		252.S16

Figure 4.2 Restriction enzyme site maps of the inserts of P element-bearing clones from *w m* f^1 *su(f)*$^{1-MS252}$*_1*

Shown are the restriction enzyme site maps of the λ clones corresponding to the four insertion site groups cloned from *w m* f^1 *su(f)*$^{1-MS252}$*_1 / FM6* :B1, S5, S7 and S10. For the B1 and S10 groups, three λ clones were mapped; the limits of each are indicated below their respective restriction site maps. Restriction sites for BamHI, SalI, HindIII and EcoRI were mapped in all the inserts; in addition, the inserts of clones from groups B1 and S5 were also mapped with XhoI and XbaI. Shown above the maps are the intervals from each group which were subcloned into plasmid vectors. The thick black lines represent P element sequences. The dotted line and associated restriction site map in λ252.S14 represents sequences which do not match λ252.B1 or λ252.S12. Presumably λ252.S14 is a chimaeric recombinant of two Sau3aI fragments.

Restriction site key:	B : BamHI	S : SalI
	H : HindIII	Xh : XhoI
	R : EcoRI	Xb : XbaI

FIGURE 4.2


pB1.L
pB1.4
pB1.R
B1 GROUP
Xh B Xh Xh RR R B
Xb Xb H H H H H S Xb Xb H Xb S S
λ252.B1
λ252.S12
λ252.S14
B RB R R
S S S
pS5.L
S5 GROUP
R Xh B R Xh Xh R R
H S H H Xb
λ252.S5
pS7.SR
S7 GROUP
B R Xh R Xh Xh XhB
H H S H H
λ252.S7
pS10.L
S10 GROUP
R R B B R R
S H S H H H H H H
λ252.S10
λ252.S6
λ252.S16
1 2 3 4 5 6 kb

Two of the sixteen clones could not be assigned to any of the above four groups, namely λ252.S4 and λ252.S11. These may represent new insertion sites following transposition during the propagation of the *su(f)*$^{l\text{-}MS252}$*-1* subline, as this stock has at least apparently intact P element (Figure 4.1). Indeed, the P element in the S5 group has the restriction site map of an intact 2.9kb element; but the other three are incomplete elements, on the basis of their restriction site maps and the sizes of the relevant restriction fragments.

4.2.2 IDENTIFICATION OF THE P ELEMENT INSERTED IN *su(f)*$^{l\text{-}MS252}$

The flanking genomic DNA subclones were used as probes in DNA blotting experiments on DNA from mutant and revertant flies in order to discover if any of the cloned P elements were associated with the mutation *su(f)*$^{l\text{-}MS252}$. Figure 4.3a shows that the P element of the B1 group behaves exactly in this fashion. The subclone pB1.L was used to probe a DNA blot filter bearing a number of DNA samples digested with both BamHI and SalI. In those stocks which carry the mutation *su(f)*$^{l\text{-}MS252}$, there is an additional hybridising band migrating 1.1kb larger than the wild-type band, corresponding to the chromosome carrying the 1.1kb B1 group P element insertion. However, the larger band is absent in all six independent revertant lines examined. When a flanking DNA subclone from the S7 group insertion site (pS7.SR) was later used as a hybridisation probe on the same filter, all the revertant lines retain the corresponding larger band resulting from the the insertion of the S7 group P element (Figure 4.3b).

The presence of the lethal allele *su(f)*$^{l\text{-}MS252}$ in a given stock correlates with the presence of the P element of the B1 group; this P element is absent in each of the *su(f)*$^{l\text{-}MS252}$ revertant lines. A similar correlation does not hold for the S7 group P element. These results provide strong evidence that the B1 group P element is responsible for the lethal phenotype of *su(f)*$^{l\text{-}MS252}$, demonstrating the usefulness of revertants when attempting to link a mutant phenotype to an observed DNA lesion.

Figure 4.3 DNA blot analysis of *su(f)*$^{l\text{-}MS252}$ lines and their revertants

Panel A shows DNA samples digested with BamHI and SalI and probed with pB1.L: a flanking probe from the B1 P element insertion site. Panel B shows the same filter probed with pS7.SR: a flanking probe from the S7 P element insertion site. Below each panel is a diagram showing the relationship of the probes to their respective P element insertion sites. The sizes of the bands are marked in kb on the right of each panel.

Lane 1 : *FM6*

Lane 2 : Canton S

Lane 3 : *f*^{1} *su(f)*$^{l\text{-}MS252}$ */ FM7*

Lane 4 : *w m f*^{1} *su(f)*$^{l\text{-}MS252}$*-1 / FM6*

Lane 5 : *w m f*^{1} *su(f)*$^{l\text{-}MS252}$*-1.RP1*

Lane 6 : *w m f*^{1} *su(f)*$^{l\text{-}MS252}$*-1.RP2b*

Lane 7 : *w m f*^{1} *su(f)*$^{l\text{-}MS252}$*-1.RP3*

Lane 8 : *w m f*^{1} *su(f)*$^{l\text{-}MS252}$*-2 / FM6*

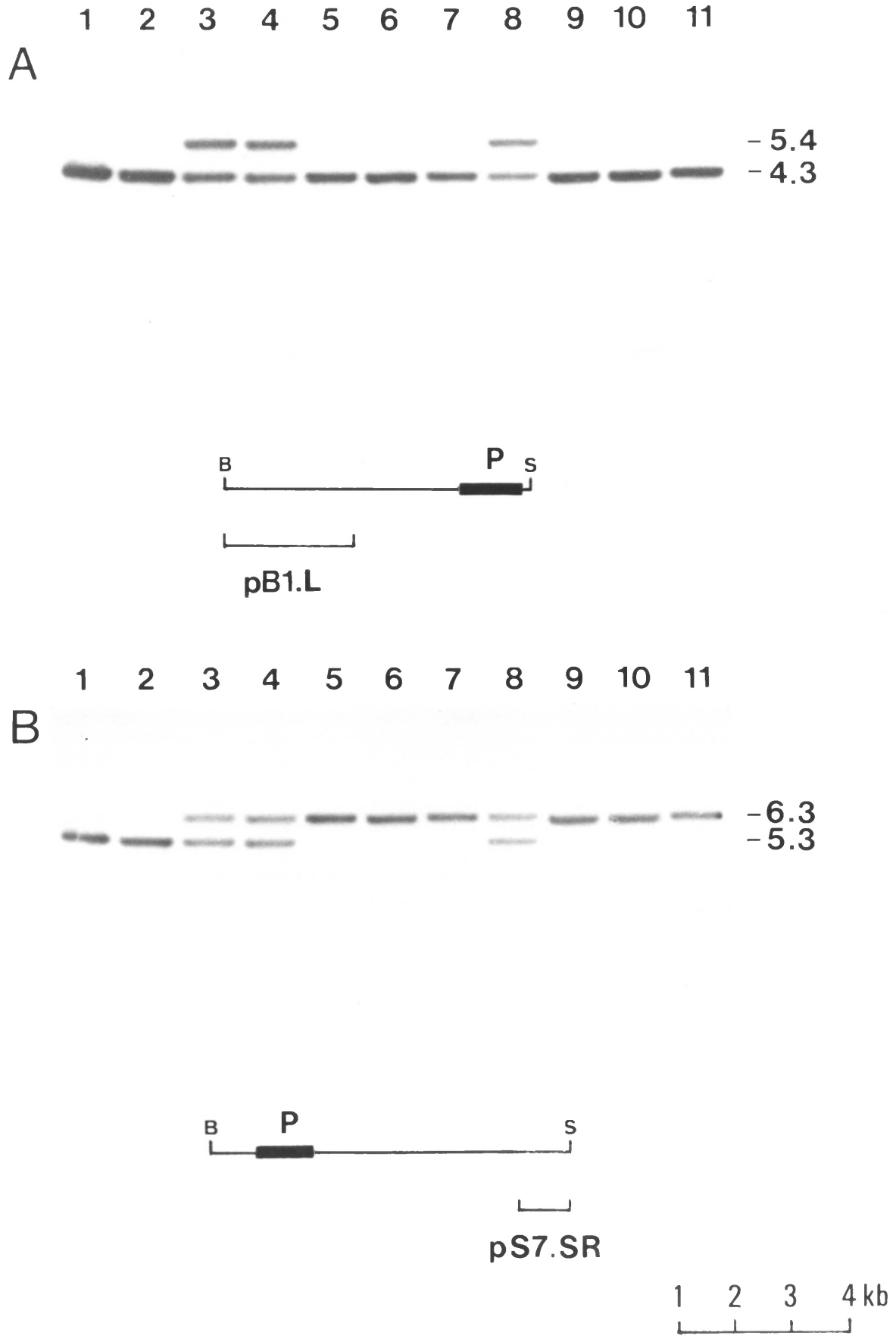
Lane 9 : *w m f*^{1} *su(f)*$^{l\text{-}MS252}$*-2.R3a*

Lane 10 : *w m f*^{1} *su(f)*$^{l\text{-}MS252}$*-2.R6a*

Lane 11 : *w m f*^{1} *su(f)*$^{l\text{-}MS252}$*-2.RM4*

Each of the six revertants arose independently; *RP1*, *RP2b* and *RP3* from *w m f*^{1} *su(f)*$^{l\text{-}MS252}$*-1 / FM6*; and *R3a*, *R6a* and *RM4* from *w m f*^{1} *su(f)*$^{l\text{-}MS252}$*-2 / FM6*. The DNA samples in lanes 3, 4 and 8 were prepared from heterozygous females, and show two bands, corresponding to the two different X chromosomes in each stock. Each of the revertant DNA samples were prepared from hemizygous males, and therefore show only one chromosome-characteristic band.

FIGURE 4.3


1 2 3 4 5 6 7 8 9 10 11
A
-5.4
-4.3
B
P
S
pB1.L
1 2 3 4 5 6 7 8 9 10 11
B
-6.3
-5.3
B
P
S
pS7.SR
1 2 3 4 kb

If $su(f)^{l-MS97}$ is also a P element insertion allele, then we might reasonably expect to see evidence of an insertion in $su(f)^{l-MS97}$ in the region of DNA covered by the B1 group. DNA blotting of samples from the original $su(f)^{l-MS252}$ and $su(f)^{l-MS97}$ stocks and their respective sublines reveals that this is indeed the case. Figure 4.4 shows that the $su(f)^{l-MS97}$ stock and its derived sublines give the same hybridisation pattern as the $su(f)^{l-MS252}$ stocks. This suggests that $su(f)^{l-MS97}$ is also associated with a similarly sized insertion of DNA in the same 4.3kb interval. Whether or not the DNA lesion observed in $su(f)^{l-MS97}$ is due to the insertion of a P element cannot be established by DNA blotting experiments alone. However, the genetic provenance of $su(f)^{l-MS97}$ suggests that this is probably the case; a supposition confirmed by cloning experiments described later in this chapter.

4.2.3 CLONING OF WILD-TYPE *su(f)* DNA SEQUENCES

The flanking DNA fragment subclones pB1.L and pB1.4 (Figure 4.2) were used as hybridisation probes to screen a Canton S DNA Sau3aI partial-digest library cloned in λEMBL4 (Mariani *et al.*, 1985). Five positively-hybridising phage clones were isolated and purified; λCSP.1 to λCSP.5. DNA was prepared from large-scale liquid lysate cultures of each and the restriction enzyme recognition sites within their inserts were mapped using six different restriction enzymes.

The five overlapping wild-type phage clones cover a 32kb region of the Canton S genome. Figure 4.5 shows the restriction enzyme site map of the 32kb interval, along with the limits of each wild type λ clone. Also shown are a number of restriction fragments which were subcloned into the plasmid vectors pBR327 or pBluescribe M13-. The SalI site ~150bp to the right of the site of the P element insertion in $su(f)^{l-MS252}$ was designated 0.0 in the coordinate system, with the + direction to the right and the - direction to the left. Restriction enzyme site maps of loci on the X chromosome are conventionally drawn such that the telomere is

Figure 4.4 *su(f)$^{l-MS97}$* has an insertion in the same region as *su(f)$^{l-MS252}$*

Shown is a DNA blot of samples from *su(f)$^{l-MS252}$* and *su(f)$^{l-MS97}$* and their respective sublines, digested with BamHI and SalI and probed with pB1.L. The sizes of the bands are marked in kb on the right of the panel.

Lane 1 : Canton S

Lane 2 : *FM7*

Lane 3 : *FM6*

Lane 4 : *f^{1} su(f)$^{l-MS252}$ / FM7*

Lane 5 : *w m f^{1} su(f)$^{l-MS252}$-1 / FM6*

Lane 6 : *w m f^{1} su(f)$^{l-MS252}$-2 / FM6*

Lane 7 : *su(f)$^{l-MS97}$ / Df(1)Basc*

Lane 8 : *w m f^{1} su(f)$^{l-MS97}$-1 / FM6*

Lane 9 : *w m f^{1} su(f)$^{l-MS97}$-2 / FM6*

Each of the two *su(f)$^{l-MS252}$* sublines retains the band migrating 1.1kb larger than the wild type band, corresponding to the 1.1kb P element insertion of the B1 insertion site. Both *su(f)$^{l-MS97}$* and its two sublines have a similar sized band, suggesting that this second PM hybrid dysgenesis induced allele may also have a P element insertion in the same 4.3kb BamHI/SalI interval.

FIGURE 4.4

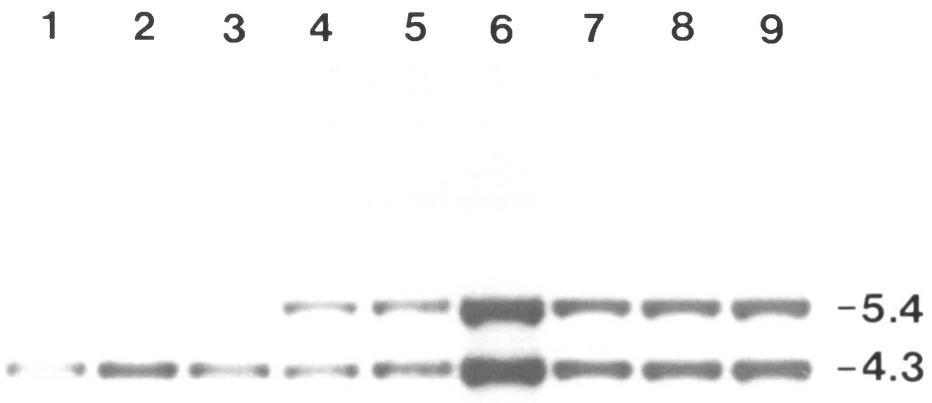


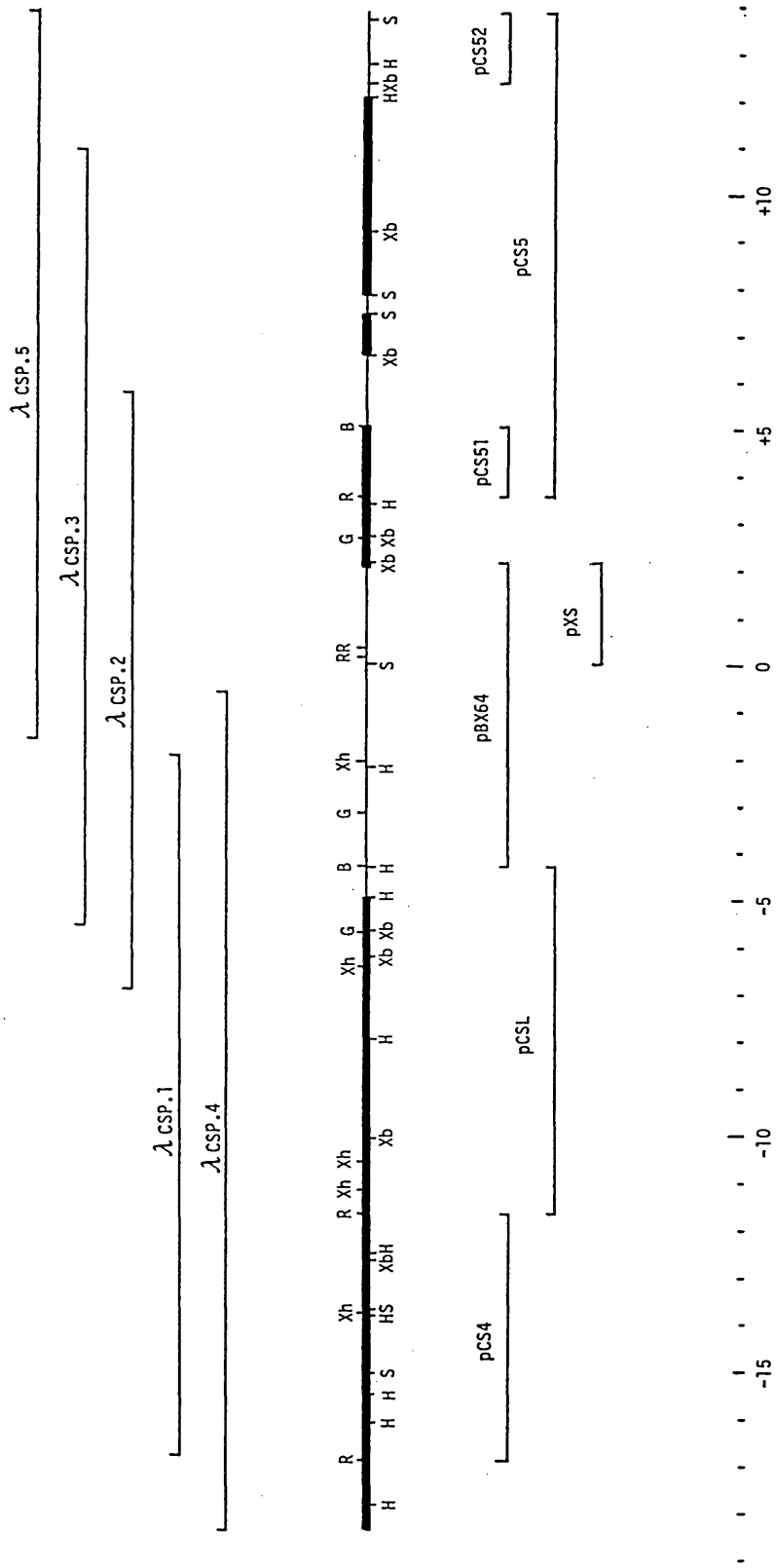
Figure 4.5 Restriction enzyme site map of the *su(f)* region from Canton S

Shown is the restriction enzyme site map of the *su(f)* region cloned from Canton S. Unique DNA sequences are represented by thin lines; repetitive DNA sequences are represented by the thick lines (see Figure 4.10). Above the map are shown the five overlapping λ clones which cover a total of 32kb. At the bottom of the Figure is shown the coordinate system marked in kb, where coordinate 0.0 is the SalI site near the centre of the interval. Restriction enzyme sites for BamHI, EcoRI, SalI, HindIII, XhoI and XbaI were mapped over the whole interval; BglII sites were only mapped in the interval between coordinates -6.4 and +2.7. Below the map are shown fragments from the wild type λ clones which were subcloned into plasmid vectors.

Restriction site key:

B : BamHI	S : SalI
H : HindIII	Xh : XhoI
R : EcoRI	Xb : XbaI
G : BglII	

FIGURE 4.5


λ CSP.5
λ CSP.3
λ CSP.2
λ CSP.1
λ CSP.4
pCS4
pCSL
pBX64
pXS
pCS51
pCS5
pCS52
-15
-10
-5
0
+5
+10

in the left hand direction, and the centromere in the right hand direction. However, none of the data contained in this thesis reveals the correct orientation of the 32kb cloned interval, so the orientation as drawn should not be taken as suggesting that the centromere is to the right.

4.3 REPETITIVE DNA IN THE *su(f)* REGION

Plasmid subclones covering almost all of the 32kb interval were digested with varying combinations of a number of restriction enzymes to give a DNA blot filter carrying restriction fragments of the cloned interval. The filter was first hybridised to the M13 clone 2(8), which carries multiple copies of the Bkm satellite GATA repeat (Singh *et al.*, 1984). Using GATA repeat-containing probes in in situ hybridisation to polytene chromosomes, GATA repeats have been found in the base of the X chromosome in the regions 19F and 20AB (Singh *et al.*, 1981; Singh *et al.*, 1984). Although this region is close to the cytological location of *su(f)*, the 2(8) probe failed to hybridise to DNA fragments from the cloned region around *su(f)* (data not shown).

In view of the chromosomal location of *su(f)*, the cloned wild-type interval was next examined for the presence of sequences repeated within the interval. After removal of the previous probe, the filter was probed successively with five plasmid subclones covering almost the whole interval (pCS4, pCSL, pBX64, pB1.R and pCS5). For each hybridisation, the filter was first stripped of the previous probe.

4.3.1 THE "1.5kb" REPEATS

When pCS4, pBX64 and pCS5 were used as probes, the only positively hybridising fragments were those contained within the clones themselves (data not shown). However, both pCSL and pB1.R hybridised to fragments other than themselves (Figures 4.6 and 4.7), demonstrating the presence of sequence reiteration within the cloned interval.

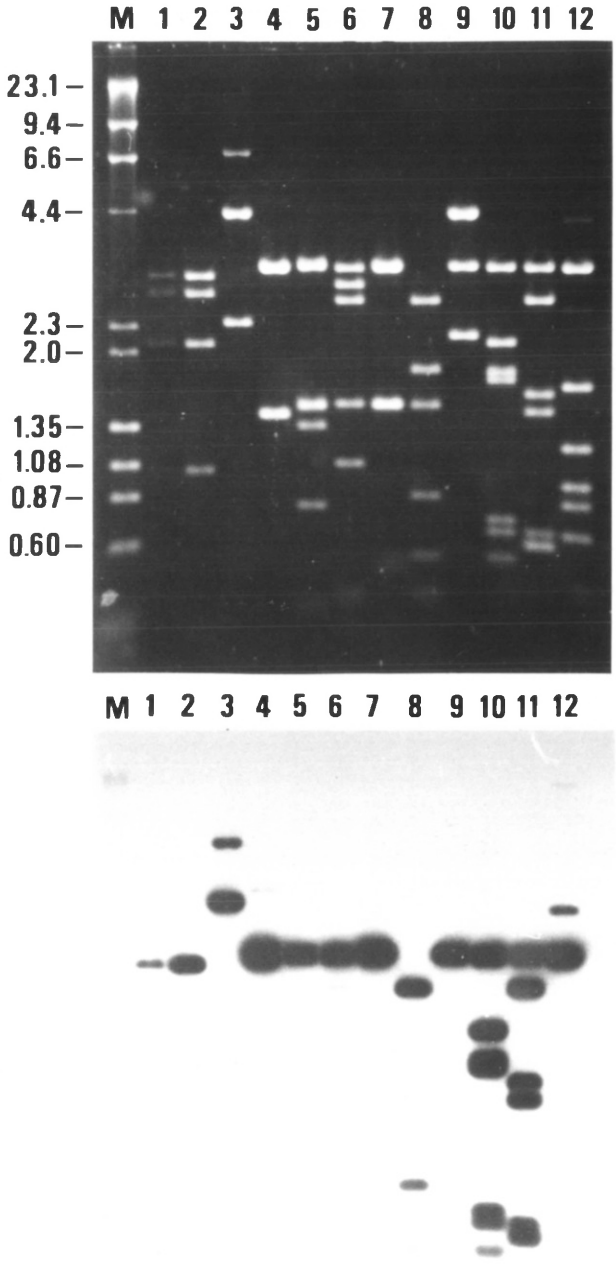
Figure 4.6 Internal repetition of DNA sequences in the *su(f)* region: pCSL probe

The upper panel shows an ethidium-bromide stained gel carrying restriction enzyme digests of plasmid subclones covering nearly all of the 32kb *su(f)* region from Canton S. The sizes of the marker bands in kb are shown on the left (lane M). The fainter bands visible in lanes 3 and 12 are partial digest products. 500ng of each plasmid digest was loaded per lane, except for lane 1, where only 50ng was loaded.

Lane M : size markers - λ digested with HindIII; ΦX174 digested with HaeIII.
Lane 1 : pB104"Sal" digested with SalI, BamHI and XhoI (50ng DNA)
Lane 2 : pB104"Sal" digested with SalI, BamHI and XhoI (500ng DNA)
Lane 3 : pm11.3 digested with EcoRI and HindIII (*white* DNA)
Lane 4 : pCS52 digested with EcoRI and XbaI
Lane 5 : pCS5 digested with EcoRI, XbaI and SalI
Lane 6 : pCS5 digested with EcoRI, XbaI and HindIII
Lane 7 : pCS51 digested with XbaI and BamHI
Lane 8 : pB1.R digested with EcoRI, XbaI, BamHI and SalI
Lane 9 : pBX64 digested with BamHI and SalI
Lane 10 : pCSL digested with EcoRI, BamHI, HindIII and XbaI
Lane 11 : pCSL digested with EcoRI, BamHI, HindIII and XhoI
Lane 12 : pCS4 digested with EcoRI and HindIII

The lower panel shows the DNA blot filter of this gel probed with pCSL (coordinates -11.6 to -4.3). In all lanes the plasmid backbone fragment hybridises. In addition to hybridising to itself (lanes 10 and 11), pCSL also hybridises to two of the insert fragments from pB1.R (lane 8). These are the 0.55kb XbaI fragment (coordinates +2.15 to +2.7) and the 0.9kb EcoRI-XbaI fragment (coordinates +2.7 to +3.6).

FIGURE 4.6



pCSL

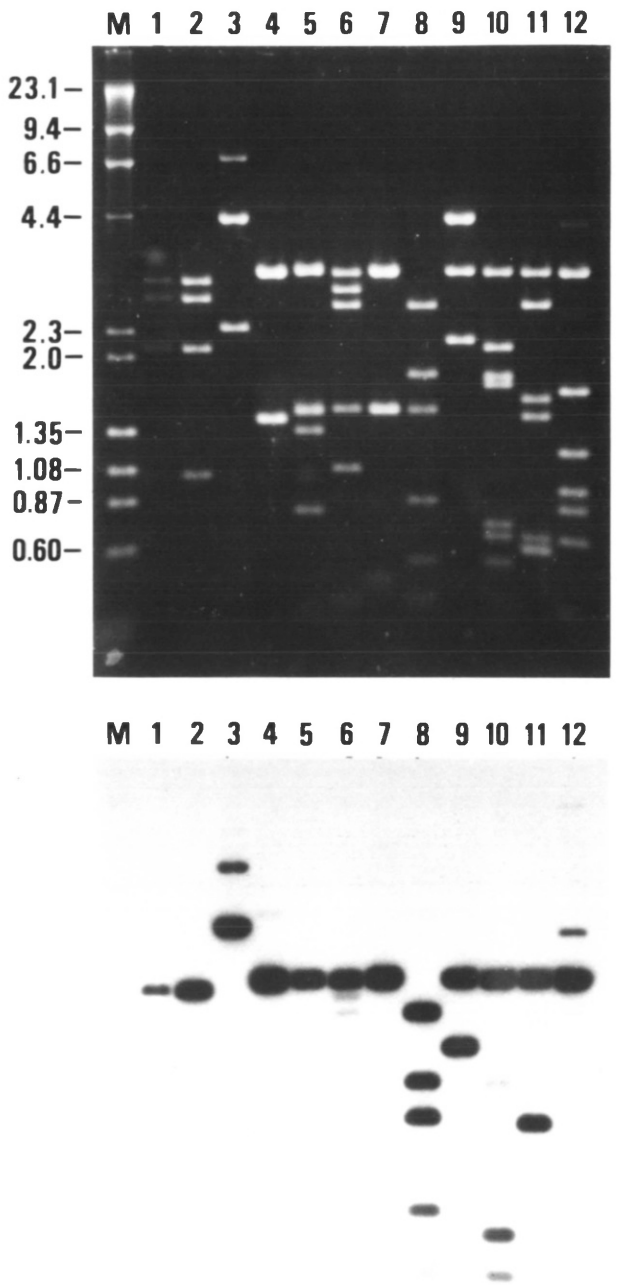
Figure 4.7 Internal repetition of DNA sequences in the *su(f)* region: pB1.R probe

The upper panel shows an ethidium-bromide stained gel carrying restriction enzyme digests of plasmid subclones covering nearly all of the 32kb *su(f)* region from Canton S. The sizes of the marker bands in kb are shown on the left (lane M). The fainter bands visible in lanes 3 and 12 are partial digest products. 500ng of each plasmid digest was loaded per lane, except for lane 1, where only 50ng was loaded.

Lane M : size markers - λdigested with HindIII; ΦX174 digested with HaeIII.
Lane 1 : pB104"Sal" digested with SalI, BamHI and XhoI (50ng DNA)
Lane 2 : pB104"Sal" digested with SalI, BamHI and XhoI (500ng DNA)
Lane 3 : pm11.3 digested with EcoRI and HindIII (*white* DNA)
Lane 4 : pCS52 digested with EcoRI and XbaI
Lane 5 : pCS5 digested with EcoRI, XbaI and SalI
Lane 6 : pCS5 digested with EcoRI, XbaI and HindIII
Lane 7 : pCS51 digested with XbaI and BamHI
Lane 8 : pB1.R digested with EcoRI, XbaI, BamHI and SalI
Lane 9 : pBX64 digested with BamHI and SalI
Lane 10 : pCSL digested with EcoRI, BamHI, HindIII and XbaI
Lane 11 : pCSL digested with EcoRI, BamHI, HindIII and XhoI
Lane 12 : pCS4 digested with EcoRI and HindIII

The lower panel shows the DNA blot filter of this gel probed with pB1.R (coordinates 0.0 to +5.1). In all lanes the plasmid backbone fragment hybridises. In addition to hybridising to itself (lane 8), pB1.R also hybridises to parts of pCSL (see lane 10). These are the 0.7kb HindIII-XbaI fragment (coordinates -5.0 to -5.65), the 0.55kb XbaI fragment (coordinates -5.65 to -6.2) and the 1.7kb XbaI-HindIII fragment (coordinates -6.2 to -7.9).

FIGURE 4.7



pB1.R

The cross-hybridising sequences correspond approximately to the regions between coordinates -6.4 to -5.0 and coordinates +2.15 to +3.5. The extent of the cross-homology is in the order of 1.5kb. These regions have a similar restriction site map indicating that each repeat copy is orientated in the same direction (Figure 4.5); furthermore, the 0.55kb XbaI-XbaI and the adjoining 0.7kb XbaI-HindIII fragments from each were shown to co-migrate in agarose gels (data not shown). The 0.55kb XbaI fragment from coordinates -5.65 to -6.2 was subcloned into an M13 vector to give the clone X5L (see Figure 4.9); this was used as a hybridisation probe on a DNA blot filter carrying DNA from the cloned interval (described above). The result is shown in Figure 4.8, and shows that X5L hybridises to itself and also to the 0.55kb XbaI fragment from the right hand copy of the 1.5kb repeat (coordinates +2.15 to +2.7). What restriction map data there is available for these repeats does not match any published map of *D.melanogaster* transposable elements (Finnegan and Fawcett, 1986).

Additional M13 subclones from the left hand copy and one M13 subclone from the right hand copy were analysed so as to obtain nucleotide sequence data covering most of the repeat. Figure 4.9a shows the position of the M13 subclones and extent of the sequence data obtained. Analysis of the data failed to reveal any features characteristic of transposable elements such as inverted or direct repeats or possible target site duplications. Neither did the nucleotide sequence show any homology to known sequences of *D.melanogaster* repetitive DNA. In addition, sequence data from the left hand copy was compared to the sequence of the M13 subclone from the right hand copy. Confirmation that at least part of the left hand and right hand copies of this repeat share identity at the sequence level is shown in Figure 4.9b: a comparison of the nucleotide sequences of the left hand ends of each of the repeat copies. In the 148bp region of homology shown, there are only two mismatches.

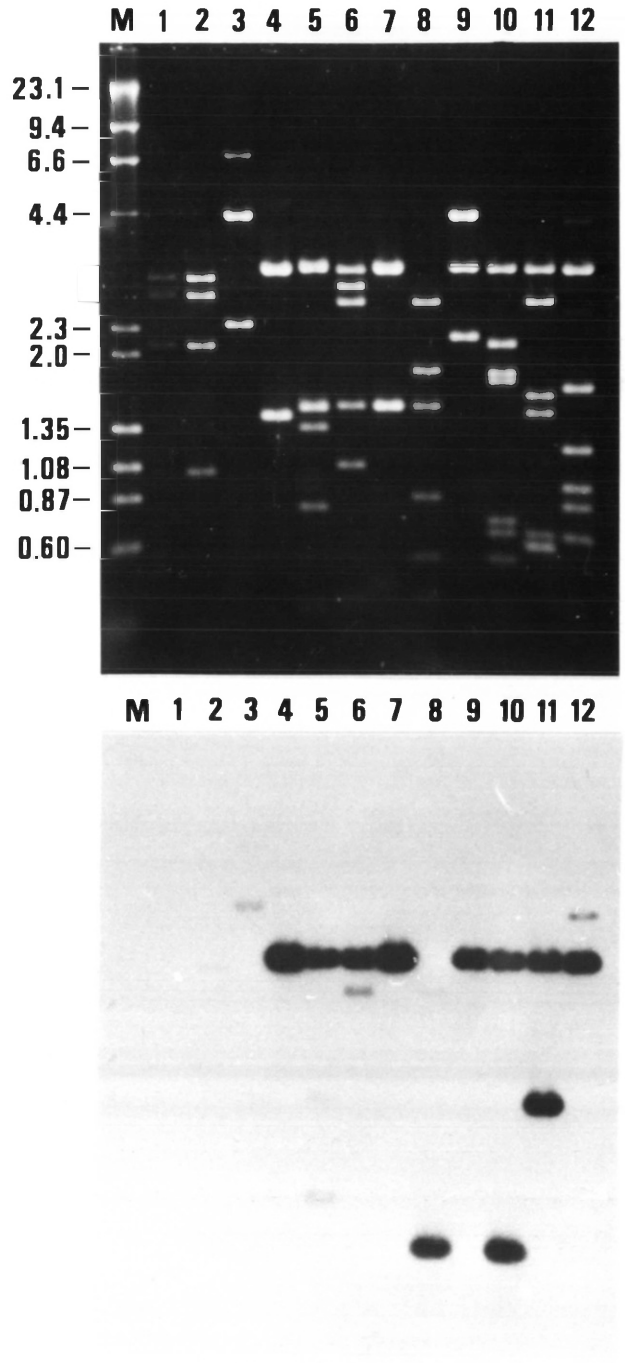
Figure 4.8 Internal repetition of DNA sequences in the *su(f)* region: X5L probe

The upper panel shows an ethidium-bromide stained gel carrying restriction enzyme digests of plasmid subclones covering nearly all of the 32kb *su(f)* region from Canton S. The sizes of the marker bands in kb are shown on the left (lane M). The fainter bands visible in lanes 3 and 12 are partial digest products. 500ng of each plasmid digest was loaded per lane, except for lane 1, where only 50ng was loaded.

Lane M : size markers - λ digested with HindIII; ΦX174 digested with HaeIII.
Lane 1 : pB104"Sal" digested with SalI, BamHI and XhoI (50ng DNA)
Lane 2 : pB104"Sal" digested with SalI, BamHI and XhoI (500ng DNA)
Lane 3 : pm11.3 digested with EcoRI and HindIII (*white* DNA)
Lane 4 : pCS52 digested with EcoRI and XbaI
Lane 5 : pCS5 digested with EcoRI, XbaI and SalI
Lane 6 : pCS5 digested with EcoRI, XbaI and HindIII
Lane 7 : pCS51 digested with XbaI and BamHI
Lane 8 : pB1.R digested with EcoRI, XbaI, BamHI and SalI
Lane 9 : pBX64 digested with BamHI and SalI
Lane 10 : pCSL digested with EcoRI, BamHI, HindIII and XbaI
Lane 11 : pCSL digested with EcoRI, BamHI, HindIII and XhoI
Lane 12 : pCS4 digested with EcoRI and HindIII

The lower panel shows the DNA blot filter of this gel probed with the M13 subclone, X5L (coordinates -5.65 to -6.2; see Figure 4.9). This probe hybridises to vector fragments which carry *lac* or M13 sequences. In addition to hybridising to itself (lane 10), X5L also hybridises to the identically sized XbaI fragment in pB1.R (lane 8; coordinates +2.15 to +2.7).

FIGURE 4.8



X5L

Figure 4.9 Sequence analysis of the "1.5kb" repeats

Part A shows a restriction enzyme site map of part of the *su(f)* region. The positions of the two tandemly orientated copies of the "1.5kb" repeats are represented by thick black lines. Below the map are shown the extent of the nucleotide sequence readings of the "1.5kb" repeats; the arrow indicates the direction in which the sequence was read. Also marked are the two clones which read across the left hand ends of each repeat, XHH9 and SX9; and the position of X5L (probe in Figure 4.8).

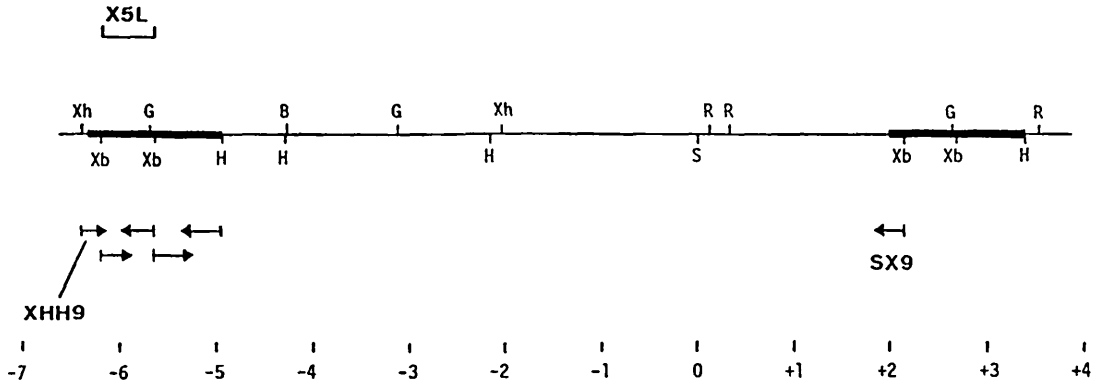
Restriction site key:

B : BamHI	S : SalI
H : HindIII	Xh : XhoI
R : EcoRI	Xb : XbaI
G : BglII	

Part B shows a nucleotide sequence comparison of XHH9 and SX9; reading from flanking DNA sequences into the left hand ends of each copy of the "1.5kb" repeat. The extent of the homology is shown. Also shown are the positions of the XhoI site in XHH9, and the XbaI sites in both XHH9 and SX9.

FIGURE 4.9

A.



B.

```
            XhoI      .         20        .         .         50
XHH9  :     CTCGAGGCCATCGGGTTGGCATCAAACATCAAGCCATAGTCATCTTAACA

SX9   :     TTGTGTTTACTTTTTTGTAAAATAATTCAATATGCAACCATTACCACGAA

                      .         70        .         .         100
XHH9  :     GCTACCATTCTTTTGCTTGGACTAGAAAAAGCATACGAAATTAGAGAAAT
                      ||||||||||||||||||||||||||||||||||| |||||
SX9   :     ATCAGTTGTTCTTTGCTTGGACTAGAAAAAGCATACGAAATTAGGGAAAT

                      .         120       .         .         150
XHH9  :     ACCTATACAAAATTAGTTAATTAAATATAAGAGGCGGTGCCACGGCCACA
            |||||||||||||||| ||||||||||||||||||||||||||||||||
SX9   :     ACCTATACAAAATTAATTAATTAAATATAAGAGGCGGTGCCACGGCCACA

                      .         170       .         .         200
XHH9  :     ATTTTTGGTATCATAAAAGATAATGGCAAAAAAATTTAAATATCTTTATC
            ||||||||||||||||||||||||||||||||||||||||||||||||||
SX9   :     ATTTTTGGTATCATAAAAGATAATGGCAAAAAAATTTAAATATCTTTATC

              XbaI
XHH9  :     CGTCTAGA
            ||||||||
SX9   :     CGTCTAGA
              XbaI
```

4.3.2 "REVERSE SOUTHERN" DNA BLOTTING EXPERIMENTS

The filter bearing restriction enzyme digest fragments covering almost all the cloned interval was stripped of probe and hybridised to nick-translated Oregon R DNA. In this type of experiment, those fragments containing sequences which occur frequently within the genome will give a stronger autoradiographic signal than fragments bearing only unique sequences. To aid interpretation of the results, similar amounts of DNA of two control clones were included on the filter. The first is pm11.3, a 2.3kb EcoRI-HindIII fragment of unique DNA from the *white* locus carrying the region from +6.7 to +9.0 in the coordinate system of Levis *et al.* (1982). The second is pB104"Sal", the 5.7kb internal SalI fragment from a B104 transposable element (R. Karess, unpublished). The copy number of B104 elements in the haploid genome of *D.melanogaster* is approximately 80 (Scherer *et al.*, 1982).

The results of the "reverse-Southern" DNA blot are shown in Figure 4.10. It is immediately apparent that most of the restriction enzyme fragments from the *su(f)* region give a stronger autoradiographic signal than that of the fragment of *white* DNA, demonstrating the presence of repetitive DNA. Below are listed those fragments which do not give a stronger signal than pm11.3, and are therefore designated "unique":

- the interval from the BamHI site at coordinate +5.1 to the XbaI site at coordinate +6.6

- the interval from the SalI site at coordinate +7.4 to the SalI site at coordinate +7.8

- the interval from the HindIII at coordinate +12.1 to the end of the cloned region at coordinate +13.9

In addition, the fragments covering the interval from the HindIII site at coordinate -5.0 to the XbaI site at coordinate +2.15 give only a faint signal, suggesting that this interval consists mostly of unique sequences. This region of

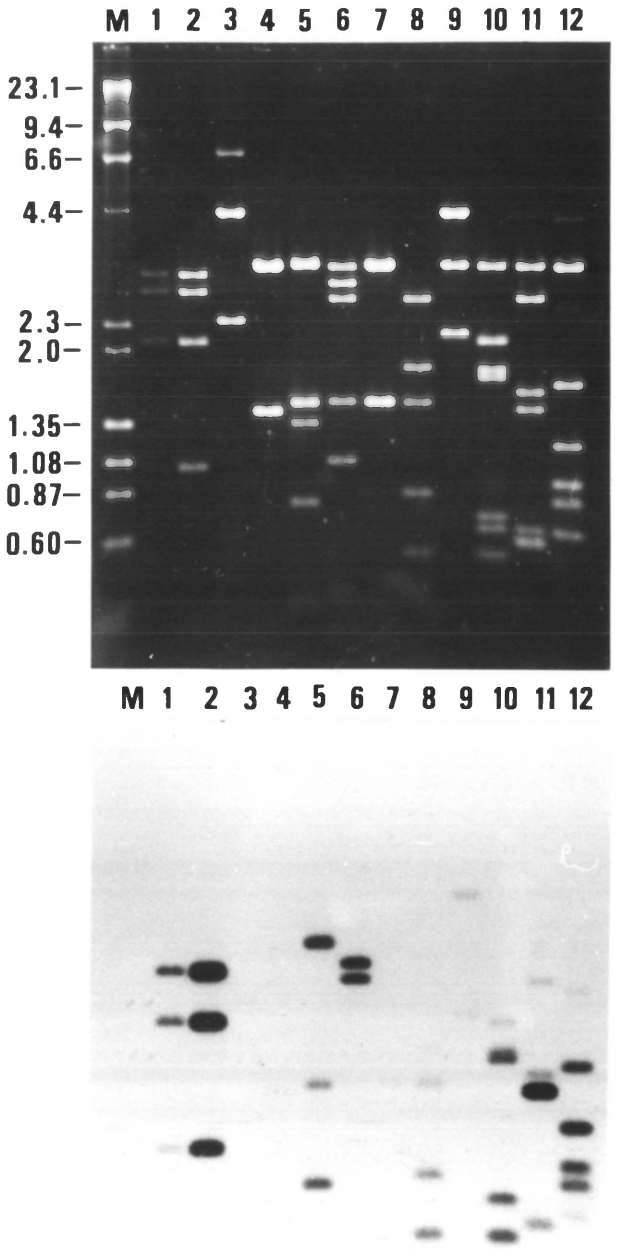
Figure 4.10 Sequences from the *su(f)* region are found elsewhere in the genome

The upper panel shows an ethidium-bromide stained gel carrying restriction enzyme digests of plasmid subclones covering nearly all of the 32kb *su(f)* region from Canton S. The sizes of the marker bands in kb are shown on the left (lane M). The fainter bands visible in lanes 3 and 12 are partial digest products. 500ng of each plasmid digest was loaded per lane, except for lane 1, where only 50ng was loaded.

Lane M : size markers - λ digested with HindIII; ΦX174 digested with HaeIII.
Lane 1 : pB104"Sal" digested with SalI, BamHI and XhoI (50ng DNA)
Lane 2 : pB104"Sal" digested with SalI, BamHI and XhoI (500ng DNA)
Lane 3 : pm11.3 digested with EcoRI and HindIII (*white* DNA)
Lane 4 : pCS52 digested with EcoRI and XbaI
Lane 5 : pCS5 digested with EcoRI, XbaI and SalI
Lane 6 : pCS5 digested with EcoRI, XbaI and HindIII
Lane 7 : pCS51 digested with XbaI and BamHI
Lane 8 : pB1.R digested with EcoRI, XbaI, BamHI and SalI
Lane 9 : pBX64 digested with BamHI and SalI
Lane 10 : pCSL digested with EcoRI, BamHI, HindIII and XbaI
Lane 11 : pCSL digested with EcoRI, BamHI, HindIII and XhoI
Lane 12 : pCS4 digested with EcoRI and HindIII

The lower panel shows the DNA blot filter of this gel probed with nick-translated Oregon R genomic DNA. Repetitive DNA fragments give a strong hybridisation signal (e.g. the B104 fragments in lanes 1 and 2); unique fragments give no or a barely detectable signal (e.g. the pm11.3 *white* DNA insert in lane 3). Most of the fragments from the *su(f)* region hybridise to a greater or lesser extent; but some do not (e.g. pCS52 - lane 4, and pCS51 - lane 7), and are therefore unique.

FIGURE 4.10



Oregon R

approximately 7kb includes the whole of pBX64, and is bounded at each end by tandemly orientated copies of the "1.5kb" repeat. It includes the site of insertion of the P element in $su(f)^{l\text{-}MS252}$. The 2.1kb fragment from the SalI site at coordinate 0.0 to the XbaI site at coordinate +2.15 gives a detectable signal. This is presumably due to the ~150bp of DNA at the XbaI end from the "1.5kb" repeat sequences previously described. When pBX64 is used as a hybridisation probe for genomic DNA blots, the signal patterns obtained are consistent with it bearing unique DNA (see for example Figures 4.11 and 4.12), although some fainter bands are visible after long exposures.

For the restriction enzyme fragments that give a signal characteristic of repetitive DNA, the technique cannot of course detect regions within these fragments that are in fact unique. By comparison to the signals yielded by the B104 fragments, most of the fragments from the *su(f)* region apparently contain repetitive sequences of a low copy number, or with poor homology to other sequences. However, some of the fragments give a much stronger signal, suggesting a copy number of perhaps 20-40. This latter group of fragments in fact correspond to those bearing DNA from the two "1.5kb" repeat sequences described previously.

4.4 DNA BLOTTING ANALYSIS OF MUTANT *su(f)* ALLELES

Eighteen mutant *su(f)* alleles were studied by DNA blotting experiments to examine their restriction maps in the interval from the XhoI site at coordinate -6.4 to the SalI site at coordinate +7.4. This interval includes that covered by pBX64 and into which the P element of $su(f)^{l\text{-}MS252}$ is inserted. In the case of the four viable *su(f)* alleles ($su(f)^{1}$, $su(f)^{l\text{-}ts67g}$, $su(f)^{l\text{-}ts76a}$ and $su(f)^{l\text{-}mad.ts}$) DNA from homozygous flies was used, but each of the fourteen lethal alleles was first balanced over FM6 and DNA prepared from heterozygous female flies. The FM6 chromosome has the same restriction enzyme site map as Canton S in this interval (data not shown). The complete list of *su(f)* alleles studied by blotting

experiments is shown below, together with the nature of the mutagenic agent (if any) used to generate them:

Allele	Mutagen
su(f)^{1}	X-ray
su(f)$^{l-ts67g}$	EMS
su(f)$^{l-ts76a}$	EMS
su(f)$^{l-mad.ts}$	EMS
su(f)$^{R-9-18}$	EMS
su(f)$^{l-L26}$	X-ray
su(f)$^{l-D13}$	EMS
su(f)$^{l-3DES}$	diethyl sulphate
Df(1)su(f)GA130	X-ray
su(f)$^{l-S1}$	spontaneous
su(f)$^{l-S2}$	spontaneous
su(f)$^{l-X1}$	X-ray
su(f)$^{l-X2}$	X-ray
su(f)$^{l-X3}$	X-ray
su(f)$^{l-16.1.85}$	neutron
su(f)$^{l-16.3.162}$	neutron
su(f)$^{l-MS97\_1}$	PM hybrid dysgenesis
su(f)$^{l-MS252\_1}$	PM hybrid dysgenesis

The DNA samples were digested with either SalI and XhoI, or with BamHI and EcoRI, and the resulting filters were hybridised with pBX64. The results are shown in Figure 4.11 (SalI+XhoI digests) and Figure 4.12 (BamHI+EcoRI digests). For the nine alleles which showed a pattern of bands differing from Canton S, additional DNA blotting experiments with a wider range of restriction enzymes were performed to define the mutant restriction site map further (data not shown).

The results of the genomic DNA blot mapping of the *su(f)* alleles are summarised in Figure 4.13. None of the four viable alleles showed a map which differed detectably from Canton S. In addition, five of the lethal alleles were

Figure 4.11 Alterations in DNA from *su(f)* mutants : SalI and XhoI digests

The two panels show DNA blots of 20 samples digested with SalI and XhoI and probed with pBX64. Approximately 1.5ug of each genomic DNA sample was loaded per lane. On the right of each panel the sizes of the wild type SalI/XhoI fragments are marked in kb. On the left of each panel the sizes of the marker bands are shown in kb. Size marker bands were provided by λ digested with HindIII and ΦX174 digested with HaeIII (lane M).

Sample 1 : Canton S
Sample 2 : *su(f)*^{1}
Sample 3 : *su(f)*$^{l\text{-}ts67g}$
Sample 4 : *su(f)*$^{l\text{-}ts76a}$
Sample 5 : *su(f)*$^{l\text{-}mad.ts}$
Sample 6 : *su(f)*$^{R\text{-}9\text{-}18}$
Sample 7 : *su(f)*$^{l\text{-}L26}$
Sample 8 : *su(f)*$^{l\text{-}D13}$
Sample 9 : *su(f)*$^{l\text{-}3DES}$
Sample 10 : *Df(1)su(f)GA130*

Sample 11 : Canton S
Sample 12 : *su(f)*$^{l\text{-}S1}$
Sample 13 : *su(f)*$^{l\text{-}S2}$
Sample 14 : *su(f)*$^{l\text{-}X1}$
Sample 15 : *su(f)*$^{l\text{-}X2}$
Sample 16 : *su(f)*$^{l\text{-}X3}$
Sample 17 : *su(f)*$^{l\text{-}16.1.85}$
Sample 18 : *su(f)*$^{l\text{-}16.3.162}$
Sample 19 : *su(f)*$^{l\text{-}MS97\_2}$
Sample 20 : *su(f)*$^{l\text{-}MS252}$

All DNA samples were made from female flies balanced over *FM6*, except for those in lanes 1, 2, 3, 4, 5 and 11, which were made from unsexed flies of homozygous stocks. In sample 20, the *su(f)*$^{l\text{-}MS252}$ stock was balanced over *FM7*. All other genetic markers in the stocks have been omitted for clarity.

Beneath the lower panel is a map of the *su(f)* region, showing the extent of pBX64 and the relevant restriction enzyme sites.

Restriction site key: B : BamHI S : SalI
X : XhoI R : EcoRI

FIGURE 4.11

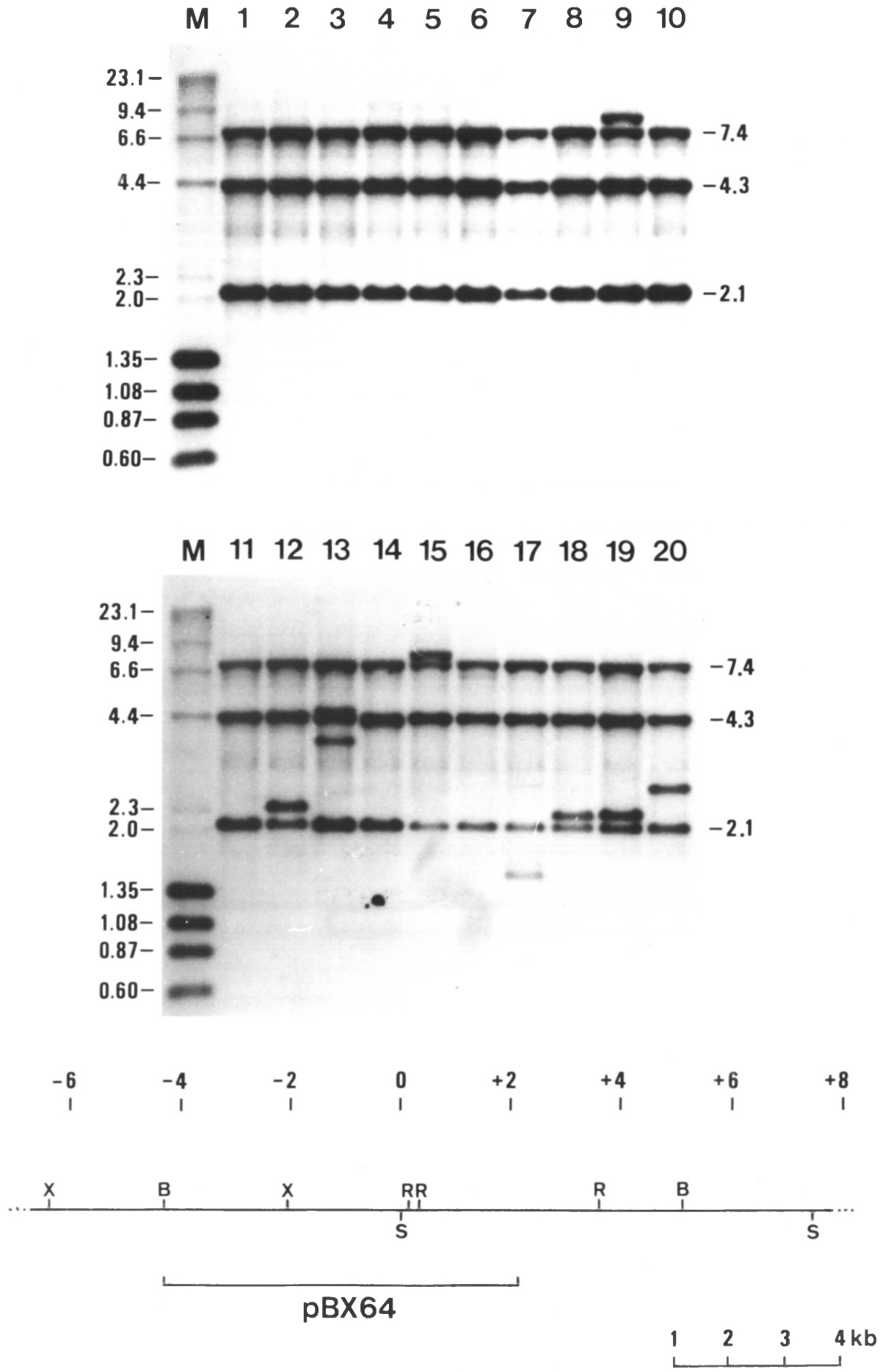

M 1 2 3 4 5 6 7 8 9 10
23.1–
9.4–
6.6–
4.4–
2.3–
2.0–
1.35–
1.08–
0.87–
0.60–
–7.4
–4.3
–2.1
M 11 12 13 14 15 16 17 18 19 20
23.1–
9.4–
6.6–
4.4–
2.3–
2.0–
1.35–
1.08–
0.87–
0.60–
–7.4
–4.3
–2.1
-6 -4 -2 0 +2 +4 +6 +8
X B X RR R B
S S
pBX64
1 2 3 4 kb

Figure 4.12 Alterations in DNA from *su(f)* mutants : BamHI and EcoRI digests

The two panels show DNA blots of 20 samples digested with BamHI and EcoRI and probed with pBX64. Approximately 1.5ug of each genomic DNA sample was loaded per lane. On the right of each panel the sizes of the wild type BamHI/EcoRI fragments are marked in kb. On the left of each panel the sizes of the marker bands are shown in kb. Size marker bands were provided by λ digested with HindIII and ΦX174 digested with HaeIII (lane M).

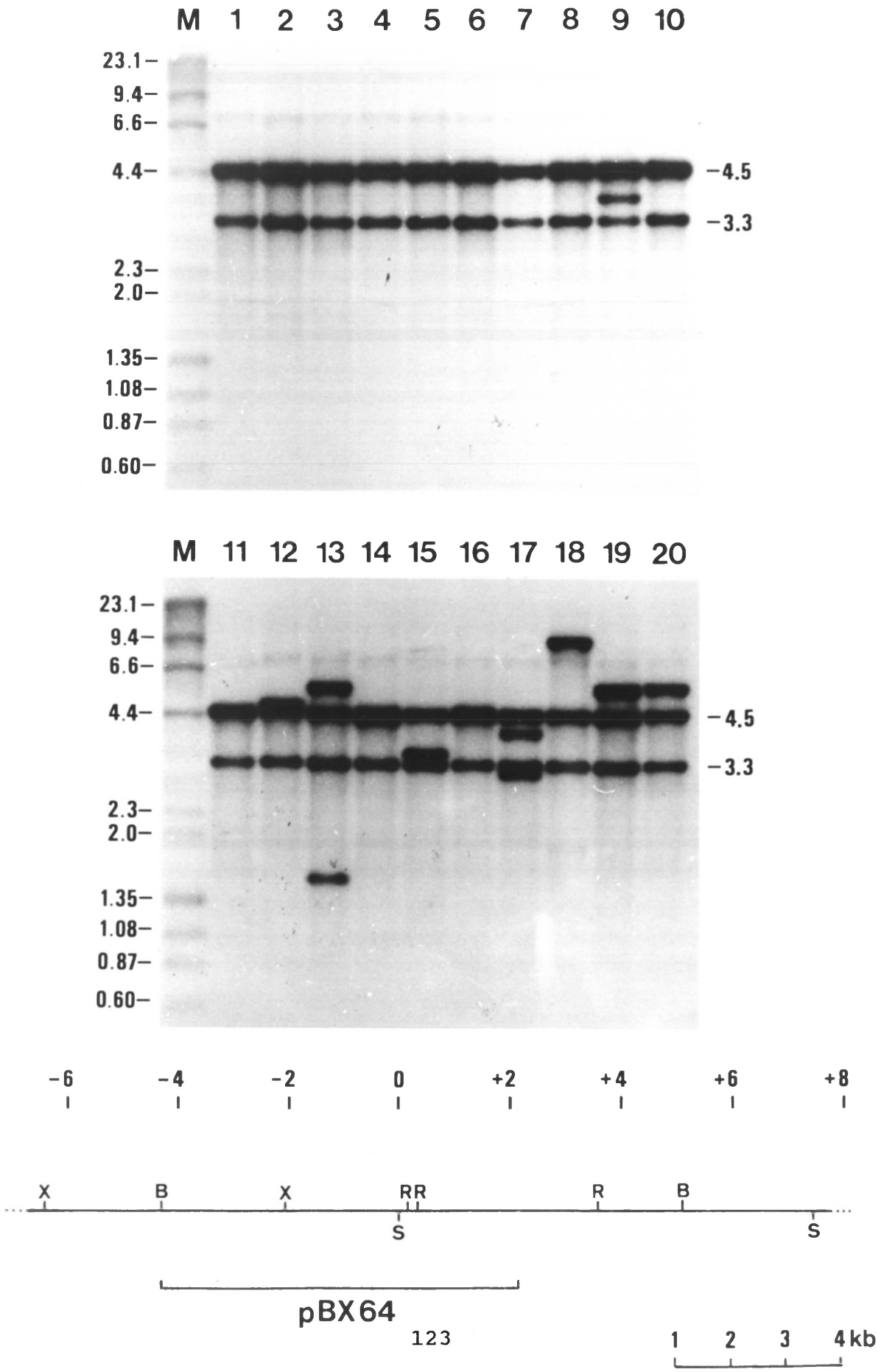
Sample 1 : Canton S
Sample 2 : $su(f)^{1}$
Sample 3 : $su(f)^{l\text{-}ts67g}$
Sample 4 : $su(f)^{l\text{-}ts76a}$
Sample 5 : $su(f)^{l\text{-}mad.ts}$
Sample 6 : $su(f)^{R\text{-}9\text{-}18}$
Sample 7 : $su(f)^{l\text{-}L26}$
Sample 8 : $su(f)^{l\text{-}D13}$
Sample 9 : $su(f)^{l\text{-}3DES}$
Sample 10 : *Df(1)su(f)GA130*
Sample 11 : Canton S
Sample 12 : $su(f)^{l\text{-}S1}$
Sample 13 : $su(f)^{l\text{-}S2}$
Sample 14 : $su(f)^{l\text{-}X1}$
Sample 15 : $su(f)^{l\text{-}X2}$
Sample 16 : $su(f)^{l\text{-}X3}$
Sample 17 : $su(f)^{l\text{-}16.1.85}$
Sample 18 : $su(f)^{l\text{-}16.3.162}$
Sample 19 : $su(f)^{l\text{-}MS97\_2}$
Sample 20 : $su(f)^{l\text{-}MS252}$

All DNA samples were made from female flies balanced over *FM6*, except for those in lanes 1, 2, 3, 4, 5 and 11, which were made from unsexed flies from homozygous stocks. In sample 20, the $su(f)^{l\text{-}MS252}$ stock was balanced over *FM7*. All other genetic markers in the stocks have been omitted for clarity.

Beneath the lower panel is a map of the *su(f)* region, showing the extent of pBX64 and the relevant restriction enzyme sites.

Restriction site key: B : BamHI S : SalI
X : XhoI R : EcoRI

FIGURE 4.12


M 1 2 3 4 5 6 7 8 9 10
23.1
9.4
6.6
4.4
2.3
2.0
1.35
1.08
0.87
0.60
4.5
3.3
M 11 12 13 14 15 16 17 18 19 20
23.1
9.4
6.6
4.4
2.3
2.0
1.35
1.08
0.87
0.60
4.5
3.3
-6 -4 -2 0 +2 +4 +6 +8
X B X RR R B
S S
pBX64
123
1 2 3 4kb

indistinguishable from Canton S. However, the nine lethal alleles which did differ show a range of mutant structures, described in turn below:

su(f)$^{l-X1}$:

A ~150bp deletion in the interval between coordinates -2.2 and -4.3.

su(f)$^{l-X2}$:

A ~1kb deletion which removes the SalI site at coordinate 0.0 but not the EcoRI site at coordinate +0.35.

su(f)$^{l-MS252}$:

A 1.1kb P element insertion 150bp to the left of the SalI site at coordinate 0.0.

su(f)$^{l-MS97}$:

A 1.1kb insertion ~200bp to the left of the SalI site at coordinate 0.0, whose restriction enzyme site map is consistent with that of a non-autonomous, 1.1kb P element.

su(f)$^{l-S2}$:

A 4.5kb insertion in the interval between coordinates -2.2 and -4.3. The insert as drawn in Figure 4.13 has two EcoRI recognition sites, but the nature of these blotting experiments means that there may be one or more EcoRI sites between these two. The restriction enzyme site map is similar to that of a Doc element (Bender *et al.*, 1983).

su(f)$^{l-S1}$:

A small insertion or tandem duplication of ~250bp in the interval between coordinates 0.0 and -2.1.

su(f)$^{l-3DES}$:

The size of the interval between the HindIII site at coordinate +3.3 and the EcoRI site at coordinate +3.5 is larger than that of Canton S by some 500bp, either due to an insertion or a tandem duplication.

su(f)$^{l-16.1.85}$:

In Figure 4.11, the 2.1kb SalI-XhoI band is submolar, to be replaced by two new bands of ~1.5 and ~1.6kb. Two new bands are also visible in the BamHI-EcoRI digest (Figure 4.12). These and other DNA blotting data suggest that this allele is associated either with the insertion of foreign DNA sequences or an inversion breakpoint in the interval between coordinates 0.0 and -2.1.

su(f)$^{l-16.3.162}$:

In this allele, the 2.1kb SalI-XhoI band is submolar, and one new band is visible (Figure 4.11). Two new bands are visible in the BamHI-EcoRI digest; one of ~10kb and a faint band at ~1.8kb (Figure 4.12). As for the previous allele, these data suggest that this allele is associated either with a large insertion of DNA or an inversion breakpoint in the interval between coordinates 0.0 and -2.1.

In addition to these alleles, the bands in the *su(f)*$^{l-L26}$*/FM6* sample appear submolar. This suggests that the DNA region covered by the probe, pBX64, may be completely deleted in this mutant. In the absence of any control for the relative amounts of DNA loaded in the lanes, this conclusion remains tentative, but it should be noted that the other samples of *su(f)*^{lethal}*/FM6* DNA all show very similar band intensities (Figures 4.11 and 4.12).

The striking feature of these lesions is that all but one fall in the interval between the SalI site at coordinate 0.0 and the BamHI site at coordinate -4.3. The exception is *su(f)*$^{l-3-DES}$, whose restriction site map differs from wild type between coordinates +3.3 and +3.5. The possibility remains that the true lesion in this chemically-induced mutant is a point mutation lying within the same region where the other mutations cluster, and that the observed difference is phenotypically silent. Examining the structure of a revertant of this allele could perhaps answer this question; indeed, a revertant of *su(f)*$^{l-3-DES}$ was once isolated (Voss and Falk, 1973), but is now apparently lost.

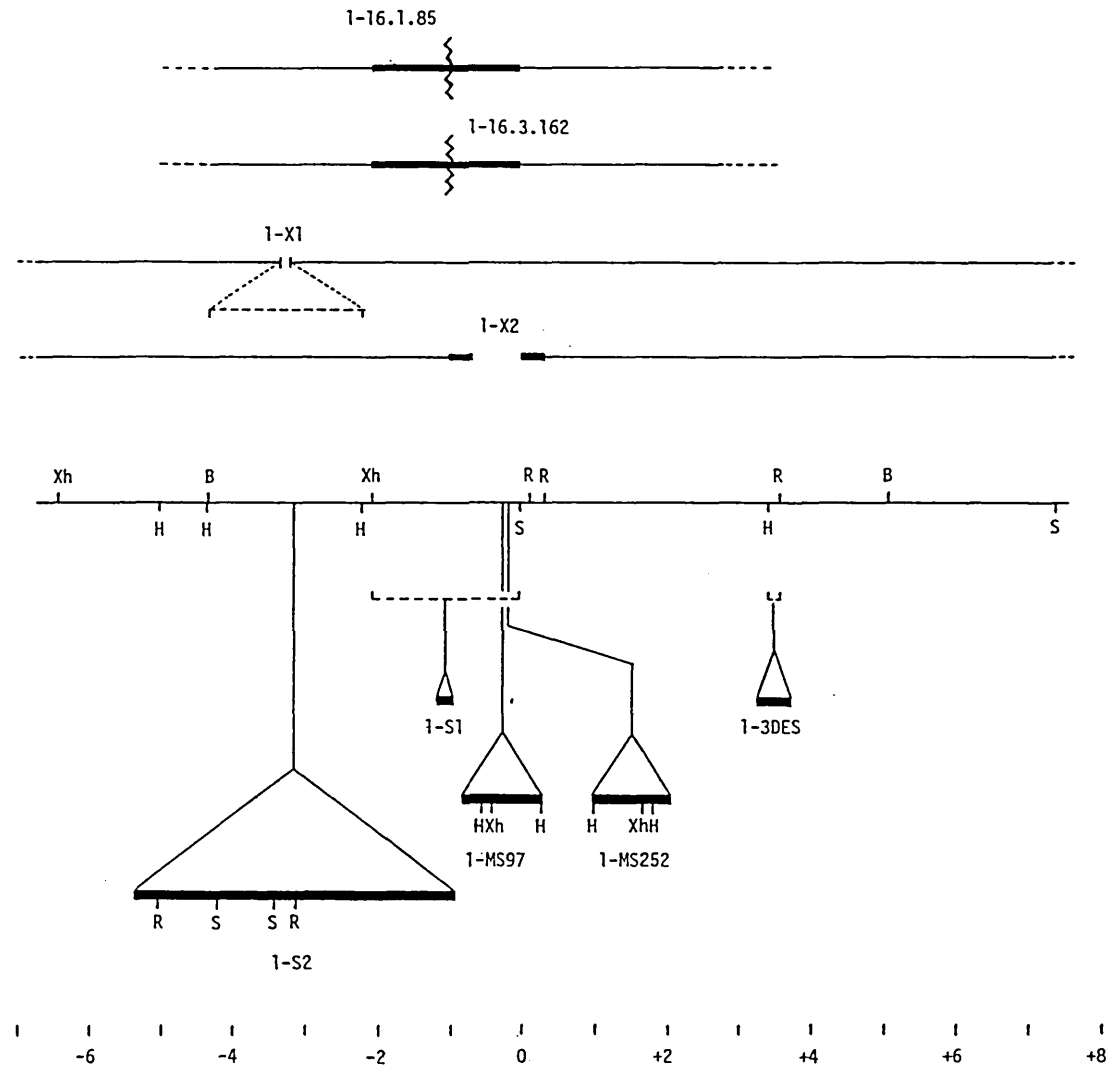
Figure 4.13 DNA lesions associated with *su(f)* mutant alleles

Shown is a diagram of the *su(f)* region indicating the nature and position of DNA lesions associated with various mutant alleles of *su(f)*. Above the line are marked the two deletion mutants, $su(f)^{l-X1}$ and $su(f)^{l-X2}$; as well as $su(f)^{l-16.1.85}$ and $su(f)^{l-16.3.162}$ which are either large insertions or chromosomal breakpoints. The extent of uncertainty in the positions of these lesions are indicated by the thick black lines, or by the dotted line interval. Below the line are shown the lesions associated with insertions. Two of these are P element insertions: $su(f)^{l-MS97}$ and $su(f)^{l-MS252}$. The ~4.5kb insertion in $su(f)^{l-S2}$ has a restriction enzyme site map of a Doc element. The two smaller lesions found in $su(f)^{l-S1}$ and $su(f)^{l-3DES}$ are either insertions of foreign DNA or small duplications.

Restriction site key:

B : BamHI
Xh : XhoI
H : HindIII
S : SalI
R : EcoRI

FIGURE 4.13


1-16.1.85
1-16.3.162
1-X1
1-X2
Xh
B
Xh
R R
R
B
H H
H
S
H
S
1-S1
1-3DES
HXh H
H XhH
1-MS97
1-MS252
R S S R
1-S2
-6
-4
-2
0
+2
+4
+6
+8

4.5 CLONING AND ANALYSIS OF *su(f)* DNA FROM OTHER *su(f)* MUTANTS

4.5.1 CLONING OF *su(f)* DNA FROM $su(f)^{l\text{-}MS97}$

It is already known that the lethal phenotype of $su(f)^{l\text{-}MS252}$ is due to the insertion of a P element. In the case of $su(f)^{l\text{-}MS97}$, the genetic data from Chapter 3 suggests that this allele might also be a P element insertion; a conclusion considerably supported by the DNA blotting data just described. From DNA blotting, it is apparent that the insertion in $su(f)^{l\text{-}MS97}$ is of a similar size to that in $su(f)^{l\text{-}MS252}$, but the data from the BamHI-EcoRI digest shows that the insertions are not identical (Figure 4.12). DNA from the *su(f)* region of this mutant allele was thus cloned for further analysis.

A BamHI limit-digest library of DNA from *w m* f^1 $su(f)^{l\text{-}MS97}$*-2/FM6* females was constructed in λ EMBL4. Following in vitro packaging and plating out of the recombinants using *E.coli* Q359, the library filters were hybridised to pBX64. A number of positively hybridising clones were obtained. On the justifiable assumption that the lesion is a P element insertion, the filters were also counterscreened with pπ25.7BWC to exclude clones deriving from the FM6 chromosome. One of the clones hybridising to both probes (λ 97.14) was analysed by mapping restriction enzyme sites and performing DNA blotting experiments using P element hybridisation probes (data not shown). The restriction enzyme site map of the insert in λ 97.14 is shown in Figure 4.14a. The data shows that there is indeed a P element lying within the *su(f)* locus in $su(f)^{l\text{-}MS97}$. It is of the same size as that in $su(f)^{l\text{-}MS252}$, but inserted in the opposite orientation to the latter, about 100bp further to the left.

4.5.2 SEQUENCE ANALYSIS OF THE P ELEMENT INSERTIONS IN $su(f)^{l\text{-}MS97}$ AND $su(f)^{l\text{-}MS252}$

To further characterise the P element insertions in the mutations $su(f)^{l\text{-}MS97}$ and $su(f)^{l\text{-}MS252}$, suitable fragments from λ 97.14 and λ 252.B1 were subcloned into

Figure 4.14 Nucleotide sequence analysis of the P element insertions in *su(f)*$^{l\text{-}MS97}$ and *su(f)*$^{l\text{-}MS252}$

Part A shows the restriction enzyme site maps of the inserts in λ252.B1 and λ97.14, compared to the map of the corresponding region from Canton S (pBX64). P element sequences are represented by thick black lines. Underneath each map are marked the extent of the nucleotide sequences determined; the arrow indicates the direction of the sequencing.

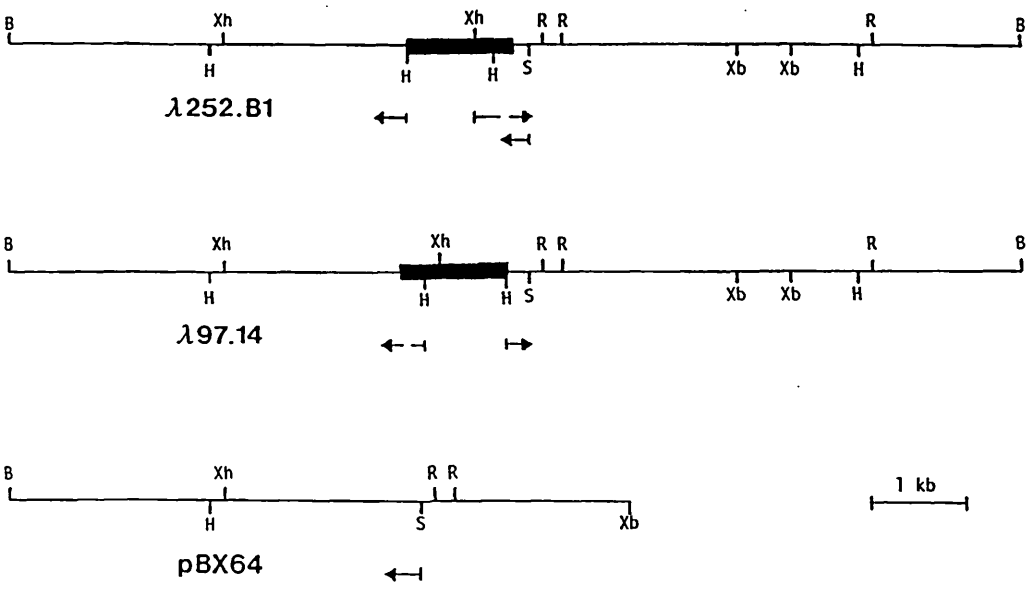
Part B shows the nucleotide sequences reading across the left and right hand ends of the two P elements, compared to the sequence from pπ25.1. Sequences from *su(f)* are in lower case letters; P element sequences are in capital letters. The sequences underlined are the 8bp target site duplications formed on insertion of the P elements. The asterisks (*) mark the single base pair polymorphism at P element position 33 seen in both elements.

Part C shows the position of the internal deletion of the P elements in *su(f)*$^{l\text{-}MS97}$ and *su(f)*$^{l\text{-}MS252}$, by comparison to the sequence of an intact P element. Both have identical deletions, removing P element sequences from 894 to 2685 inclusive. The numbering system of the sequence of pπ25.1 is that of the P element sequence reported by O'Hare and Rubin (1983).

Part D is the Canton S nucleotide sequence reading from the SalI site at position 0.0 in the "-" direction, showing the sites of insertion of the P elements in *su(f)*$^{l\text{-}MS97}$ and *su(f)*$^{l\text{-}MS252}$. Nucleotide -1 is the first nucleotide of the sequence given; the "+"-most G residue of the SalI site. The sequences underlined are the 8bp target site duplications of the two P elements.

FIGURE 4.14

A.



B.

Left hand termini:

su(f)$^{l\text{-}MS97}$
ttacgagtcgctaggtCATGATGAAATAACATAAGGTGGTCCCGTCGATAGCCGAAGCTT

su(f)$^{l\text{-}MS252}$
tcatgacactccaggaCATGATGAAATAACATAAGGTGGTCCCGTCGATAGCCGAAGCTT

10 20 30 40
pπ25.1 CATGATGAAATAACATAAGGTGGTCCCGTCGAAAGCCGAAGCTT
HindIII

Right hand termini:

su(f)$^{l\text{-}MS97}$
CGACGGGACCACCTTATGTTATTTCATCATGcgctaggtaatgtgtttccgacaaccgct

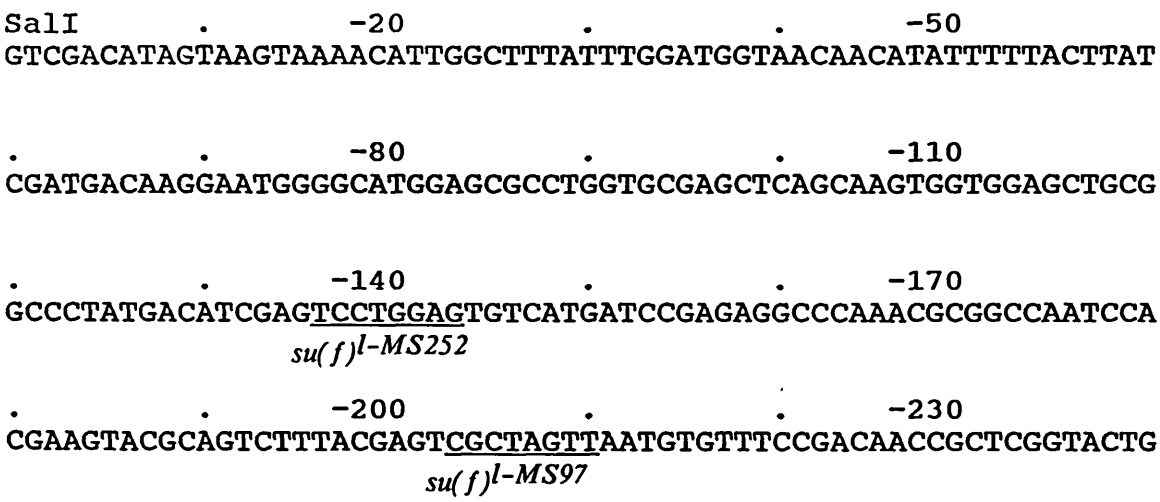
su(f)$^{l\text{-}MS252}$
CGACGGGACCACCTTATGTTATTTCATCATGctccaggactcgatgtcatagggccgcag

2880 2890 2900
CGACGGGACCACCTTATGTTATTTCATCATG pπ25.1

FIGURE 4.14 (cont.)

C.

D.



M13 vectors in order to sequence across the termini of each P element. For the left hand termini, the clones sequenced began at the left hand HindIII site of each P element, reading outwards. Other clones examined began at either the internal HindIII site or the nearby XhoI site, finishing at a genomic restriction enzyme site beyond the right hand end of the elements. Use was made of a special 17-mer sequencing primer homologous to P element sequences from position 2847 to position 2865 (Roiha *et al.*, 1988) in order to determine the nucleotide sequence of the right hand ends and DNA flanking each element. Using the normal M13 dideoxy sequencing primer, these clones also gave the nucleotide sequence reading across the position of the internal deletion in each of these defective P elements.

The M13 clones examined and the extent of nucleotide sequence obtained are shown diagrammatically in Figure 4.14a. Figure 4.14b gives the nucleotide sequence reading across the left hand and the right hand termini of each element. Figure 4.14c shows the sequence reading acrosss the internal deletion of each P element, compared to the sequence of the complete P element in pπ25.1 (O'Hare and Rubin, 1983).

The sequence data shows that the P elements from *su(f)*$^{l\text{-}MS97}$ and *su(f)*$^{l\text{-}MS252}$ are inserted in opposite orientations, confirming previous restriction enzyme site mapping data. The sequence from the clones reading across the termini of each P element was then compared to that of genomic DNA. Figure 4.14d shows the nucleotide sequence of wild type DNA from the SalI site at coordinate 0.0, reading to the left. On it are marked the insertion sites of the P elements of *su(f)*$^{l\text{-}MS97}$ and *su(f)*$^{l\text{-}MS252}$. Each element has created an 8bp duplication upon insertion, in keeping with the known properties of P elements (O'Hare and Rubin, 1983). In the P elements from both *su(f)*$^{l\text{-}MS97}$ and *su(f)*$^{l\text{-}MS252}$, the internal deletion is in exactly the same position, removing P elements sequences from nucleotide 894 to 2685 inclusive (Figure 4.14c). In addition, both P elements have an identical

single base polymorphism consisting of an A to T transversion at position 33, just to the right of the left hand 31bp terminal repeat (Figure 4.14b).

With respect to the polymorphism at position 33 and the position and extent of the internal deletion, the P element from *su(f)*$^{l-MS97}$ is identical in structure to that from *su(f)*$^{l-MS252}$. In fact, a P element of exactly this structure has previously been described: that associated with the mutation *white*$^{\#6}$ (O'Hare and Rubin, 1983). The property that *w*$^{\#6}$, *su(f)*$^{l-MS97}$ and *su(f)*$^{l-MS252}$ have in common is that all were isolated from dysgenic crosses where the mutagenising P strain was π_2. Indeed, two P elements of precisely this structure have been found in π_2; in the clones $\lambda\pi17$ and $\lambda\pi23$ (O'Hare and Rubin, 1983; S.McGrath and K.O'Hare, unpublished).

4.5.3 CLONING OF *su(f)* DNA FROM *su(f)*$^{l-S2}$

From DNA blotting data, *su(f)*$^{l-S2}$ is associated with an insertion approximately 4-5kb in size within the interval from coordinates -2.2 to -4.3. The restriction enzyme site map of the insertion resembles that of a Doc transposable element (Bender *et al.*, 1983). DNA from the *su(f)* region in *su(f)*$^{l-S2}$ was cloned in order to analyse more closely the molecular structure of this allele.

In cloning DNA covering the *su(f)*$^{l-S2}$ lesion, use was made of the fact that the insertion does not have a BamHI recognition site, and is thus carried on a BamHI fragment of ~14kb as opposed to the wild type sized fragment of 9.5kb. DNA from *su(f)*$^{l-S2}$*/FM6* heterozygous females was digested to completion with BamHI and the fragments separated on a low-gelling-temperature agarose gel. The gel portion containing fragments 12-16kb in size was excised and the DNA purified and ligated to λ EMBL4 digested with BamHI and SalI. Following in vitro packaging, the library was plated out using *E.coli* strain Q359. Filter lifts from the library were screened by hybridising to pBX64, and several positively hybridising phage clones were obtained. Following purification and large-scale liquid lysate

culture, DNA was prepared from two (λ S2.1 and λ S2.2). Restriction enzyme site mapping showed that both had identical inserts, differing from Canton S by the insertion of a 4.5kb region of DNA whose map matched that of a Doc element (Figure 4.15a). Furthermore, hybridisation of DNA blot filters to $pw^{1.8}$, a subclone carrying Doc sequences derived from the mutation w^1 (K.O'Hare, unpublished), confirmed that the 4.5kb insertion is a Doc element (data not shown).

4.5.4 SEQUENCE ANALYSIS OF THE DOC INSERTION IN *su(f)*$^{1-S2}$

To examine the insertion site of the *su(f)*$^{1-S2}$ Doc element, a number of M13 clones were subcloned from λ S2.1 in order to sequence across each end of the insertion. The positions of these are shown diagrammatically in Figure 4.15a, along with the M13 clones from pBX64 used to determine the wild type sequence of the insertion site.

The sequence data obtained is shown in Figure 4.15b. The Doc element does not have a terminal repeat, either direct or inverted. Instead, the most striking feature is a run 20 A residues at the left hand end of the element. The sequences of the termini of the inserted element match those of a previously published Doc element (O'Hare *et al.*; cited in Finnegan and Fawcett, 1986). The insertion of the Doc element in *su(f)*$^{1-S2}$ has produced a 13bp target site duplication. Analysis of several Doc element insertions at the sequence level has shown that all create a target site duplication, but that this can vary in size from 3 to 14bp (S.Lacey, A.Driver and K.O'Hare, unpublished).

The 13bp duplication created by the Doc element insertion in *su(f)*$^{1-S2}$ contains a recognition site for the restriction enzyme BglII. As there are no recognition sites for BglII within the Doc element (Figure 4.15a), the entire Doc element can be excised from λS2.1 with an absolute minimum of flanking DNA. This fragment has been subcloned into pBluescribe M13-, giving the clone pDoc (Figure 4.15a).

Figure 4.15 Nucleotide sequence analysis of the Doc insertion in *su(f)*$^{l-S2}$

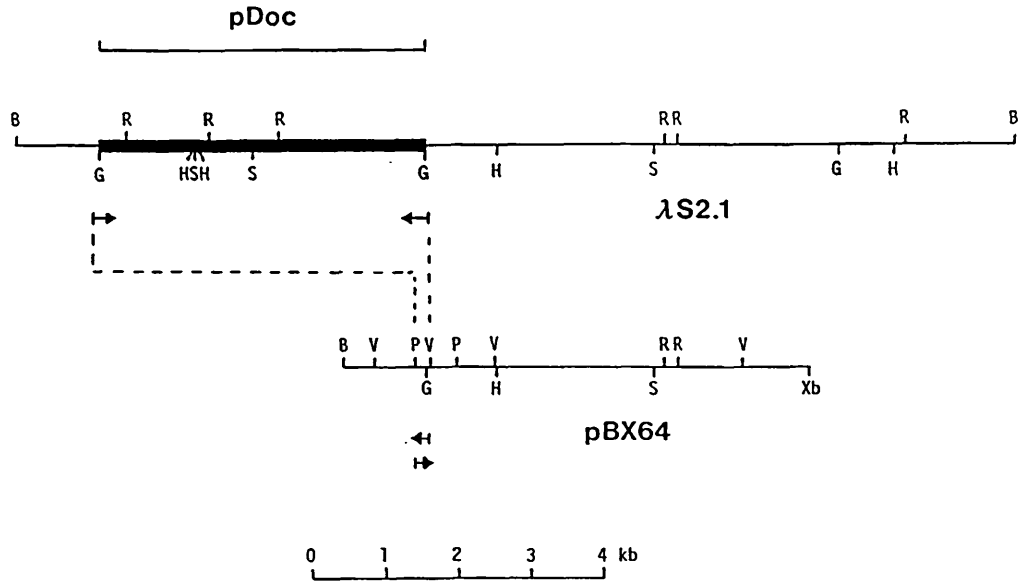
Part A shows the restriction enzyme site map of λ S2.1, compared to the map of Canton S (pBX64). Doc sequences are represented by thick black lines. Also shown are the extents of the nucleotide sequence readings across both ends of the Doc element and of the wild type insertion site from pBX64. The arrows indicate the direction of sequencing.

Restriction site key:	B : BamHI	S : SalI
	H : HindIII	V : PvuII
	R : EcoRI	G : BglII
	P : PstI	Xb : XbaI

Part B shows the nucleotide sequences of the right and left hand ends of the Doc insertion in comparison to Canton S sequence (right and left refer to the orientation of the Doc as drawn in part A). Sequences in lower case letters are from *su(f)*; those in upper case letters are Doc sequences. The underlined sequences are the 13bp target site duplication generated by the Doc insertion; as indicated, this duplicated region contains a BglII site.

FIGURE 4.15

A.



B.

su(f)$^{l-S2}$ Doc insertion : right hand end

```
          .   BglII  .
gaattgtcgcctaagatcttcgatctggccaagtcgg  : Canton S
|||||||||||||||||||||||||||||||||||    .    .    .    .
gaattgtcgcctaagatcttcgatcGAATTGATTCGGCATTCCACAGTCTTCGGGTGGAG

          .    .    .    .    .    .    .
ACGTGTTTCTTTCAAGCTACGAATAGCAAGTTCTAAAAACTACAACTAGTATAGTGAAAG
```

su(f)$^{l-S2}$ Doc insertion : left hand end

```
          .    .    .    .    .    .    .
AGTTAGATTAAGATTTGAACACATTATTGTTAGTCTCTTAACACAAAGGGAAGATTCAAT

          .    .    .         BglII    .    .
AAATAATAAAAAATTAAAAAAAAAAAAAAAAAAAAaagatcttcgatctggccaagtcgg
                                    |||||||||||||||||||||||||
          Canton S :  gaattgtcgcctaagatcttcgatctggccaagtcgg
```

4.6 THE LIMITS OF THE *su(f)* LOCUS

One way to define the extent of a gene within a cloned DNA region is to map the ends of breakpoints which define the gene genetically. For genetically well-studied euchromatic regions this approach can be reasonably straightforward; by these means the *white* gene was positioned within a 14kb interval (Levis *et al.*, 1982). Although the proximal X has been well characterised genetically and many deficiencies in this region exist, there are several problems in applying this approach to *su(f)*. Firstly, as the most proximal known gene, breakpoints proximal to *su(f)* might map anywhere in the heterochromatic half of the X chromosome. Secondly, although several breakpoints separate *su(f)* from the next distal complementation group, *sparse hairs*, it is possible that the cloned region does not extend far enough to reach them. Thirdly, in any approach using DNA blotting experiments, unique probes must be used. Given the distribution of repeated sequences in the cloned interval, it may be impossible to detect breakpoints within long stretches of repetitive DNA sequences.

The apparent submolarity of the bands of $su(f)^{l-L26}$ DNA in Figures 4.11 and 4.12 suggest that this allele may be a candidate for a deletion of the *su(f)* locus. The $su(f)^{l-L26}$ mutation is sph^{+}, so if it is a deletion, finding the breakpoint between *su(f)* and *sph* would be very useful. The allele *Df(1)su(f)GA130* is also sph^{+}, but whether or not it is truly a deficiency as its name implies is unclear; especially as the bands in the *Df(1)su(f)GA130/FM6* lanes in Figures 4.11 and 4.12 do not appear submolar.

A more generally applicable approach to determining the physical limits of a gene is that of P element mediated germ line transformation (Spradling and Rubin, 1982; Rubin and Spradling, 1982). This powerful technique provides a means for re-introducing cloned DNA fragments into the *D.melanogaster* genome, and thus can be used to establish whether a defined region of cloned DNA bearing a gene possesses

all those sequences which are necessary and sufficient to rescue mutant alleles of that gene (Spradling and Rubin, 1983; Goldberg *et al.*, 1983; Scholnick *et al.*, 1983).

Germ line transformation experiments were attempted in order to establish if a cloned wild type DNA fragment from the *su(f)* locus could, when re-established in the genome, rescue the full range of mutant phenotypes shown by alleles of *su(f)*. The interval chosen was the 7.8kb XbaI fragment from coordinates -5.7 to +2.15, which includes the region where almost all lesions associated with *su(f)* alleles are clustered, together with small amounts of the "1.5kb" repeats that flank it.

The 7.8 kb XbaI fragment from λCSP2 was cloned into two different vectors suitable for P element mediated transformation. Cloning into pDM23 (D.Mismer and G.M.Rubin, unpublished), a vector carrying the *D.melanogaster rosy* gene as a selectable marker, gave pry.X78. Cloning into phshyg B (S.Parkhurst, unpublished), a vector which uses a bacterial gene encoding resistance to the drug hygromycin B (Gritz and Davies, 1983) under the control of a *D.melanogaster* heat shock promoter, gave phshyg.X78. The helper P element used in each case was pπ25.7wc.

Four separate series of injections, twice each with pry.X78 and phshyg.X78, failed to give any transformed progeny. The survival rates of ry^{506} embryos injected with pry.X78 and pπ25.7wc helper were very poor: the two experiments gave 277 hatched embryos out of a total of 2562 injected (11%). These 277 hatched *ry* embryos resulted in a total of 32 fertile adults, none of which gave ry^{+} transformed progeny. At least some evidence that the mix of vector and helper DNAs was successfully injected and that they could successfully promote biological activity comes from the observation of G_0 *rosy* expression in both series of *ry* embryo injections. Both the ry^{506} and *bw:st* injection stocks were conclusively demonstrated to lack P element sequences by DNA blotting experiments (data not

shown).

The corresponding *bw;st* embryo survival rates in the two hygromycin B vector experiments were much higher; 858 hatched out of 2154 embryos injected (40%). These resulted in a total of 300 fertle adults. In the second series of injections using the hygromycin-based vector, some apparently transformed progeny were obtained. However, these turned out to be flies "escaping" the toxic effects of the selective food. These flies and their descendants consistently survived, albeit with low fecundity, on hygromycin B-containing food which was shown to be lethal to Canton S and *bw;st* flies. Why this should be the case remains unclear; perhaps some form of genetic background associated "hybrid vigour" raised their threshold of tolerance to the drug.

4.7 CONCLUDING REMARKS

This chapter has described the cloning of DNA sequences from the *su(f)* locus by means of the *su(f)* P element insertion allele $su(f)^{1-MS252}$. Using subclones of the region from the P element allele, some 32kb of DNA from Canton S flies was subsequently cloned. Analysis of the wild type interval using hybridisation techniques has demonstrated the presence of extensive amounts of repetitive DNA; however, there are regions of unique DNA. DNA blotting experiments have shown that the largest of these, at 7kb, contains not only the insertion site of the P element in $su(f)^{1-MS252}$, but also the interval where eight out of fourteen different lethal *su(f)* alleles tested have detectable chromosomal lesions.

This impressive clustering suggests firstly that each observed lesion is probably responsible for its particular mutant phenotype. This causal correlation has only been conclusively demonstrated for $su(f)^{1-MS252}$, with the use of revertants of the lethality. Secondly, it suggests that the region encompassing the sequences necessary for correct vital *su(f)* function may be physically limited to perhaps 5kb

or so.

Three *su(f)* alleles have been characterised at the sequence level. The two P element alleles, $su(f)^{l\text{-}MS97}$ and $su(f)^{l\text{-}MS252}$, were both shown to be insertions of a non-autonomous P element, inserted 75bp apart in opposite orientations. Both P elements share identical structures; a structure found in another *D.melanogaster* P element insertion allele, and at least twice in the strain π_2. DNA from the *su(f)* region of the allele $su(f)^{l\text{-}S2}$ was also cloned, and subsequently shown to have an insertion of a Doc transposable element.

None of the four viable *su(f)* alleles differed detectably from Canton S in the interval examined. Whether or not these mutants lie in another region where chromosomal lesions result in a visible *su(f)* phenotype remains unclear. However, three of these were induced with EMS, a mutagen which usually causes point mutations or small lesions difficult or impossible to detect in DNA blotting experiments. The same considerations apply to the allele $su(f)^{R\text{-}9\text{-}18}$, which shows intra-allelic complementation of lethality with several other *su(f)* alleles. None of the data so far obtained suggests that the *su(f)* gene is larger than the interval examined; one gene may have pleiotropic effects if mutated in different ways. Unfortunately, attempts at obtaining flies transformed with a 7.8kb fragment of DNA of the wild type *su(f)* region failed. Whether the 7.8kb XbaI fragment carries the whole of the *su(f)* locus, and whether this fragment has the ability to rescue all the mutant phenotypes shown by various alleles of *su(f)*, remains unknown.

Confirmation that the cloned DNA region does indeed come from the proximal region of the X chromosome has been provided by in situ hybridisation to polytene chromosomes. Using the probes pB1.L and pB1.4 (Figure 4.2), hybridisation was detected to well-stretched X chromosomes in a region between the most proximal

polytene bands visible (~20AB) and the chromocentre (M.Yamamoto and G.Miklos, unpublished). The fact that a signal was detectable suggests that *su(f)* is at least partially replicated in polytene chromosomes, unlike α-heterochromatin.

CHAPTER 5

ANALYSIS OF POLY-A^+ RNA TRANSCRIPTS AND cDNA CLONES FROM THE SUPPRESSOR-OF-FORKED LOCUS

5.1 INTRODUCTION

The previous chapter described the DNA structure of the *su(f)* chromosomal region. The experiments described in this chapter examine the *su(f)* transcription unit, focussing on the cloned interval where DNA lesions found in various *su(f)* mutants are clustered. In the first series of experiments, RNA transcripts from the *su(f)* region were studied by RNA blot analysis of poly-A^+ RNA from both wild-type and *su(f)* mutant flies. In further experiments, several cDNA clones representing transcripts from the *su(f)* region were isolated and analysed to further define the structure of the transcripts.

5.2 RNA BLOT ANALYSIS OF POLY-A+ RNA TRANSCRIPTS FROM *su(f)*

5.2.1 PRELIMINARY RNA BLOTTING EXPERIMENTS

The first series of RNA blot experiments consisted of a preliminary analysis of the number, extent and transcriptional orientation of poly-A^+ transcripts from the *su(f)* region. Figure 5.1 shows a RNA blot experiment using poly-A^+ RNA purified from first instar Oregon R larvae. Each filter was hybridised with a different single stranded radiolabeled probe made from M13 subclones of the genomic DNA intervals shown in Figure 5.1. Probes A1 and A2 cover the same interval from coordinates -2.2 to -4.3, but will recognise RNA transcripts from opposite strands of this interval. No transcripts were detected by probe A1, even though the Southern blot strip added to the hybridisation bag shows a discrete band, demonstrating the efficacy of the probe. Re-probing with a *Drosophila* ras64B clone (Mozer *et al.*, 1984) shows that this filter does indeed carry poly-A^+ RNA (Figure 5.1). On the other hand, probe A2 recognises a doublet of poly-A^+ RNA bands; a major band at 2.6kb and a lesser band at 2.9kb. Probes B and C also recognise these two transcripts. The results from probes A2, B and C reveal the presence of a doublet of transcripts whose 5' to 3' transcriptional orientation is from + to - with respect to the DNA map; i.e. their transcription proceeds leftwards. Together, probes A2, B and C cover the genomic interval from coordinates +2.2 to -

Figure 5.1 The unique central region of the cloned sequences is transcribed

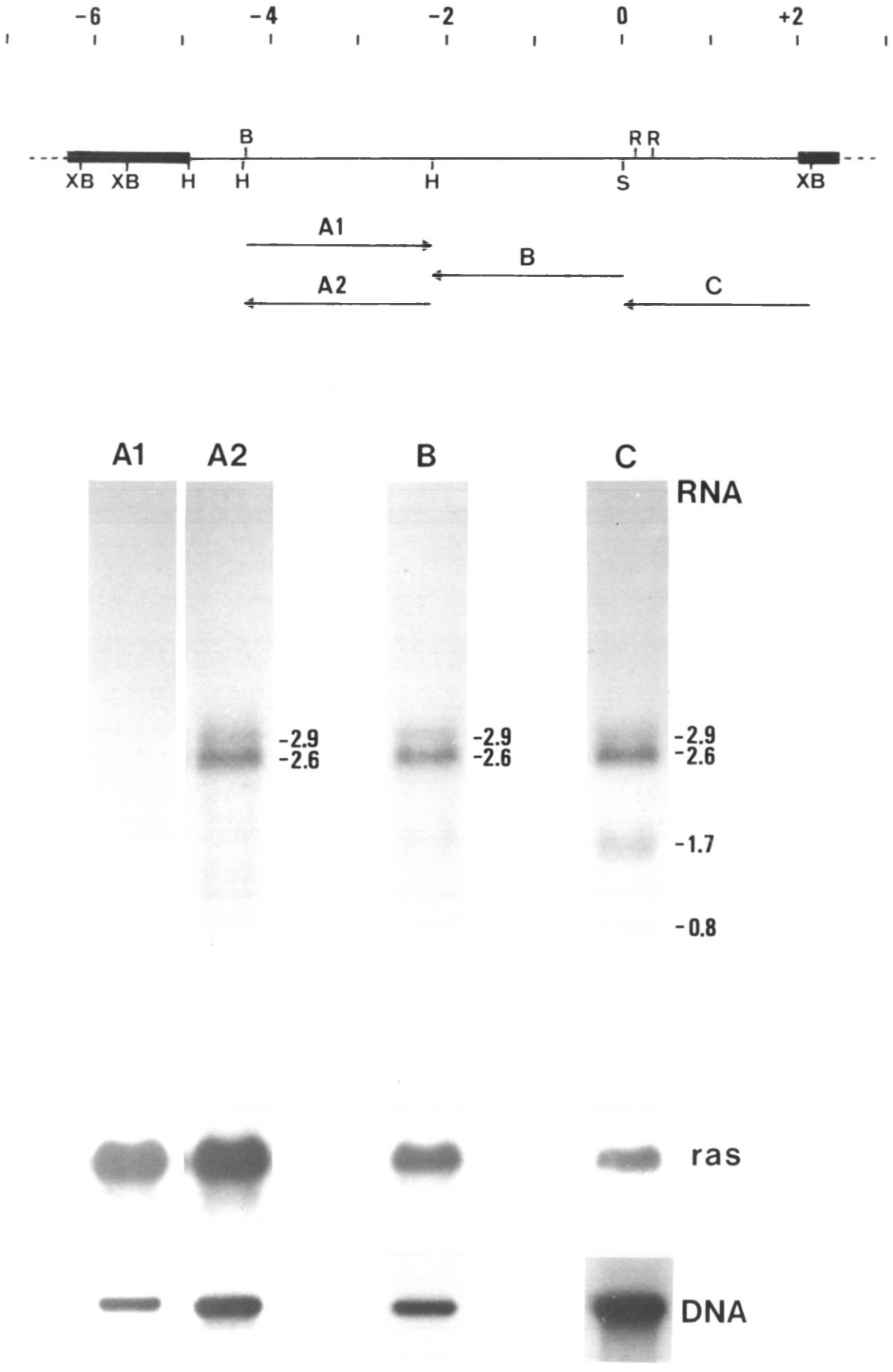
Shown are blots of Oregon R first instar poly-A^+ RNA samples probed with four single stranded DNA probes from the *su(f)* region. Approximately 5ug of RNA was loaded per lane. The extent of each probe is indicated on the restriction map at the top of the figure, where the arrows indicate the direction of transcription of any messages recognised by each probe. The thick black lines represent repetitive DNA sequences. The sizes of the RNA transcripts detected are marked in kb. The probes A2, B and C all detect RNA transcripts, but not probe A1; showing that the transcripts detected are all transcribed from right to left (5' to 3') with respect to the map.

Each RNA blot hybridisation also contained a DNA blot strip of Canton S DNA digested with XhoI and SalI to demonstrate the efficacy of the probe. At a later stage, the RNA filters were then hybridised to a *ras*64B probe to show that each contained poly-A^+ RNA sample.

Restriction enzyme key :

B : BamHI
R : EcoRI
S : SalI
H : HindIII
XB : XbaI

FIGURE 5.1


-6
-4
-2
0
+2
B
R R
XB XB H H H S XB
A1
B
A2
C
A1
A2
B
C
RNA
-2.9
-2.6
-2.9
-2.6
-2.9
-2.6
-1.7
-0.8
ras
DNA

4.3. In addition to the 2.6 and 2.9kb transcripts, probe C (coordinates +2.2 to 0.0) hybridises to two other transcripts of 1.7 and 0.8kb.

To delimit the extent of these transcripts, a similar series of RNA blots was carried out using additional probes. Equal amounts of Oregon R first instar poly-A^+ RNA were loaded onto the same gel and the resulting filter cut into strips, each of which was hybridised to a different probe. Figure 5.2 shows the results obtained and the genomic intervals covered by each probe. The 2.6 and 2.9kb doublet is recognized by probes E, F and G, but not by probes D and H. This narrows down the genomic interval from which the doublet is transcribed to between coordinates +1.25 to -4.3, unless either or both of probes D and H correspond to intron sequences. Probe G hybridises to the 1.7 and 0.8kb transcripts, unlike any other probe. The genomic interval covered by probe G is only 1.2kb, therefore the 1.7kb transcript must derive at least in part from sequences elsewhere.

In Figure 5.2, the 1.3kb transcript is detected only by probe F. Using a long autoradiographic exposure, this transcript is also detected by a probe covering the adjacent interval from coordinates 0.0 to -2.2 as used in Figure 5.4 (see next section). The results of these preliminary RNA blot analyses of transcripts from the *su(f)* region have been combined in Figure 5.3, showing diagramatically their size, extent and transcriptional orientation. The results from the previous chapter showed that DNA lesions associated with various lethal *su(f)* alleles were clustered in the interval from coordinates 0.0 to -4.3. The poly-A^+ transcripts from this interval are therefore the most likely candidates for *su(f)* mRNA species; that is the 2.9, 2.6 and 1.3kb messages.

5.2.2 THE DEVELOPMENTAL PROFILE OF TRANSCRIPTION FROM *su(f)*

Poly-A^+ RNA samples were prepared from various developmental stages of Oregon R : embryos, first instar, second instar, third instar, early pupae, late pupae and

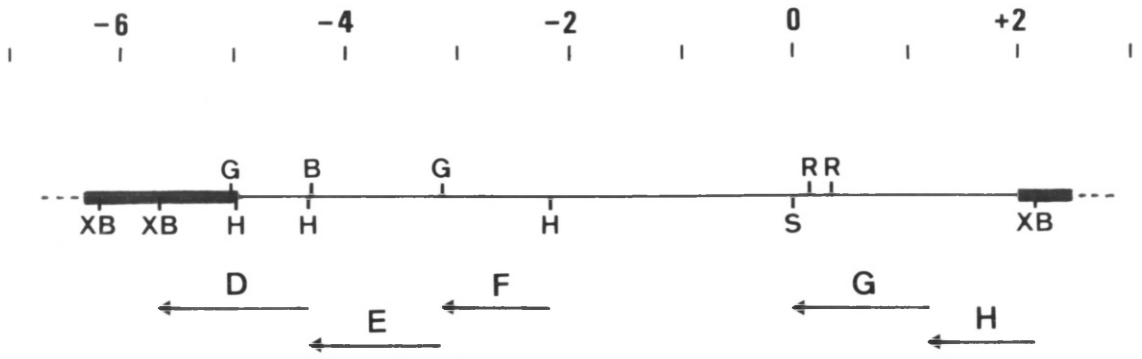
Figure 5.2 The extent of the *su(f)* transcription unit

Equal aliquots (5ug) of the same Oregon R first instar poly-A^+ RNA sample as used in Figure 5.1 were loaded on a gel and the resulting RNA blot filter cut into five strips. Each was hybridised to a different single stranded DNA probe from the *su(f)* region : D, E, F, G and H. The extent of each probe is indicated on the restriction map at the top of the figure, where the arrows indicate the direction of transcription of any messages recognised by each probe. The thick black lines represent repetitive DNA sequences.

The sizes of the RNA transcripts detected are shown in kb on the right of the RNA blot. Each RNA blot hybridisation also contained a DNA blot strip of Canton S DNA digested with XhoI and SalI to demonstrate the efficacy of the probe. The multiply hybridising bands and smear in the DNA blot strip using probe D is due to repetitive sequences contained in the probe.

Restriction enzyme key :	B : BamHI	H : HindIII
	R : EcoRI	XB : XbaI
	S : SalI	G : BglII

FIGURE 5.2



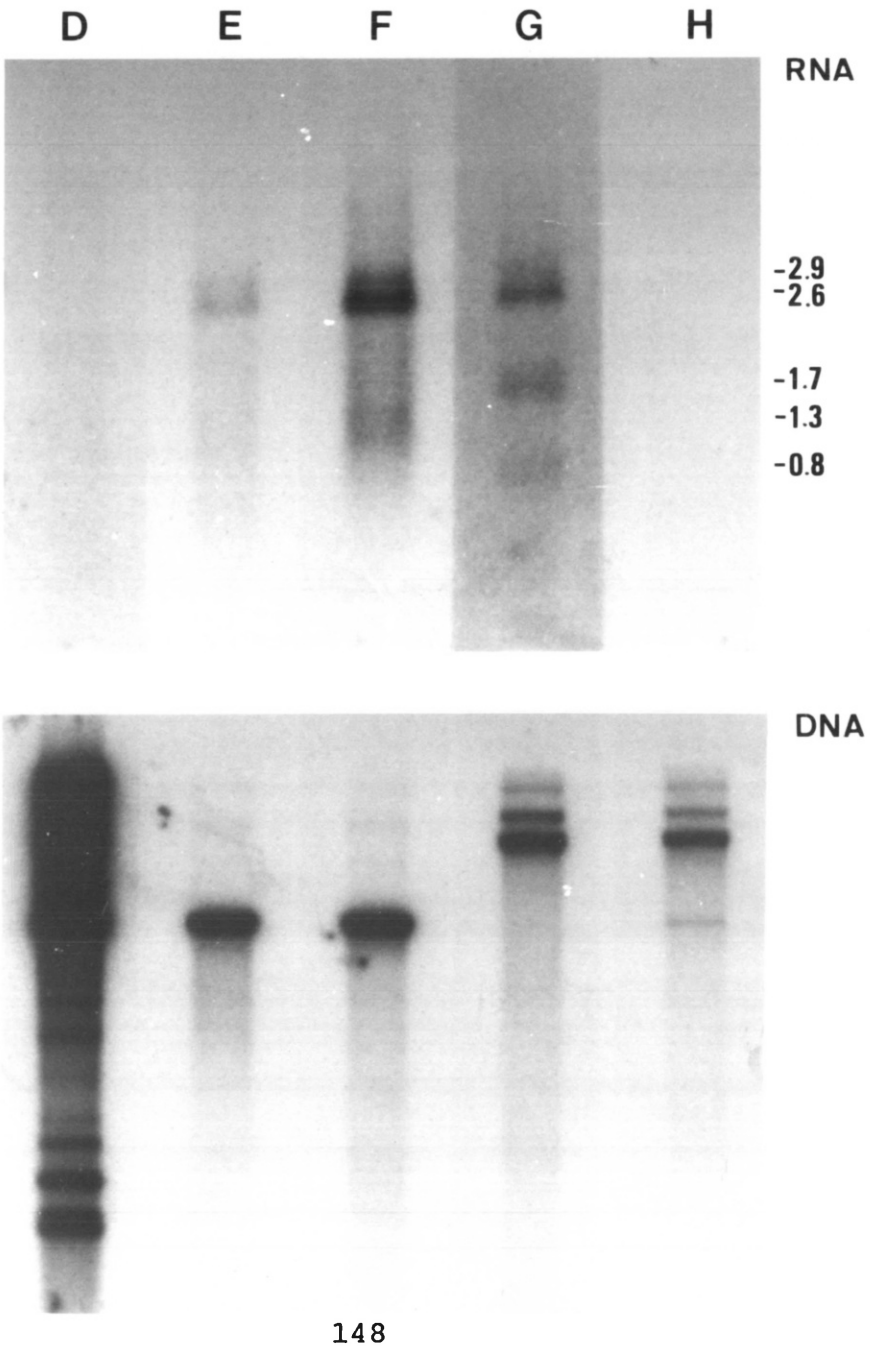


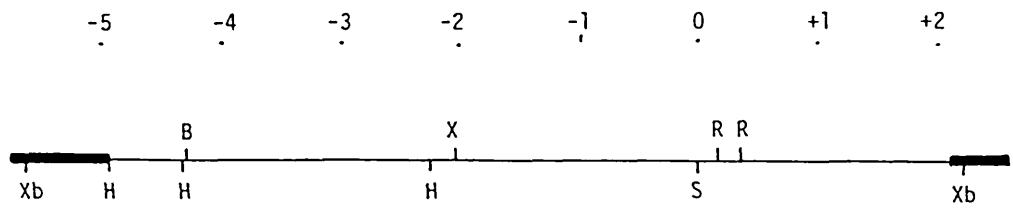
Figure 5.3 Diagram of poly-A^+ RNA transcripts in the *su(f)* region

Shown is a restriction enzyme site map of the *su(f)* region, indicating the size (in kb) and extent of the poly-A^+ RNA transcripts detected in the *su(f)* region. The arrows indicate the direction of transcription. The thick black lines represent repetitive DNA sequences.

Restriction enzyme key :

B : BamHI
R : EcoRI
S : SalI
H : HindIII
Xb : XbaI

FIGURE 5.3



← — — — — — — — — — — 2.9

← — — — — — — — — — — 2.6

← — — — — — 1.3

← — — 1.7

← — — 0.8

male and female adults. Approximately 10ug of each were loaded onto a gel and the resulting RNA blot filter probed with a single stranded DNA probe covering the interval from coordinate 0.0 to -2.2 (probe B in Figure 5.1). The result is shown in Figure 5.4. The filter was later re-probed with a *Drosophila* ras64B probe (Mozer *et al.*, 1984) to control for the amount of poly-A^+ RNA loaded in each lane. All three transcripts (2.9, 2.6 and 1.3kb) are present in all stages, but their abundances vary, suggesting that *su(f)* is subject to transcriptional regulation during development. In all stages, the most abundant transcript is the 2.6kb RNA. Relative to each other, the levels of the the 2.9 and 2.6kb RNAs remain constant; they are most abundant in the embryo, early pupal and adult female samples. The same is true for the 1.3kb RNA, but it differs in that, in relative terms, it is not as abundant as the larger doublet in the embryo sample as it is in early pupae. This suggests that the developmental regulation of levels of the 1.3kb RNA differs slightly from that of the two larger RNA species.

That transcripts from the *su(f)* region are present throughout development is not surprising in view of the range of phenotypes displayed by temperature sensitive *su(f)* alleles. Their ts lethal phases extend from late first instar/early second instar to just after pupariation (Dudick *et al.*, 1974; Russell, 1974, Jurgens and Gateff, 1979; Wilson, 1980). The ts period for suppression of f^1 lasts for 24h in early pupal life (Dudick *et al.*, 1974), during the period of bristle formation (Lees and Waddington, 1942). In addition, ts *su(f)* alleles show reversible female sterility when placed at the restrictive temperature. In all these periods of development the genetic data show that *su(f)* wild type function is required. The reversible sterility found in temperature sensitive *su(f)* alleles is particularly interesting in view of the greater transcript abundance in adult females by comparison to adult males, which may reflect *su(f)* transcription within the ovary. Ovarian follicle cell abnormalities and/or cell death have been seen in ts *su(f)* alleles at non-permissive temperatures (Dudick *et al.*, 1974; Wilson, 1980; Lineruth

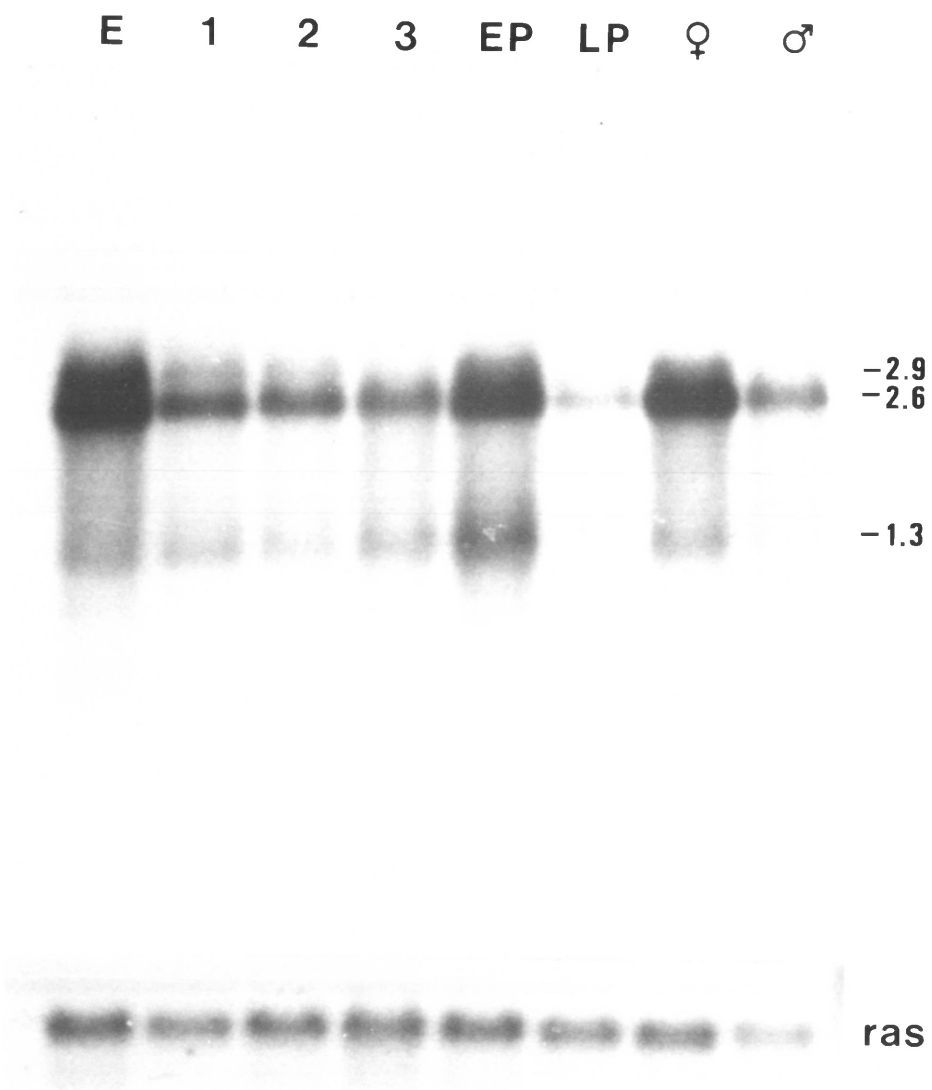
Figure 5.4 Developmental profile of poly-A^+ RNA transcripts in the *su(f)* region

Shown is a blot of Oregon R poly-A^+ RNA samples from different developmental stages probed with a single stranded DNA probe covering the interval between coordinate 0.0 and -2.2 (probe B in Figure 5.1). Approximately 10ug of poly-A^+ RNA was loaded in each lane. The sizes of the major transcripts detected are marked in kb to the right of the panel. The developmental stages used are as follows:

E : 0-16h embryos
1 : first instar larvae
2 : second instar larvae
3 : third instar larvae
EP : early pupae (7 days)
LP : late pupae (11 days)
♀ : adult females
♂ : adult males

Oregon R flies were grown at 25°C. Below the main panel shows the signal obtained on re-probing the RNA blot filter with *ras*64B to control for the relative amounts of poly-A^+ RNA loaded in each lane.

FIGURE 5.4



and Lambertsson, 1986), suggesting that in adult females *su(f)* transcription should at least be found in ovarian follicle cells. The abundance of the 2.6 and 2.9kb RNAs in embryos makes it tempting to speculate that these transcripts are maternally derived, although any maternal effects of *su(f)* mutations are impossible to determine since homozygous mutant germ line clones are lethal [N.Perrimon, pers.comm.(see Chapter 1)].

5.2.3 ANALYSIS OF POLY-A+ RNA TRANSCRIPTS FROM *su(f)* MUTANTS

RNA samples from *su(f)* mutant flies were examined in an attempt to correlate their mutant phenotypes with quantitative and/or qualitative alterations in transcripts from the *su(f)* region. Four alleles were studied: $su(f)^{1}$, $su(f)^{l\text{-}ts76a}$, $su(f)^{l\text{-}S2}$ and $su(f)^{l\text{-}MS252}$. The first two of these show no difference from wild type at the DNA level in the *su(f)* region; and both are viable as homozygotes. The alleles $su(f)^{l\text{-}S2}$ and $su(f)^{l\text{-}MS252}$ are both associated with transposable element insertions; the former has a Doc element inserted at coordinate -3.3 and the latter has a P element inserted at coordinate -0.15. For the two viable alleles, RNA was prepared from homozygous adult males and females grown at 25°C; but for the two lethal alleles, RNA was prepared from adult female $su(f)^{lethal}$ / *FM6* heterozygotes, again grown at 25°C. Poly-A^+ RNA samples from each of the four were loaded onto a gel, along with a sample of adult female Oregon R poly-A^+ RNA. The resulting RNA blot filter was hybridised to a single-stranded DNA probe covering the interval between coordinates 0.0 and -2.2 (probe B in Figure 5.1).

The results of the RNA blot are shown in Figure 5.5, along with the results obtained from re-probing with a *D.melanogaster* ras64B probe (Mozer *et al.*, 1984) to control for the amounts of poly-A^+ RNA loaded in each sample. In comparison to the adult female Oregon R sample, a number of the mutant samples show new transcripts.

In $su(f)^{1}$, both the 2.6 and 2.9kb wild type species are present, but the 1.3kb

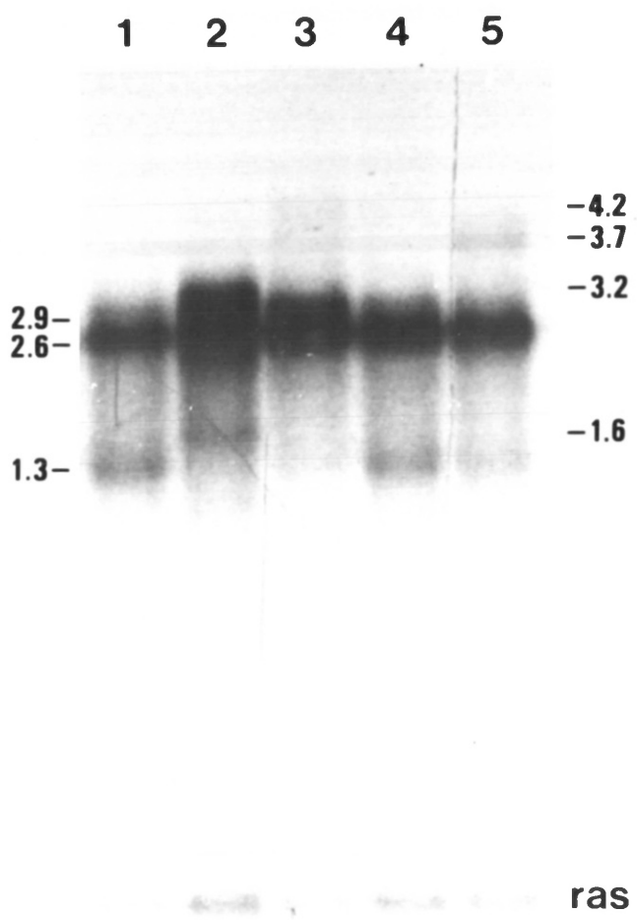
Figure 5.5 Analysis of poly-A^+ RNA in *su(f)* mutants

Shown is a blot of poly-A^+ RNA samples from *su(f)* mutants probed with a single stranded DNA probe covering the interval between coordinates 0.0 to -2.2 (probe B in Figure 5.1). Approximately 5ug of poly-A^+ RNA was loaded per lane. The sizes of the three wild type transcripts are marked on the left of the panel; on the right of the panel are marked the sizes of new transcript bands.

Lane 1 : Oregon R adult females
Lane 2 : $su(f)^{1}$ adult males and females
Lane 3 : $su(f)^{l\text{-}ts76a}$ adult males and females
Lane 4 : $su(f)^{l\text{-}S2}$ / *FM6* adult females
Lane 5 : $su(f)^{l\text{-}MS252\text{-}1}$ / *FM6* adult females

All flies were grown at 25°C. Below the main panel is shown the results obtained on re-probing the filter with *ras*64B to control for the amount of poly-A^+ RNA loaded in each lane.

FIGURE 5.5



transcript is not detected. There are two new bands of 3.2kb and 1.6kb. Since the 1.3kb and the two larger RNA species all derive from the same transcription unit, these novel transcripts might result from a mutation affecting overall mRNA length, such as a point mutation in a splice site which acted to decrease splicing efficiency of a particular intron. This would result in larger-than-wild type transcripts such as the two novel bands: i.e. the 1.3kb RNA becomes 1.6kb in size, and the 2.9kb transcript becomes 3.2kb in size. If the same occurred to the 2.6kb RNA, any new band at 2.9kb would be obscured.

In the $su(f)^{l\text{-}ts76a}$ sample the abundance of the 1.3kb RNA is decreased by comparison to wild type. There also appears a barely detectable transcript of 4.2kb in length. In the absence of a detectable DNA lesion in $su(f)^{l\text{-}ts76a}$, the origin of the larger transcript remains mysterious. At 25°C $su(f)^{l\text{-}ts76a}$ is fully viable, but does suppress f^{1}. It is possible that adult flies aged for several days at 29°C might show a more abnormal pattern of transcription.

In the poly-A^+ samples from the two lethal *su(f)* alleles, interpretation of the results is complicated by the fact that the *su(f)* copy on the balancer chromosome should give a wild type transcription pattern. In these samples it is therefore probably only worthwhile looking for novel transcripts, and not alterations in transcript abundance. The $su(f)^{l\text{-}S2}$ sample gives a pattern indistinguishable from wild type. The Doc element inserted in the transcription unit that gives rise to the 2.6 and 2.9kb RNAs is oriented such that its poly-A^+ addition site is in the appropriate strand for terminating transcription prematurely. This would result in a chimaeric transcript lacking an unknown proportion of its wild type 3' end, but including sequences from the 4.5kb Doc element which would make it larger than the 2.6 or 2.9kb RNAs. No such transcript is detected. However, in the $su(f)^{l\text{-}MS252}$ sample there is in addition to the three wild type sized transcripts a novel band of 3.7kb. The P element insertion in this allele is oriented in the opposite

transcriptional orientation from the transcription unit, so the termination signals in the P elements are on the non-coding strand. The 3.7kb RNA in *su(f)*$^{l-MS252}$ probably therefore results from transcription reading through the 1.1kb P element to give an transcript larger than wild type by the size of the element.

The differences observed in the pattern of transcription from this genomic region in the various *su(f)* alleles provides almost conclusive evidence that this transcription unit is that of *su(f)*. Which of the 1.3, 2.6 and 2.9kb RNAs is the *su(f)*$^{+}$ transcript, or does more than one contribute to *su(f)* function? Consider the Doc element insertion in *su(f)*$^{l-S2}$, which only interrupts the two larger transcripts. If the Doc element is the lesion responsible for the lethal phenotype of *su(f)*$^{l-S2}$, then at least as far as the lethal function of *su(f)* is concerned, it is more likely that either or both of the 2.6 and 2.9kb transcripts are *su(f)*$^{+}$ RNAs. There remains the possiblity that the Doc insertion in *su(f)*$^{l-S2}$ might affect the termination processes involved in producing the 1.3kb RNA, in which case this transcript too may have a role in *su(f)* function.

5.3 ISOLATION AND PRELIMINARY ANALYSIS OF cDNA CLONES FROM THE *su(f)* LOCUS

To investigate in more detail the structure of the transcripts from the *su(f)* region, several cDNA clones representing these transcripts were isolated and analysed. The clones were obtained by screening a cDNA library made from 1st and 2nd instar poly-A^{+} RNA (Poole *et al.*, 1985). This cDNA library is an amplified library, and has a representation of $2x10^{5}$ clones (L.Kauvar, B.Drees, S.Poole and T.Kornberg, pers.comm.). A total of 210,000 plaques were plated out using *E.coli hflA* and the resulting filters hybridised to nick-translated pBX64. The library gave 24 positively hybridising clones, varying in signal strength, which were designated λcK1 to λcK24. After three rounds of plaque purification, a pure sample of each clone was obtained. Filter hybridisation experiments with nick-

translated pBR322 DNA revealed that eight of the 24 were contaminant clones carrying plasmid DNA sequences: cK2, 7, 9, 11, 15, 17, 20 and 21 (data not shown). These were discarded.

The remaining sixteen clones were grown up in large scale liquid cultures and DNA prepared from the purified phage. The size of the inserts in each clone were determined by agarose gel electrophoresis of the λ DNA samples digested with EcoRI, and are as follows:

cK1	1.35kb	cK13	1.25kb, <0.2kb
cK3	1.45kb	cK14	1.9kb
cK4	1.9kb	cK16	1.35kb
cK5	1.7kb	cK18	0.8kb
cK6	1.9kb	cK19	0.45kb
cK8	1.35kb	cK22	2.5kb, 0.7kb
cK10	0.55kb	cK23	1.45kb
cK12	0.9kb	cK24	0.9kb

A number of the EcoRI insert fragments co-migrate with one another in agarose gels (data not shown), and may therefore be multiple re-isolates of the same recombinant from the library. This was confirmed by subcloning the EcoRI fragments of the cDNA clones into an M13 vector and performing a single deoxyTTP/dideoxyTTP sequencing (T-track) reaction on the clones obtained (data not shown). On the basis of identical insert size and identical T-tracks of insert termini, the sixteen clones were arranged into ten cDNA clones of unique structure:

cK1, 8, 16	1.35kb	cK12, 24	0.9kb
cK3, 23	1.45kb	cK13	1.25kb, <0.2kb
cK4, 6, 14	1.9kb	cK18	0.8kb
cK5	1.7kb	cK19	0.45kb
cK10	0.55kb	cK22	2.5kb, 0.7kb

In a crude-scale mapping experiment, DNA samples of each of the ten different λ cDNA clones were nick-translated and used individually as hybridisation probes on

nitrocellulose strips to which had been fixed samples of genomic M13 clones. These M13 clones carried various intervals of genomic DNA subcloned from pBX64 and pCS4. These are shown in Figure 5.6, which also shows the results of the dot-blot analysis. The cDNA clones fall into two groups on the basis of which genomic intervals they hybridise to. Only clones of the first group (cK3, cK5 and cK10) hybridise to sequences between coordinates +1.0 and +2.15. The clones from the second group (cK1, cK4, cK12, cK13, cK18, cK19 and cK22) apparently arise from a transcription unit lying between coordinates +0.35 and -4.3. A number of the clones appear to hybridise to genomic interval J (Figure 5.6), coordinates +0.35 to +1.0. However, subsequent sequence analysis of these cDNA clones showed that they do not carry DNA from this interval (see later in this Chapter). The M13 DNA sample of interval J may be contaminated with another clone, possibly that of interval H (Figure 5.6).

5.4 NUCLEOTIDE SEQUENCE ANALYSIS OF cK22

5.4.1 DETERMINATION OF THE NUCLEOTIDE SEQUENCE OF cK22

The second group of cDNA clones hybridises to the genomic interval between coordinates +0.35 and -4.3. It is this region in which lesions associated with mutant *su(f)* alleles (except $su(f)^{l-3DES}$) are clustered. Furthermore, from this interval arise the three transcripts of 1.3, 2.6 and 2.9kb some or all of which are qualitatively altered in *su(f)* mutant flies. This group of cDNA clones thus appears to derive from the *su(f)* transcription unit.

The largest cDNA clone of this group is cK22. It has two EcoRI insert fragments of 0.7kb and 2.5kb, but only the larger fragment corresponds to a transcript from the *su(f)* region. The 0.7kb EcoRI fragment from cK22 does not hybridise to DNA from the 32kb cloned interval around the *su(f)* locus (data not shown). Presumably the smaller EcoRI fragment derives from an RNA transcript unlinked to *su(f)*, and that the presence of both fragments in one λ recombinant is a cloning artefact.

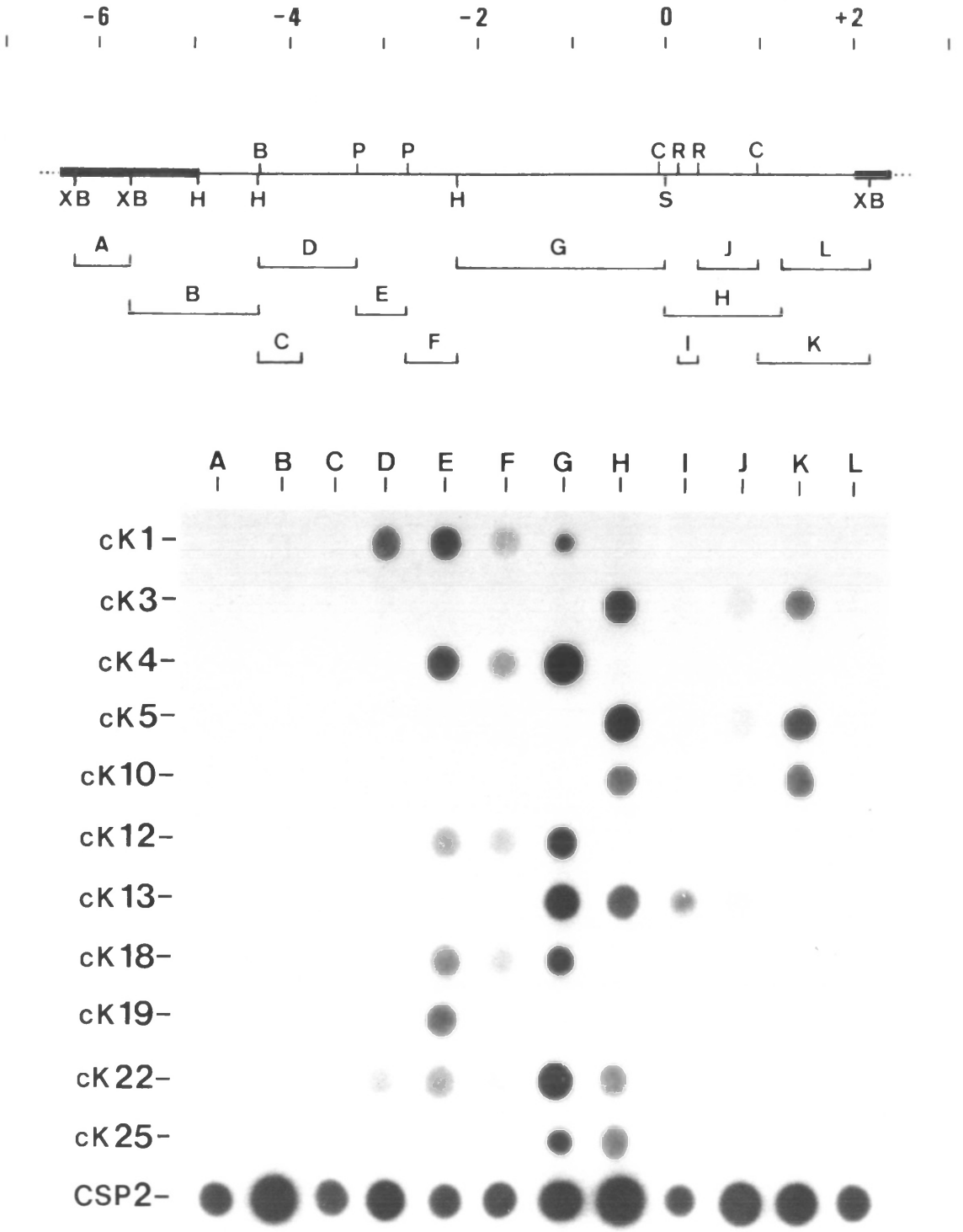
Figure 5.6 Location of the extent of cDNA clones in the genomic region

Identical filter strips bearing M13 genomic subclones from the *su(f)* region were hybridised to nick-translated cDNA λ clones. The limits of the M13 sublcones A to L are indicated on the restriction map above the panel. The thick black lines represent repetitive DNA sequences. As a control, one strip was hybridised to the Canton S genomic clone λCSP.2, which covers the entire interval.

Restriction enzyme key :

B : BamHI
R : EcoRI
S : SalI
C : ClaI
H : HindIII
XB : XbaI
P : PstI

FIGURE 5.6


-6
-4
-2
0
+2
B
P
P
C R R
C
XB
XB
H
H
H
S
XB
A
D
G
J
L
B
E
H
C
F
I
K
A
B
C
D
E
F
G
H
I
J
K
L
cK1-
cK3-
cK4-
cK5-
cK10-
cK12-
cK13-
cK18-
cK19-
cK22-
cK25-
CSP2-

Henceforth, the name cK22 will apply only to the 2.5kb insert fragment from this clone. The 2.5kb EcoRI fragment from cK22 was purified from an LGT agarose gel and aliquots of the purified DNA digested with four different restriction enzymes with 4bp recognition sites: Sau3aI, MspI, TaqI and HaeIII. Fragments from these digests were cloned into appropriately digested M13 vectors and recombinants screened by performing a single deoxyTTP/dideoxyTTP sequencing reaction (T-track) on each clone. In addition, a number of subfragments of the 2.5kb EcoRI fragment were generated by digesting with restriction enzymes with 6bp recognition sites; these were purified from LGT agarose gels and cloned into appropriately digested M13 vectors. Selected clones from the Sau3aI, MspI, TaqI and HaeIII cloning experiments and the collection of fragments generated by digesting with enzymes with 6bp recognition sites were sequenced using the Sanger dideoxy chain-terminating method. The nucleotide sequences obtained were assembled using the Microgenie package of computer programs (Beckman) to generate a contig of the complete sequence of cK22. In the sequence, the identity of each nucleotide was determined from at least one gel reading on each strand.

The complete nucleotide sequence of the 2.5kb cK22 cDNA is shown in Figure 5.7, oriented 5' to 3' in the direction of transcription. At the 5' end of the sequence is a run of 16 C residues (position 7 to 22), presumably resulting from the method used to make the cDNA library. After the first-strand synthesis step, the first cDNA strand was tailed with dGTP and the synthesis of the second strand primed with an oligo-C primer (Poole *et al.*, 1985). At the 3' end of the sequence is a tract of 53 A residues (position 2418-2470), preceded by an AATAAA polyadenylation signal (position 2399 to 2404) (Proudfoot and Brownlee, 1976). A search of the EMBL and GenBank databases with the nucleotide sequence of cK22 failed to yield any matches of obviously significant homology. The sequence of cK22 also contains the target sites of the P element insertions in $su(f)^{l-MS97}$ and $su(f)^{l-MS252}$, and of the Doc insertion in $su(f)^{l-S2}$. These are shown underlined in Figure 5.7.

Figure 5.7 Nucleotide sequence of the 2.5kb insert of cK22

Shown is the nucleotide sequence of the 2.5kb insert of cK22, orientated 5' to 3' in the direction of transcription of the message it represents. The sequences in lower case at the 5' and 3' ends are sequences not found in the corresponding genomic positions; i.e. the 3' poly-C leader resulting from the cDNA cloning method and the 5' poly-A tail. Selected restriction enzyme recognition sites sre indicated above the sequence. The amino acid sequence of the putative product of the long open reading frame is shown beneath the nucleotide sequence; it begins at the most 5' ATG codon and ends at a TAA termination codon (marked ***). The AATAAA polyadenylation signal just 5' to the start of the poly-A tail is boxed. The positions of the eight introns are indicated by the vertical arrows. Also shown are the target sites of the P element insertions of *su(f)*$^{l-MS97}$ and *su(f)*$^{l-MS252}$, as well as the Doc insertion found in *su(f)*$^{l-S2}$; the nucleotides underlined are those which form the respective target site duplications.

FIGURE 5.7

cK22

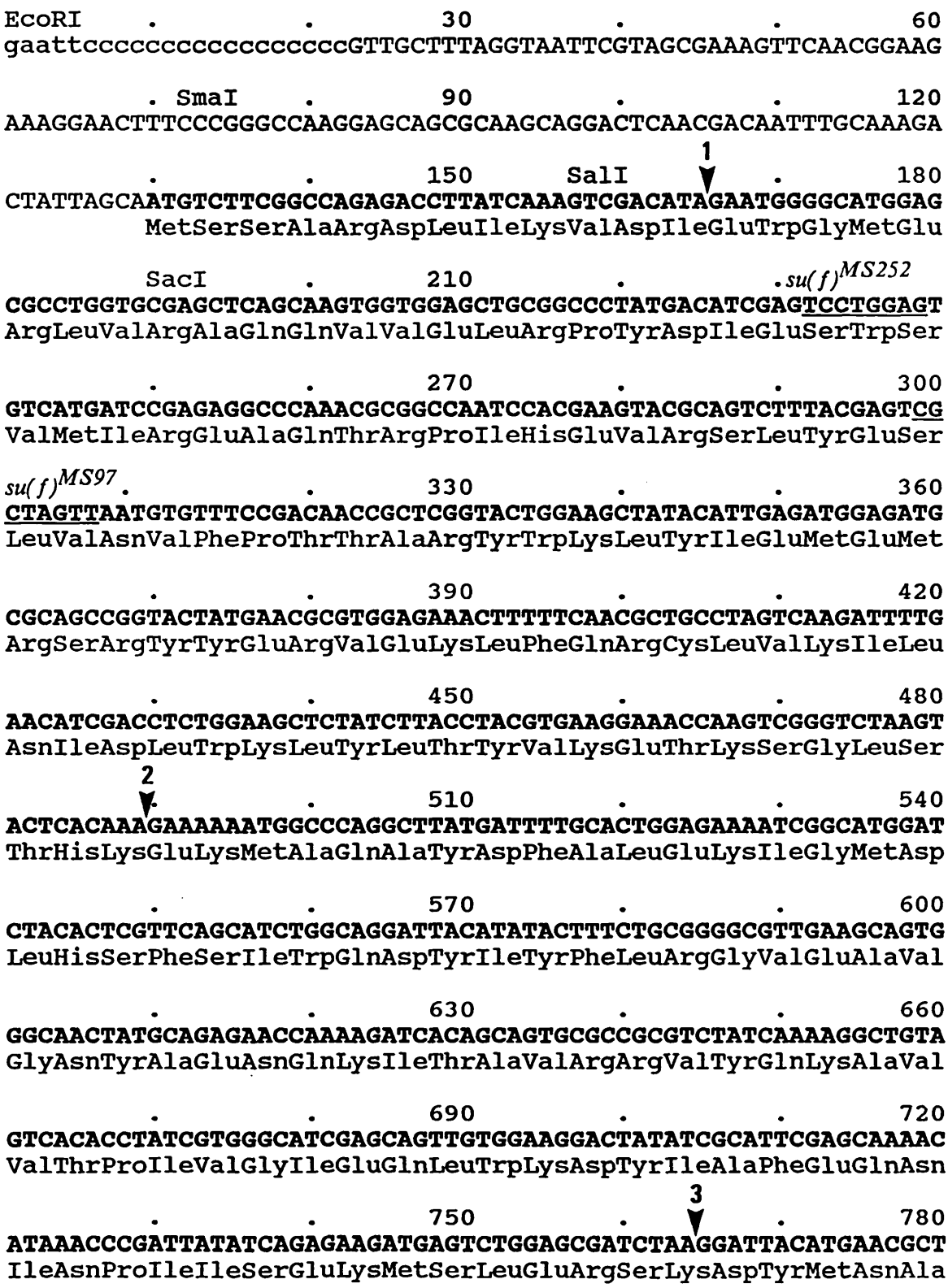


FIGURE 5.7 (cont.)

```
          .         .       810         .         .       840
CGTCGTGTTGCCAAAGAGTTGGAATACCACACAAAGGGTCTCAATCGCAATCTGCCAGCC
ArgArgValAlaLysGluLeuGluTyrHisThrLysGlyLeuAsnArgAsnLeuProAla

          .         .       870         .         .       900
GTCCCACCCACACTCACAAAAGAGGAGGTGAAGCAAGTGGAGCTGTGGAAGCGCTTTATC
ValProProThrLeuThrLysGluGluValLysGlnValGluLeuTrpLysArgPheIle

          .         .       930         .         .       960
ACGTATGAGAAGTCTAATCCCTTGCGCACTGAAGATACGGCCCTAGTAACCCGTCGTGTC
ThrTyrGluLysSerAsnProLeuArgThrGluAspThrAlaLeuValThrArgArgVal

          .         .       990         .         .      1020
ATGTTCGCCACGGAGCAATGTTTGCTGGTGCTCACCCACCACCCGGCAGTGTGGCACCAG
MetPheAlaThrGluGlnCysLeuLeuValLeuThrHisHisProAlaValTrpHisGln
                                                4
          .         .      1050         .         .      1080
GCCTCCCAATTCCTGGACACCAGTGCTCGCGTGCTCACCGAAAAAGGCGACGTCCAGGCG
AlaSerGlnPheLeuAspThrSerAlaArgValLeuThrGluLysGlyAspValGlnAla

          .         .      1110         .         .      1140
GCCAAGATTTTTGCGGACGAGTGTGCTAATATACTGGAACGGTCTATAAACGGCGTGCTA
AlaLysIlePheAlaAspGluCysAlaAsnIleLeuGluArgSerIleAsnGlyValLeu

          .         .      1170         .         .      1200
AACCGAAATGCGTTGCTATATTTTGCCTATGCGGACTTTGAAGAAGGACGGCTCAAGTAC
AsnArgAsnAlaLeuLeuTyrPheAlaTyrAlaAspPheGluGluGlyArgLeuLysTyr

          .         .      1230         .         .      1260
GAGAAGGTGCACACTATGTACAACAAGCTGTTACAGTTGCCAGACATTGATCCCACATTG
GluLysValHisThrMetTyrAsnLysLeuLeuGlnLeuProAspIleAspProThrLeu
5
          .         .      1290         .         .      1320
GTCTATGTTCAGTACATGAAGTTCGCCCGTCGTGCTGAGGGGATCAAGTCGGCCCGTAGC
ValTyrValGlnTyrMetLysPheAlaArgArgAlaGluGlyIleLysSerAlaArgSer

          .         .      1350         .         .      1380
ATATTTAAGAAAGCTCGGGAGGACGTAAGGTCGCGCTATCACATTTTCGTGGCCGCTGCT
IlePheLysLysAlaArgGluAspValArgSerArgTyrHisIlePheValAlaAlaAla

          .         .      1410         .         .      1440
TTAATGGAGTATTATTGCTCCAAGGACAAAGAAATTGCGTTCCGCATCTTCGAGCTGGGA
LeuMetGluTyrTyrCysSerLysAspLysGluIleAlaPheArgIlePheGluLeuGly

          .         .      1470         .         .      1500
TTGAAGCGTTTTGGTGGCAGTCCAGAGTACGTGATGTGCTACATTGACTACCTGTCCCAT
LeuLysArgPheGlyGlySerProGluTyrValMetCysTyrIleAspTyrLeuSerHis

          .         .      1530         .         .      1560
CTAAACGAGGACAATAATACGCGCGTCCTGTTTGAACGGGTACTTAGTTCAGGTGGTTTG
LeuAsnGluAspAsnAsnThrArgValLeuPheGluArgValLeuSerSerGlyGlyLeu

          .         .      1590      XhoI       .      1620
TCGCCGCACAAGAGTGTTGAGGTCTGGAATCGCTTCCTCGAGTTCGAATCTAATATTGGG
SerProHisLysSerValGluValTrpAsnArgPheLeuGluPheGluSerAsnIleGly
```

FIGURE 5.7 (cont.)

```
                                                            6
     .           .      1650           .           .        ▼ 1680
GATCTCTCCAGCATCGTTAAGGTCGAACGTCGCCGTAGTGCCGTTTTCGAAAACCTTAAA
AspLeuSerSerIleValLysValGluArgArgArgSerAlaValPheGluAsnLeuLys

     .           .      PvuII          .           .        1740
GAGTATGAGGGCAAGGAGACCGCCCAGCTGGTTGACCGTTACAAATTTCTGGACCTTTAC
GluTyrGluGlyLysGluThrAlaGlnLeuValAspArgTyrLysPheLeuAspLeuTyr
                                            7
     .       SacI.      1770           . ▼         .        1800
CCCTGCACCAGCACCGAGCTCAAGTCCATAGGATATGCTGAGAATGTGGGCATTATACTG
ProCysThrSerThrGluLeuLysSerIleGlyTyrAlaGluAsnValGlyIleIleLeu

     .         SacI     1830           .           .        1860
AACAAGGTGGGTGGTGGAGCTCAAAGCCAGAACACCGGAGAGGTCGAAACAGATAGCGAG
AsnLysValGlyGlyGlyAlaGlnSerGlnAsnThrGlyGluValGluThrAspSerGlu

     .           .      1890           .           .        1920
GCAACGCCACCATTGCCACGTCCGGATTTCTCCCAGATGATTCCATTCAAGCCGCGTCCA
AlaThrProProLeuProArgProAspPheSerGlnMetIleProPheLysProArgPro

     .           .      1950           .           .        1980
TGTGCTCATCCTGGTGCCCATCCACTTGCAGGTGGTGTATTTCCGCAGCCGCCAGCTCTG
CysAlaHisProGlyAlaHisProLeuAlaGlyGlyValPheProGlnProProAlaLeu

     .           .      2010           .           .        2040
GCCGCTTTGTGCGCCACCCTGCCTCCACCGAACTCCTTCCGTGGACCGTTCGTCAGCGTA
AlaAlaLeuCysAlaThrLeuProProProAsnSerPheArgGlyProPheValSerVal
                                            8
     .           :      2070           . ▼         .        2100
GAGCTACTATTCGACATTTTCATGCGTCTCAATCTTCCGGACTCTGCACCGCAACCGAAT
GluLeuLeuPheAspIlePheMetArgLeuAsnLeuProAspSerAlaProGlnProAsn

     .           .  BglII su(f)^l-S2   .           .        2160
GGAGACAACGAATTGTCGCCTAAGATCTTCGATCTGGCCAAGTCGGTTCACTGGATCGTG
GlyAspAsnGluLeuSerProLysIlePheAspLeuAlaLysSerValHisTrpIleVal

     .           .      2190           .           .        2220
GATACAAGTACGTATACGGGGGTGCAGCACAGTGTTACGGCCGTTCCACCGCGCCGGCGA
AspThrSerThrTyrThrGlyValGlnHisSerValThrAlaValProProArgArgArg

     .           .      2250       PstI            .        2280
CGACTGCTGCCTGGGGGCGATGATAGTGACGACGAACTGCAGACCGCTGTGCCGCCCTCC
ArgLeuLeuProGlyGlyAspAspSerAspAspGluLeuGlnThrAlaValProProSer

     .           .      2310           .           .        2340
CACGATATATATCGCCTGCGTCAGCTAAAGCGTTTCGCCAAGTCTAATTAAGAAACGTTA
HisAspIleTyrArgLeuArgGlnLeuLysArgPheAlaLysSerAsn***

     .           .      2370           .           .        2400
AAAAAAAACCGATTTTTACTTTTTTCTTTCAATTAATTTACAAGAAACAAGAAATCCCAA

     .           .      2430           .           .        2460
TAAAGCAAGCCATATTAaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaagaaaaa

     . EcoRI
aaaaaaaaaaggaattc
```

5.4.2 THE INTRON/EXON STRUCTURE OF THE cK22 TRANSCRIPT

Genomic fragments from the *su(f)* region were subcloned into M13 vectors, and nucleotide sequence readings from these clones used to assemble two large contigs of genomic sequence. Two of the clones sequenced read from the appropriate termini of the inserts from λCSP4 and λCSP5. The contigs are separated by a gap of approximately 300bp where no convenient genomic restriction sites exist. The positions of the clones sequenced and the extent of the sequence data obtained are shown in Figure 5.8. Comparison of genomic sequence to the sequence of the 2.5kb cK22 cDNA revealed multiple discontinuities representing introns. There are eight introns in the cK22 transcript; their positions are indicated on the sequence of cK22 in Figure 5.7. The intron/exon structure of the cK22 cDNA is shown in diagrammatic form in Figure 5.8; the most 5' sequence represented in cK22 starts at genomic coordinate +0.13 and the most 3' ends at coordinate -3.4. The sequencing of genomic M13 subclones gave the sequence of seven of the splice donors and eight of the splice acceptors. The splice junction sequences are shown in Table 5.1; these match well the consensus splice donor and acceptor sequences derived from a large number of *D.melanogaster* genes (Mount, 1982) (Table 5.1).

5.4.3 ANALYSIS OF THE OPEN READING FRAME IN cK22

The sequence of cK22 contains a long open reading frame of 733 codons beginning at the the most 5' ATG initiation codon (position 130-132), ending at a TAA stop codon (position 2329-2331). In a compilation of translational start sequences in eukaryotic mRNAs, the 5'-most AUG triplet serves as the initiator codon in 95% of cases examined (Kozak, 1984). The long open reading frame in the cK22 cDNA sequence encodes a putative protein product 733 amino acids in length, with a predicted molecular weight of 84kD. The coding region also encompasses the target sites of the P element insertions in $su(f)^{l\text{-}MS97}$ and $su(f)^{l\text{-}MS252}$, and of the Doc insertion in $su(f)^{l\text{-}S2}$ (Figure 5.7). This would suggest that the putative protein product is important for $su(f)^{+}$ function.

Figure 5.8 The intron/exon structure of the transcript represented by cK22

Above the restriction enzyme site map is shown the intron/exon structure of the transcript represented by cK22. The open boxes represent 5' and 3' untranslated sequences; the filled boxes represent open reading frame.

The intron/exon structure of the cK22 transcript was determined by comparing cDNA to genomic DNA sequences. Below the restriction map are shown the position, direction and extent of sequencing of the Canton S genomic DNA M13 subclones used to provide the genomic sequence.

Restriction enzyme key :

B : BamHI	H : HindIII
R : EcoRI	Xb : XbaI
S : SalI	P : PstI
C : ClaI	V : PvuII
Xh : XhoI	G : BglII

FIGURE 5.8

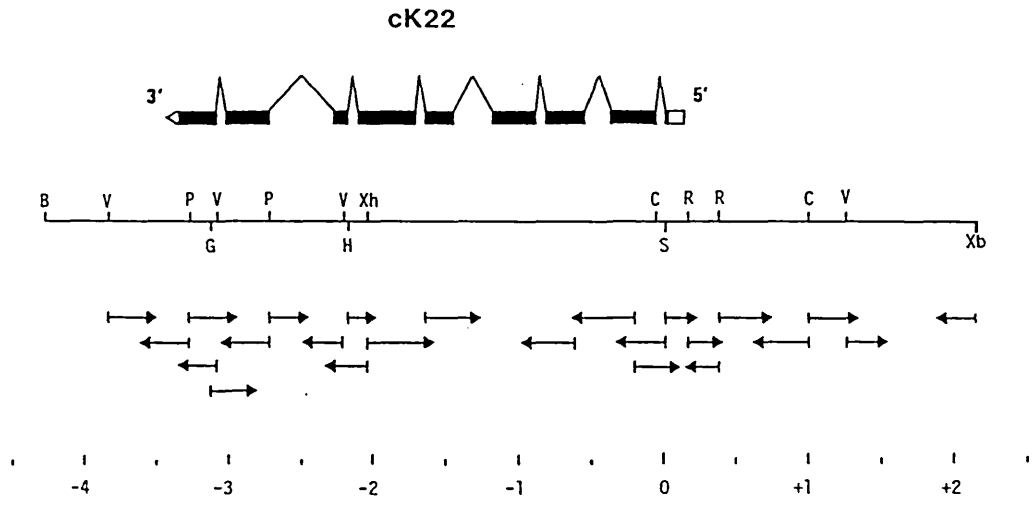


Table 5.1 Splice acceptor and donor sites in the transcript represented by cK22

Shown are the splice acceptor and donor sites of the eight introns found in the transcript represented by cK22; except for the donor site of intron 4, where no genomic sequence was obtained. Exon sequences are in capital letters; intron sequences are in lower case letters. The obligatory **gt** and **ag** intron terminal sequences are underlined. Below the main figure is given the consensus splice donor and acceptor sequences for *D.melanogaster* compiled by Mount (1982).

TABLE 5.1

Intron	Donor	Acceptor
1	..**AGTCGACATA**gtaagtaaaaca.....atcgatgacaag**GAATGGGGCA**..	
2	..**GTACTCACAA**gtaactgttata.....ctcttttctcag**AGAAAAAATG**..	
3	..**AGCGATCTAA**gtgagtgttata.....aactaaaaccag**GGATTACATG**..	
4	..**CGAAAAAGGC**.....aaaacattgcag**GACGTCCAGG**..	
5	..**TCCCACATTG**gtaggaacacac.....ttaatactccag**GTCTATGTTC**..	
6	..**TTTCGAAAAC**gtgcgttctttt.....atttgtttaaag**CTTAAAGAGT**..	
7	..**ATATGCTGAG**gtgcgtaacgga.....ctgcag**AATGTGGGCA**..	
8	..**CTTCCGGACT**gtgagtgtccct.....tatcttactcag**CTGCACCGCA**..	

Drosophila	^{C}AG$\underline{GT}^{G}$AGT	TTTTTTTT$_{N}{}^{T}\underline{AG}$G
Consensus	A A	CCCCCCCC C

An analysis of the codon usage within the predicted protein is shown in Table 5.2. On comparison to a codon usage frequency table compiled from several *D.melanogaster* proteins (Maruyama *et al.*, 1986), the predicted protein shows good codon usage and good codon preference for a *D.melanogaster* protein (Table 5.2). Searches of the EMBL, GenBank and Edinburgh Protein Sequence databases with the sequence of the predicted protein failed to produce any obviously significant homologies to known sequences. Neither does the predicted protein possess an obvious candidate for a leader peptide (Watson, 1984), suggesting that it is probably not a secreted protein. The protein sequence was also examined for 'Zinc-finger' sequence motifs which are found in a range of DNA binding proteins, including the predicted product of the *D.melanogaster su(Hw)* protein (see Chapter 1, section 1.3.4). Like the predicted protein product of the *su(w^a)* gene (Chou *et al.*, 1987), the predicted product of cK22 has no Zinc-fingers. The most striking feature of the predicted protein sequence is the distribution of the 40 proline codons; this number is not in itself unusual by comparison to consensus figures for *D.melanogaster* proteins (Maruyama *et al.*, 1986), but the prolines are unevenly distributed. Of the 40 prolines, 25 are in the C-terminal 20% of the protein, making the C terminus proline rich. As proline is a helix-breaking amino acid, this would suggest that the C terminus of the putative protein is largely devoid of alpha-helix or beta-sheet secondary structure. What implications this has for the overall structure and biological function of the predicted protein remains obscure.

5.5 FURTHER STRUCTURAL AND SEQUENCE ANALYSIS OF cDNA CLONES

5.5.1 STRUCTURAL ANALYSIS OF cK4, cK12, cK18 AND cK19

As described in section 5.3, the termini of the other cDNA clones isolated were examined by single deoxyTTP/dideoxyTTP sequencing reactions (T tracking). Several of these cDNA clones derive from the same transcription unit as cK22 (see Figure 5.6), namely cK1, cK4, cK12, cK13, cK18 and cK19.

Table 5.2 Codon usage in the open reading frame of cK22

Shown are the numbers of each codon of the 733 amino acid open reading frame. The figures in brackets are the numbers of each codon expressed as a percentage of the total. The numbers in bold type are the consensus figures for codon usage in *Drosophila* , taken from Maruyama *et al.* (1986).

TABLE 5.2

Codon	No.	%	*Drosophila* Consensus	Codon	No.	%	*Drosophila* Consensus
TTT Phe	13	(1.8)	**0.8**	TCT Ser	7	(1.0)	**0.4**
TTC Phe	19	(2.6)	**2.8**	TCC Ser	9	(1.2)	**2.4**
TTA Leu	2	(0.3)	**0.1**	TCA Ser	2	(0.3)	**0.3**
TTG Leu	13	(1.8)	**1.1**	TCG Ser	9	(1.2)	**1.5**
TAT Tyr	18	(2.5)	**0.7**	TGT Cys	3	(0.4)	**0.2**
TAC Tyr	17	(2.3)	**2.5**	TGC Cys	5	(0.7)	**1.6**
TAA End				TGA End			
TAG End				TGG Trp	10	(1.4)	**1.1**
CTT Leu	9	(1.2)	**0.3**	CCT Pro	5	(0.7)	**0.5**
CTC Leu	11	(1.5)	**1.3**	CCC Pro	6	(0.8)	**2.4**
CTA Leu	12	(1.6)	**0.3**	CCA Pro	14	(1.9)	**1.0**
CTG Leu	24	(3.3)	**5.1**	CCG Pro	15	(2.0)	**1.0**
CAT His	3	(0.4)	**0.7**	CGT Arg	17	(2.3)	**0.6**
CAC His	13	(1.8)	**1.4**	CGC Arg	18	(2.5)	**2.8**
CAA Gln	11	(1.5)	**0.5**	CGA Arg	6	(0.8)	**0.3**
CAG Gln	15	(2.0)	**3.1**	CGG Arg	10	(1.4)	**0.3**
ATT Ile	12	(1.6)	**1.6**	ACT Thr	3	(0.4)	**0.5**
ATC Ile	17	(2.3)	**4.5**	ACC Thr	13	(1.8)	**3.6**
ATA Ile	10	(1.4)	**0.2**	ACA Thr	9	(1.2)	**0.5**
ATG Met	16	(2.2)	**3.0**	ACG Thr	9	(1.2)	**0.9**
AAT Asn	14	(1.9)	**0.9**	AGT Ser	12	(1.6)	**0.3**
AAC Asn	15	(2.0)	**3.2**	AGC Ser	8	(1.1)	**1.3**
AAA Lys	12	(1.6)	**1.0**	AGA Arg	1	(0.1)	**0.3**
AAG Lys	33	(4.5)	**5.7**	AGG Arg	1	(0.1)	**0.4**
GTT Val	11	(1.5)	**0.6**	GCT Ala	17	(2.3)	**1.5**
GTC Val	14	(1.9)	**2.2**	GCC Ala	21	(2.9)	**4.6**
GTA Val	7	(1.0)	**0.5**	GCA Ala	9	(1.2)	**1.0**
GTG Val	23	(3.1)	**3.5**	GCG Ala	5	(0.7)	**0.9**
GAT Asp	14	(1.9)	**1.9**	GGT Gly	10	(1.4)	**1.6**
GAC Asp	21	(2.9)	**3.9**	GGC Gly	11	(1.5)	**3.6**
GAA Glu	20	(2.7)	**0.8**	GGA Gly	7	(1.0)	**1.9**
GAG Glu	38	(5.2)	**5.7**	GGG Gly	4	(0.5)	**0.2**

cK4, cK12, cK18 and cK19:

For these four clones, the sequences of both their termini fall within the sequence of cK22 at distances apart which agree well with the sizes of these cDNA fragments as established by agarose gel electrophoresis. These clones therefore appear to be structurally contiguous with the larger cK22 cDNA; a conclusion supported by S1 nuclease analysis of double stranded hybrid molecules formed between an M13 clone carrying the 2.5kb fragment of cK22 and appropriately orientated M13 clones of each of the smaller cDNAs. The results of these analyses are shown in Figure 5.9. The size of the S1 nuclease resistant hybrid molecule formed with cK22 again matches the size of the smaller cDNAs as determined by agarose gel electrophoresis. The structures of these four clones are shown in diagrammatic form in Figure 5.10.

5.5.2 cK1 AND THE 3' REGION OF THE *su(f)* TRANSCRIPTION UNIT

The 5' terminal sequence of cK1 falls within the sequence of cK22, but its 3' terminal sequence lies at genomic coordinate -3.7, ~250bp beyond the 3' end of cK22. The sequence of the 3' end of cK1 reads contiguously with genomic sequence up until the point where cK22 ends, after which the sequence matches cK22. These findings are displayed in Figure 5.11, which shows the genomic nucleotide sequence from the PstI site at coordinate -3.3 to the PvuII site at coordinate -3.85.
Marked on the sequence are the positions of the 3' ends of both cK22 and cK1. As cK22 itself has a 3' poly-A tail, this means that the transcript represented by cK1 uses a different polyadenylation signal. There is no poly-A tract at the 3' end of cK1; neither is there an AATAAA polyadenylation signal in the genomic sequence for 100bp in either direction from the 3' end of cK1. However, there is a run of 7 A residues ending 14bp from the 3' end of cK1 which may act as a polyadenylation signal. Alternatively, the transcript represented by cK1 may be joined by an intron to another exon further 3', although there are no good candidates for a splice donor sequence beyond the end of cK1 in the genomic sequence (Figure 5.11).

Figure 5.9 Comparison of cDNA clones by S1 nuclease analysis of M13 subclones

Various M13 subclones of cDNAs were annealed to the appropriately orientated M13 subclone of cK22 and digested with S1 nuclease. The S1-resistant double stranded products were then analysed by agarose gel electrophoresis. The panel shows a photograph of the ethidium-bromide stained gel. The sizes of the marker bands are marked in kb on the left of the panel.

Lane M : λ digested with HindIII; ΦX174 digested with HaeIII

Lane 1 : cK22 / cK25
Lane 2 : cK22 / cK13 (1.3kb EcoRI fragment)
Lane 3 : cK22 / cK19
Lane 4 : cK22 / cK18
Lane 5 : cK22 / cK12
Lane 6 : cK22 / cK16
Lane 7 : cK22 / cK14

Lane 8 : PSTA / PSTB
Lane 9 : PSTA / PSTB : no S1
Lane 10 : PSTB alone : no S1
Lane 11 : PSTA alone : no S1

The M13 clones used in the control lanes (PSTA and PSTB) were the same 550bp PstI fragment cloned in opposite orientations (genomic interval E in Figure 5.6). In lanes 9, 10 and 11, no S1 nuclease was added: the high molecular weight band in lane 9 is the annealing product of PSTA and PSTB. When digested with S1 nuclease, this band is resolved to a 550bp double stranded DNA molecule (lane 8).

FIGURE 5.9

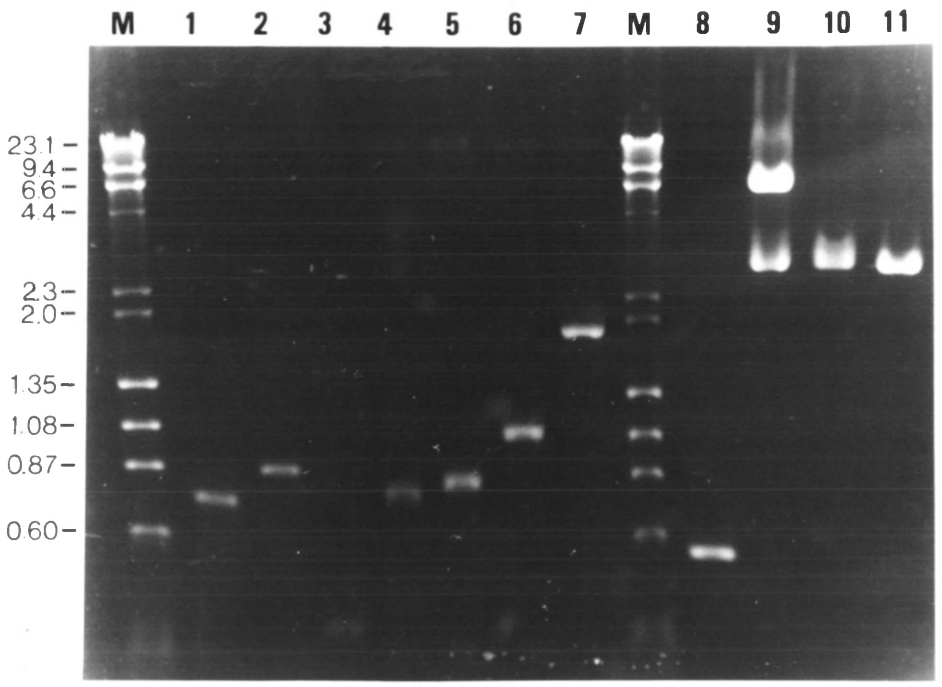


Figure 5.10 Structures of cDNAs related to cK22

Below the intron/exon structure of cK22 is shown cK22 drawn without interruptions. Compared to the latter are other cDNAs of the *su(f)* transcription unit. The open boxes represent non-translated leader; filled boxes represent open reading frame. All of these are contiguous with cK22, except cK1, whose 3' end extends ~250bp beyond that of cK22 (see text). Also shown is the first exon of cK5, which derives from a transcription unit 5' to and transcribed divergently from that of *su(f)*.

Restriction enzyme key :	B : BamHI	H : HindIII
	R : EcoRI	Xb : XbaI
	S : SalI	P : PstI
	C : ClaI	V : PvuII
	Xh : XhoI	G : BglII

n.b. : cK1, cK8 and cK16 are identical clones
cK4, cK6 and cK14 are identical clones
(see text, section 5.3)

FIGURE 5.10

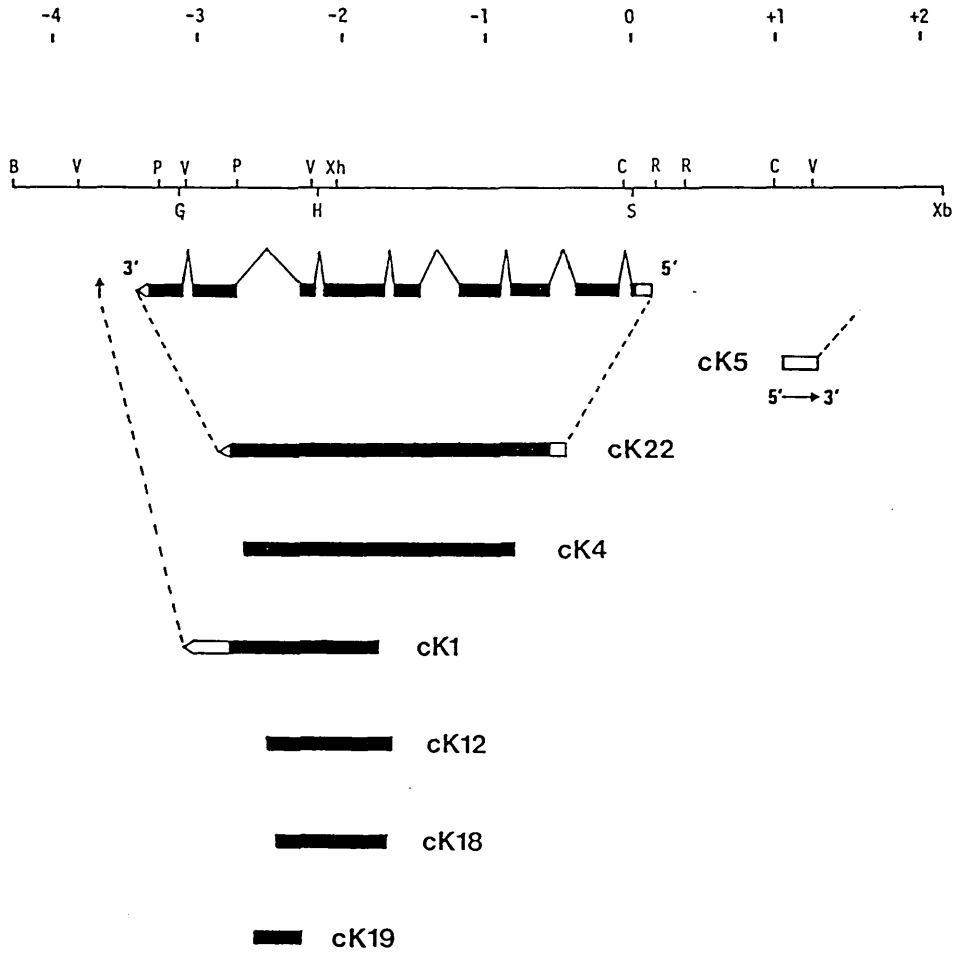
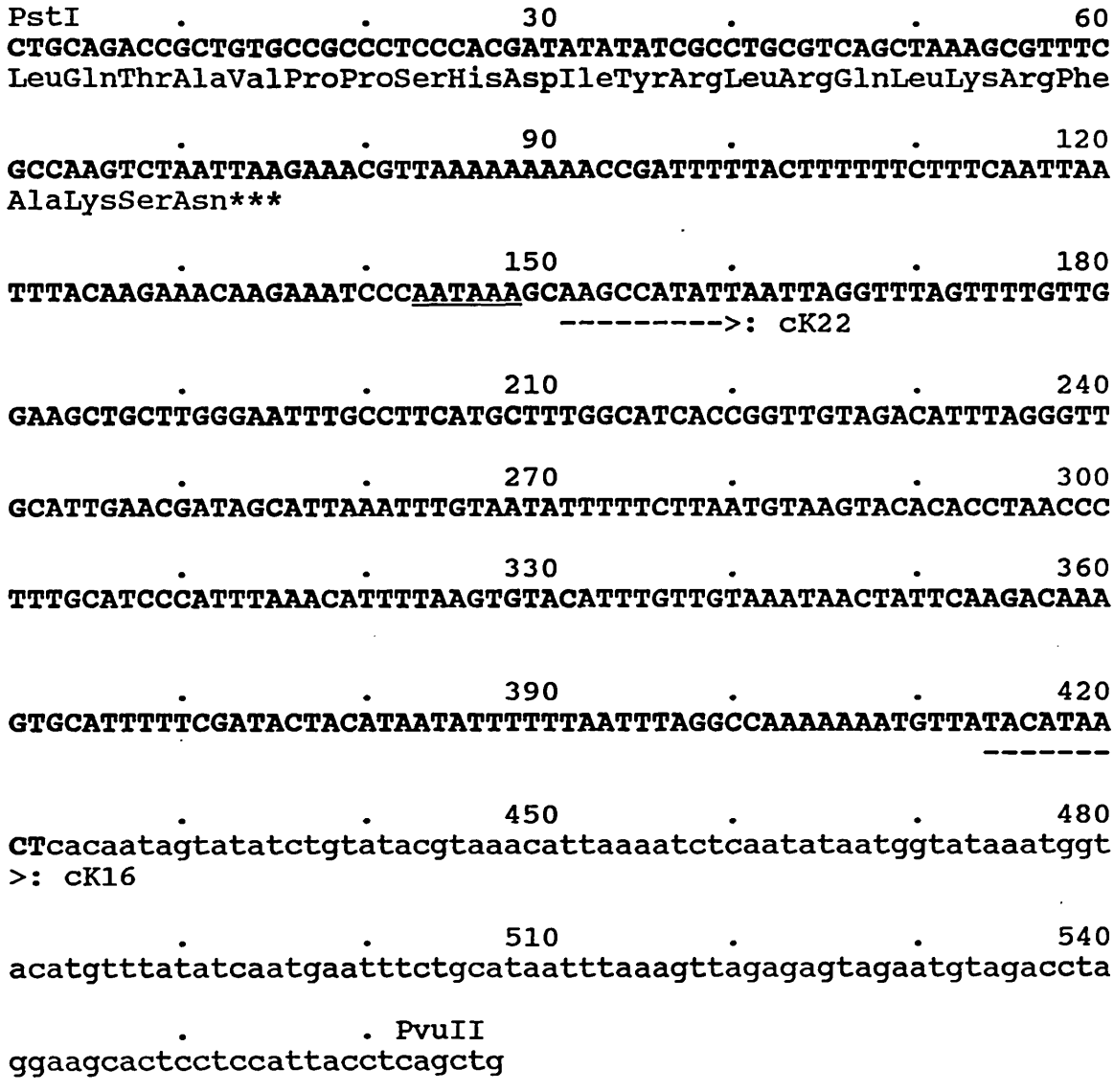

-4
-3
-2
-1
0
+1
+2
B
V
P
V
P
V
Xh
C
R
R
C
V
G
H
S
Xb
3'
5'
cK5
5'→3'
cK22
cK4
cK1
cK12
cK18
cK19

Figure 5.11 Alternate 3' ends for *su(f)* transcripts

Shown is the nucleotide sequence of the PstI/PvuII fragment from Canton S (coordinates -3.3 to -3.8). Sequences in lower type are those not found in either cK22 or cK16. The 5' end of cK22 is shown, along with part of its open reading frame and the whole of its 3' untranslated region. The presumed AATAAA polyadenylation signal for cK22 is boxed. Further 3' to the end of cK22 is shown the 3' end of cK16. There is no obvious candidate for a polyadenylation signal for cK16, the closest to AATAAA is the run of 7 A residues between nucleotides 402 to 408, which may possibly serve this function.

FIGURE 5.11


PstI
30
60
CTGCAGACCGCTGTGCCGCCCTCCCACGATATATATCGCCTGCGTCAGCTAAAGCGTTTC
LeuGlnThrAlaValProProSerHisAspIleTyrArgLeuArgGlnLeuLysArgPhe
90
120
GCCAAGTCTAATTAAGAAACGTTAAAAAAAAAACCGATTTTTACTTTTTTCTTTCAATTAA
AlaLysSerAsn***
150
180
TTTACAAGAAACAAGAAATCCCAATAAAGCAAGCCATATTAATTAGGTTTAGTTTTGTTG
---------->: cK22
210
240
GAAGCTGCTTGGGAATTTGCCTTCATGCTTTGGCATCACCGGTTGTAGACATTTAGGGTT
270
300
GCATTGAACGATAGCATTAAATTTGTAATATTTTTCTTAATGTAAGTACACACCTAACCC
330
360
TTTGCATCCCATTTAAACATTTTAAGTGTACATTTGTTGTAAATAACTATTCAAGACAAA
390
420
GTGCATTTTTCGATACTACATAATATTTTTTAATTTAGGCCAAAAAAAATGTTATACATAA

450
480
CTcacaatagtatatctgtatacgtaaacattaaaatctcaatataatggtataaatggt
>: cK16
510
540
acatgtttatatcaatgaatttctgcataatttaaagttagagagtagaatgtagaccta
. PvuII
ggaagcactcctccattacctcagctg

The different positions of the 3' termini of cK1 and cK22 may provide an explanation of the 2.6 and 2.9kb doublet of RNAs found in the *su(f)* transcription unit. If these two messages are 5' co-terminal, then the 250bp difference in length at their 3' ends would produce the transcript doublet observed on RNA blots. If the 2.9kb RNA is otherwise contiguous with the 2.6kb RNA represented by cK22, it would contain the same long open reading frame, producing the same protein product if translated.

5.5.3 THE 5' REGION OF THE *su(f)* TRANSCRIPTION UNIT

Figure 5.12 shows the genomic sequence of the region around the 5' end of the *su(f)* transcription unit from nucleotides +1500 to -180, where nucleotide +1 is the "+"-most G-C base pair of the SalI site (GTCGAC) at coordinate 0.0. Marked on the sequence is the 5' end of cK22, including its first exon, its first intron and part of its second exon. After removing the 5' poly-C tract and the 3' poly-A tract, the size of cK22 is 2.4kb, which matches well the size of the 2.6kb RNA of the *su(f)* transcription unit, allowing for a poly-A tail. However, the structural analysis of other cDNA clones shows that cK22 may not be full length at its 5' end:

cK13:

Of the cDNA clones in the *su(f)* transcription unit, cK13 is the only one which hybridises to the 200bp EcoRI fragment from coordinates +0.15 to +0.35 (fragment I in Figure 5.6). This cDNA clone has two EcoRI insert fragments of 1.25kb and <0.2kb. The smaller fragment was cloned into an M13 vector and its nucleotide sequence determined. The sequence is shown below:

```
5'
EcoRI     .         .        30        .         .        60
GAATTCCCCCGCtctatcaaaaggctgtagtcacacctatcgtgggcatcgagcagttgt

          .         .        90        .         .       120
ggaaggactatatcgcattcgagcaaaacataaacccgattatatcagagaagatgagtc

          .         .       150      EcoRI
GTCGATATCTAAGCAGCCGCATTGCAACTTCTAAAAGAATTC
                                              3'
```

Figure 5.12 Genomic nucleotide sequence around the 5' end of the *su(f)* transcription unit

Shown is the genomic nucleotide sequence around the 5' end of the *su(f)* transcription unit, from nucleotide +1500 to nucleotide -180. Nucleotide +1 is the "+-most" G residue of the SalI site at coordinate 0.0. Sequences in lower case letters are those not present in any of the cDNAs examined. Selected restriction enzyme sites are shown above the sequence. The 5' ends of cK13, cK25 and cK22 are indicated, along with the beginning of the open reading frame in the *su(f)* transcription unit (nucleotide +28), the whole of the first intron and part of the second exon. Also shown is the 5' end of cK5 (nucleotide +1047) and its first exon, continuing to the splice donor site (nucleotide +1299).

FIGURE 5.12

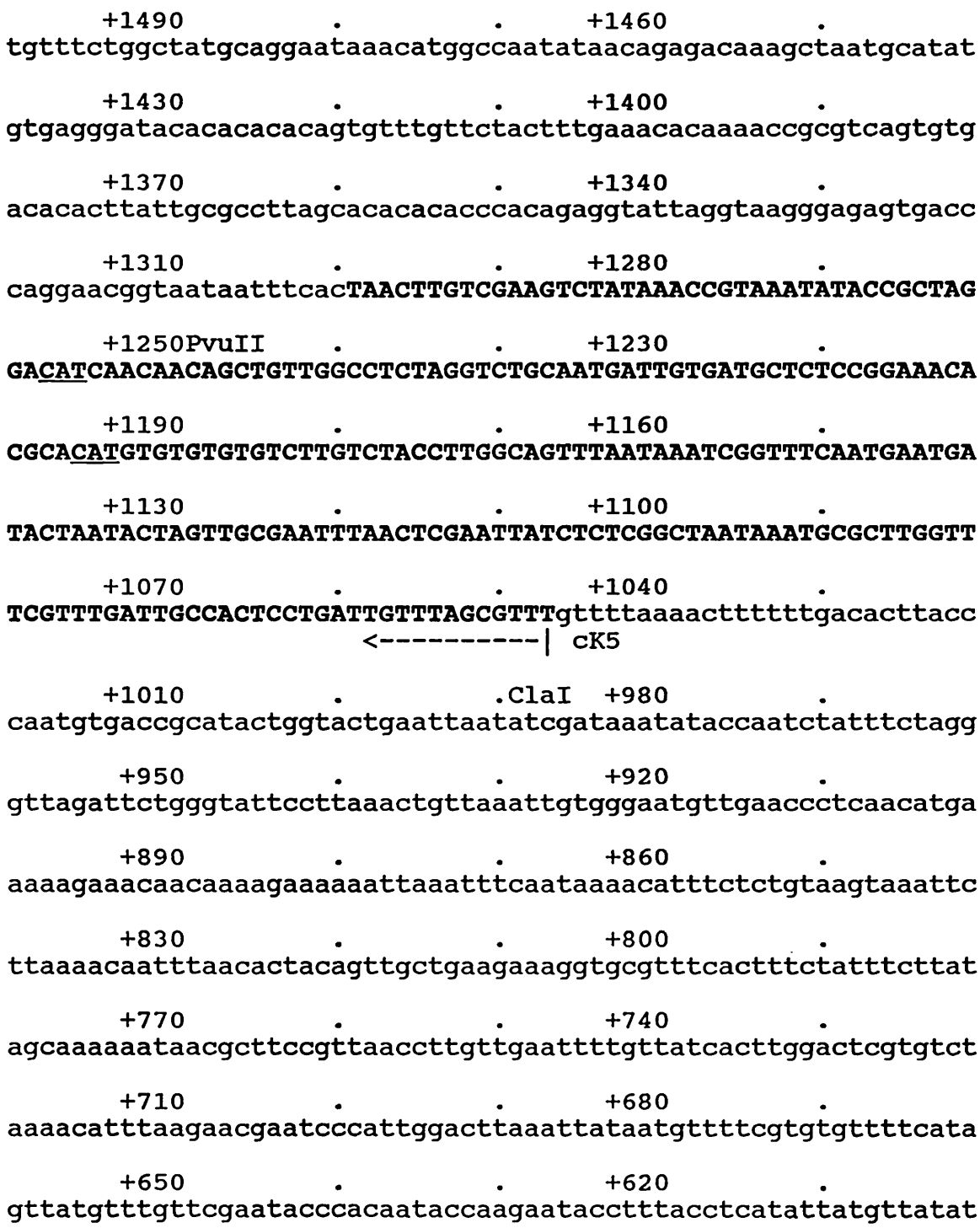
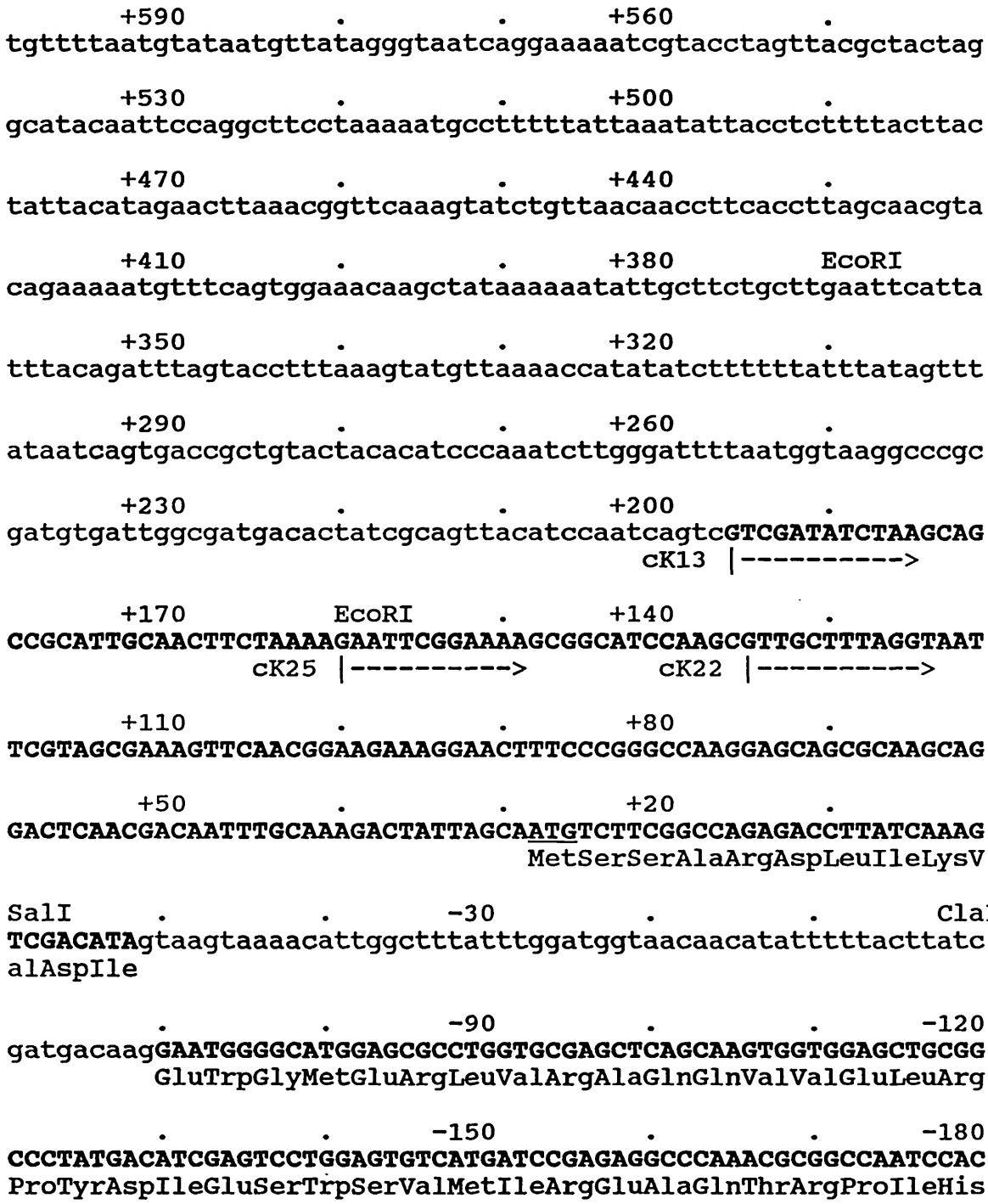

+1490 . . +1460 .
tgtttctggctatgcaggaataaacatggccaatataacagagacaaagctaatgcatat
+1430 . . +1400 .
gtgagggatacacacacacagtgtttgttctactttgaaacacaaaaccgcgtcagtgtg
+1370 . . +1340 .
acacacttattgcgccttagcacacacacccacagaggtattaggtaagggagagtgacc
+1310 . . +1280 .
caggaacggtaataatttcacTAACTTGTCGAAGTCTATAAACCGTAAATATACCGCTAG
+1250PvuII . . +1230 .
GACATCAACAACAGCTGTTGGCCTCTAGGTCTGCAATGATTGTGATGCTCTCCGGAAACA
+1190 . . +1160 .
CGCACATGTGTGTGTGTCTTGTCTACCTTGGCAGTTTAATAAATCGGTTTCAATGAATGA
+1130 . . +1100 .
TACTAATACTAGTTGCGAATTTAACTCGAATTATCTCTCGGCTAATAAATGCGCTTGGTT
+1070 . . +1040 .
TCGTTTGATTGCCACTCCTGATTGTTTAGCGTTTgttttaaaacttttttgacacttacc
<----------| cK5
+1010 . .ClaI +980 .
caatgtgaccgcatactggtactgaattaatatcgataaatataccaatctatttctagg
+950 . . +920 .
gttagattctgggtattccttaaactgttaaattgtgggaatgttgaaccctcaacatga
+890 . . +860 .
aaaagaaacaacaaaagaaaaaattaaatttcaataaaacatttctctgtaagtaaattc
+830 . . +800 .
ttaaaacaatttaacactacagttgctgaagaaaggtgcgtttcactttctatttcttat
+770 . . +740 .
agcaaaaaataacgcttccgttaaccttgttgaattttgttatcacttggactcgtgtct
+710 . . +680 .
aaaacatttaagaacgaatcccattggacttaaattataatgttttcgtgtgttttcata
+650 . . +620 .
gttatgtttgttcgaatacccacaataccaagaataccttacctcatattatgttatat

FIGURE 5.12 (cont.)



From position 121 to the end of the sequence at position 162 (sequences in upper case letters), it matches the sequence of the small genomic EcoRI fragment such that the EcoRI site at position 157 to 162 is that at genomic coordinate +0.15. The position of the 5' end of cK13 is marked on the genomic DNA sequence shown in Figure 5.12. From position 13 to position 120 (sequences in lower case letters) the sequence matches that of cK22 between positions 644 and 751 (part of exon 3). Thus cK13 results from a cloning artefact presumably occurring from anomalous priming of the complete, newly synthesised first cDNA strand at a point on the mRNA template corresponding to exon 3. With respect to the 3' end of cK13, the sequence of one end of the larger EcoRI fragment shows a poly-A tract of approximately 85bp which prevented a reading of the sequence beyond. The other end of this fragment has a nucleotide sequence which begins at the genomic EcoRI site at coordinate +0.15, and therefore carries on directly from the sequence of the small EcoRI fragment given above. S1 nuclease digestion of a hybrid between an M13 clone of the larger, 1.25kb EcoRI fragment of cK13 and an M13 clone of cK22 gave a protected double stranded molecule of 0.85kb. Even taking into account the poly-A tail of cK13, the size of the protected fragment is smaller than would be expected if cK13 were contiguous along its length with cK22. The transcript represented by cK13 may thus perhaps differ in intron/exon structure from that represented by cK22, although the data available shows that the two are mostly contiguous.

cK25:

In addition to the ten different cDNA clones obtained from the first/second instar cDNA library, another cDNA clone was later isolated after screening an adult female cDNA library (Poole *et al.*, 1985). The probe used in this case was an M13 clone, S6.22: a clone derived from the sequencing of cK22. It carries DNA from positions 145 and 249 in the cK22 sequence (Figure 5.7) and was used in order to select for cDNA clones from the *su(f)* transcription unit which were full length at the 5' end. The single clone obtained, cK25, has a single EcoRI insert fragment of ~0.75kb. On

subcloning this fragment into an M13 vector, only one orientation was obtained. The sequence of this terminus reads from the genomic EcoRI site at coordinate +0.15, 24bp further 5' to the start of cK22. Thereafter it matches the cK22 sequence. The nucleotide sequence of an M13 subclone from λcK25 reading from the 5' proximal SalI site within the inserted fragment reads into λgt10 vector sequences after the 5' EcoRI site, showing that there is not a second EcoRI cDNA fragment at this end. S1 nuclease digestion of the hybrid formed between M13 clones of cK25 and cK22 gave a protected fragment of ~0.75kb (Figure 5.9), suggesting that cK25 is contiguous along its length with cK22, except for an additional 24bp at the 5' end. The position of the 5' end of cK25 is marked on the genomic DNA sequence shown in Figure 5.12.

An examination of the genomic sequence around the 5' ends of cK13, cK25 and cK22 failed to reveal either any good candidates for a Goldberg-Hogness TATA box, or candidate for a transcriptional startsite (Busslinger *et al.*, 1980).

cK5:

Of the three cDNA clones which hybridise to genomic sequences between coordinates +1.0 and +2.15, cK5 is the longest at 1.7kb. These cDNAs appear to belong to a transcription unit lying 5' to the *su(f)* transcription unit. To discover its approximate extent, nick-translated λcK5 was hybridised to a filter carrying almost all the 32kb of the cloned genomic interval. The results of this DNA blot are shown in Figure 5.13. The cDNA hybridises to two separate genomic intervals; the first between coordinates 0.0 and +2.15 and the second between +6.7 and +9.2, suggesting the existence of an intron at least 4.5kb in size. The extent of cK5 in relation to genomic sequence is shown diagramatically in Figure 5.14. The EcoRI insert fragment from cK5 was subcloned into M13 vectors and the nucleotide sequences of both termini examined. Neither end of cK5 has a poly-A tract. The sequence of one end fails to match any available genomic sequence. The sequence of

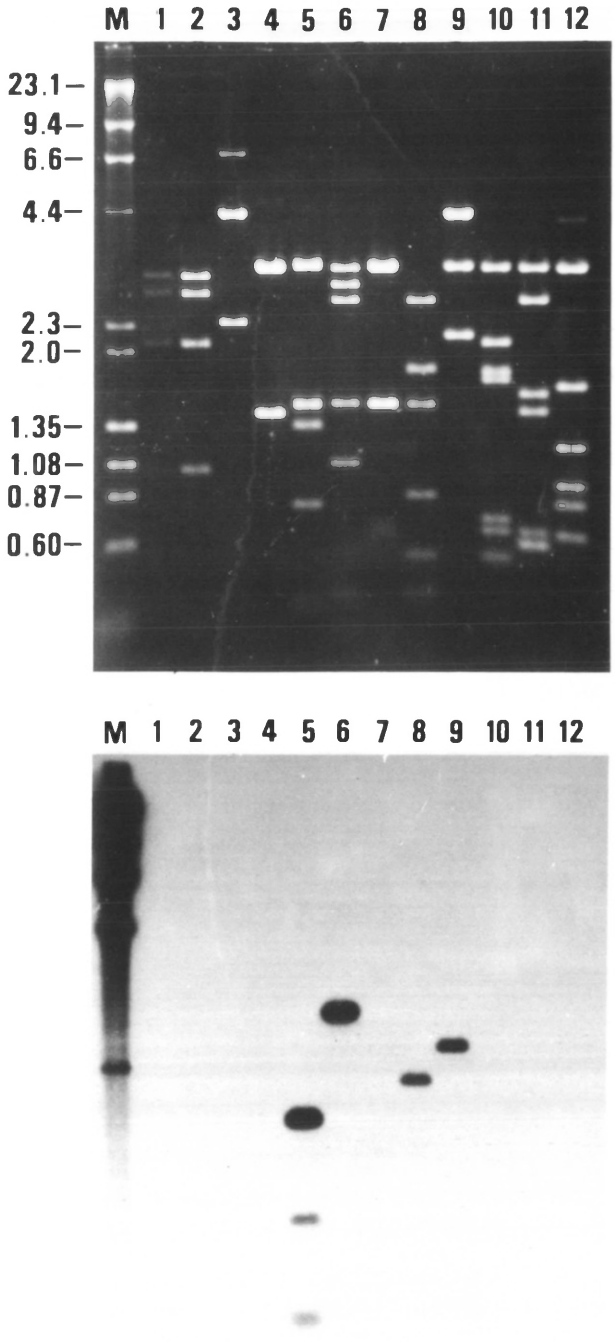
Figure 5.13 The extent of hybridisation of cK5 to genomic sequences

The upper panel shows an ethidium-bromide stained gel carrying restriction enzyme digests of plasmid subclones covering nearly all of the 32kb *su(f)* region from Canton S. The sizes of the marker bands in kb are shown on the left (lane M). The fainter bands visible in lanes 3 and 12 are partial digest products. 500ng of each plasmid digest was loaded per lane, except for lane 1, where only 50ng was loaded.

Lane M : size markers - λ digested with HindIII; ΦX174 digested with HaeIII.
Lane 1 : pB104"Sal" digested with SalI, BamHI and XhoI (50ng DNA)
Lane 2 : pB104"Sal" digested with SalI, BamHI and XhoI (500ng DNA)
Lane 3 : pm11.3 digested with EcoRI and HindIII (*white* DNA)
Lane 4 : pCS52 digested with EcoRI and XbaI
Lane 5 : pCS5 digested with EcoRI, XbaI and SalI
Lane 6 : pCS5 digested with EcoRI, XbaI and HindIII
Lane 7 : pCS51 digested with XbaI and BamHI
Lane 8 : pB1.R digested with EcoRI, XbaI, BamHI and SalI
Lane 9 : pBX64 digested with BamHI and SalI
Lane 10 : pCSL digested with EcoRI, BamHI, HindIII and XbaI
Lane 11 : pCSL digested with EcoRI, BamHI, HindIII and XhoI
Lane 12 : pCS4 digested with EcoRI and HindIII

The lower panel shows the DNA blot filter of this gel probed with λcK5. This cDNA clone hybridises to the 1.8kb EcoRI-XbaI fragment from pB1.R in lane 8 (coordinates +0.35 to +2.15) and three fragments from pCS5 in lane 5: the 0.8kb XbaI-SalI fragment (coordinates +6.6 to +7.4), the 0.4kb SalI fragment (coordinates +7.4 to +7.8) and the 1.4kb SalI-XbaI fragment (coordinates +7.8 to +9.2).

FIGURE 5.13



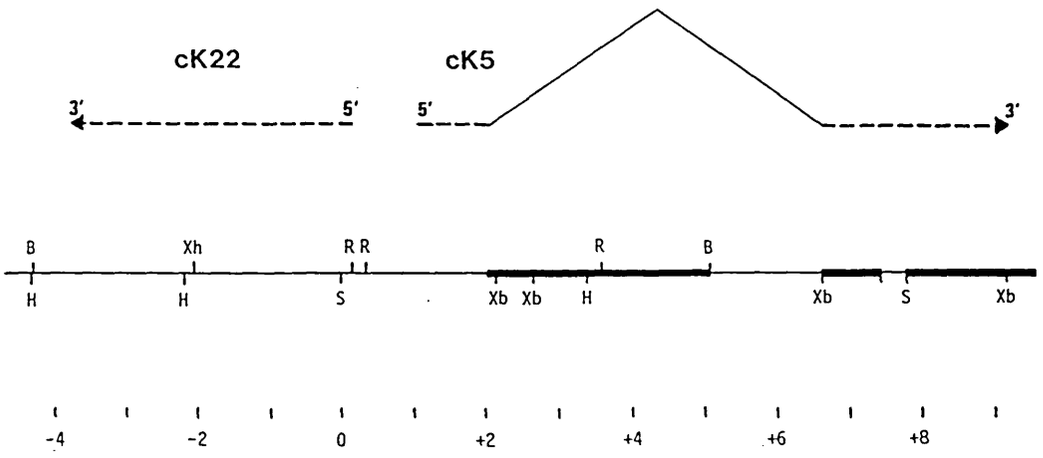
λcK5

Figure 5.14 Structure of the cK5 transcription unit

Shown above the restriction map is the structure of the cK5 transcription unit as determined by DNA blotting using λcK5 (Figure 5.13) with the approximate extent of the large intron. This transcription unit is shown in relation to the *su(f)* transcription unit; introns have been omitted from the latter for clarity. The two transcription units are divergently transcribed. Thick black lines on the restriction map represent repetitive DNA sequences; the thin lines represent unique DNA sequences. The regions represented as repetitive DNA sequences are the maximum extents of these sequences: the cK5 transcript does not necessarily contain repetitive DNA itself.

Restriction enzyme key :	B : BamHI	H : HindIII
	R : EcoRI	Xb : XbaI
	S : SalI	Xh : XhoI

FIGURE 5.14



the other end of cK5 matches genomic DNA sequence, starting at coordinate +1.05 and reading to the right with respect to the DNA map. This is also the direction of transcription, as shortly after the PvuII site at coordinate +1.25 there is a discontinuity between genomic and cDNA sequence: at the beginning of this discontinuity is a splice donor sequence. The position of this 5' exon of cK5 is shown diagramatically in Figure 5.10, and it is marked on the genomic DNA sequences shown in Figure 5.12. Examination of the 5' nucleotide sequence of cK5 shows two ATG initiation codons, but both are followed by in-phase termination codons, respectively 20 and 2 codons later. This part of the cDNA therefore probably represents 5' untranslated leader sequence. The genomic DNA sequence of the region around the 5' end of cK5 is shown below:

```
5'        .         .    ClaI  .         .         .      +1020
cctagaaatagattggtatatttatcgatattaattcagtaccagtatgcggtcacattg

          .         .      +1050        .         .      +1080
ggtaagtgtcaaaaaagttttaaaacAAACGCTAAACAATCAGGAGTGGCAATCAAACGA
                           |----------->  cK5                3'
```

The sequences in capital letters are those at the 5' end of cK5. There is a candidate Goldberg-Hogness TATA box underlined at positions +989 to +995, followed 22bp later by a candidate sequence for a transcriptional startsite: CATTGG. The consensus sequence for eukaryotic mRNA startsites is $\frac{C}{T}CATTC\frac{A}{G}$, where the A residue at the third position is the first base in the mRNA (Busslinger, 1980). In the above genomic sequence, the putative initiating A residue at position +1017 is underlined.

5.6 CONCLUDING REMARKS

This chapter has described the analysis of poly-A^+ RNA transcripts in the region where DNA lesions associated with mutant *su(f)* alleles are clustered. A number of transcripts were detected; the most likely candidates for $su(f)^+$ messenger RNAs derive from a transcription unit lying between genomic coordinates +0.35 and -4.3. These RNAs (2.9, 2.6 and 1.3kb in size) are qualitatively altered in the mutant

alleles $su(f)^1$ and $su(f)^{l-S2}$. In wild type flies, the three transcripts are present throughout development, but are most abundant in embryos, early pupae and female adults. This profile is in keeping with the range of mutant phenotypes shown by ts alleles of *su(f)* throughout development; and also raises the possibility that the abundance of *su(f)* transcripts in embryos may be due to maternal contribution. The "embryo" poly-A^+ RNA sample employed was prepared from a 0-16h egg-laying collection; performing RNA blots using samples from earlier and shorter developmental intervals might be useful in clarifying this point.

The investigation of the transcriptional products of *su(f)* continued with the isolation of several cDNA clones and their sequence analysis. This provided evidence for alternative 5' polyadenylation sites for transcripts from the *su(f)* region, as well as the presence of a transcription unit divergent to that of *su(f)*. The clone cK22 corresponds to a nearly full-length copy of the 2.6kb transcript; the most abundant RNA species in the *su(f)* transcription unit. It is not clear where the 5' end of the transcription unit lies, or indeed whether there are several alternative transcritpional startsites. The fact that the 5' ends of cK22, cK13 and cK25 lie relatively close to one another would suggest that the true 5' end is not far away. An S1 nuclease mapping or primer extension experiment would need to be carried out to resolve this issue.

The 2.6kb transcript represented by cK22 encodes a putative 84kD protein product. Evidence that this putative protein is likely to play a role in $su(f)^+$ function is provided by the positions of the insertion sites of the transposable elements associated with the lethal alleles $su(f)^{l-MS97}$, $su(f)^{l-MS252}$ and $su(f)^{l-S2}$ (Figure 5.7). All three sites lie within the open reading frame encoding the putative 84kD protein, and for $su(f)^{l-MS252}$ at least, DNA blotting analysis of revertants showed that the insertion of the P element was indeed responsible for the lethal phenotype (Chapter 4).

CHAPTER 6

GENERAL DISCUSSION

6.1 THE STRUCTURE OF THE GENOMIC REGION AROUND *su(f)*

The 32kb interval of cloned DNA around the *su(f)* region has a high content of repetitive DNA sequences, interspersed with regions of unique sequences. This structure is unlike that surrounding euchromatic genes; for example, the 195kb of DNA from the Bithorax complex and the 110kb around the *achaete-scute* complex consist essentially of unique DNA (Bender *et al.*, 1983; Campuzano *et al.*, 1985). The organisation of DNA sequences around *su(f)* is similar to other regions of β-heterochromatin that have been examined (Miklos *et al.*, 1988) in that unique regions are found embedded within regions of repetitive DNA. This structural organisation bears out the prediction that repetitive sequences may accumulate in locations where meiotic recombination rarely occurs, such as telomere or heterochromatic regions (Charlesworth *et al.*, 1986).

During the course of this work, no differences in the restriction site maps of the *su(f)* region were found between Canton S, *FM6* or π_2. Neither did any of the mutant *su(f)* alleles examined show any restriction site polymorphisms apart from the discrete DNA lesions that were found in several cases. In a study of the restriction maps of 64 wild type isofemale lines using probes from *su(f)*, no polymorphisms were identified other than six lines which appeared to have insertions of DNA (C.H.Langley, pers.comm.). The scarcity of polymorphism between widely varying strains of *D.melanogaster* is presumably due to the low crossing over rate in this heterochromatin-proximal region, which would tend to prevent the fixation of new variants in the population.

Little is known so far of the nature of the repetitive DNA sequences around *su(f)*, other than that most of the repeats are not reiterated within the cloned interval, and are present in the genome at low copy number (or have poor homology to other repeats). The exceptions are the tandemly orientated copies of the "1.5kb" repeats which flank the ~7kb unique region where the *su(f)* transcription unit lies. By

"reverse Southern" DNA blot analysis these give a stronger signal and appear to be repeated in the genome at a higher copy number. Mapping and partial sequencing of these repeats showed that they did not correspond to any known transposable element family in *D.melanogaster* (Finnegan and Fawcett, 1986); neither were any structural features characteristic of transposable elements found within them. Further analysis of the "1.5kb" repeats in the *su(f)* region and elsewhere would need to carried out to clarify their origin.

The probable *su(f)* transcription unit covering coordinates +0.2 to -3.7 fits neatly into the ~7kb unique DNA island of the cloned interval. This is not true for the divergent transcription unit lying ~900bp 5' to that of *su(f)*, which is closely interspersed with repetitive DNA sequences, including a large intron containing such sequences. Euchromatic genes which function perfectly with introns containing repetitive DNA are known: the allele w^{i+A} has an F element within the 3kb first intron of *white* but has a wild type phenotype (Karess and Rubin, 1982). The Doc element found on the bx^3 chromosome along with the gypsy insertion responsible for the mutant phenotype is still present in the wild type revertant bx^3-*rv*, yet it is contained within an intron of the *Ubx* transcription unit (Bender *et al.*, 1983; Peifer and Bender, 1986). Transcription units interwoven with repetitive DNA sequences are therefore not unique; though the transcription unit 5' to that of *su(f)* is perhaps an extreme example. Whether it is a gene with a defined function is not known. The next most proximal known complementation group to *su(f)* is *l(1)sph*, but this locus may lie some distance away, and possibly in the other direction. According to Lindsley and Zimm (1986), there are several putative deficiencies that are $su(f)^-$ and sph^+; i.e. that break between the two loci. These are *Df(1)R8*, *Df(1)R16*, *Df(1)R18* and *Df(1)16.1.85*; the last of which has been characterised in this work as $su(f)^{l-16.1.85}$ and is probably not a deficiency (Chapter 4, section 4.4). All four mutations are $bobbed^+$, and as there are no known loci proximal to *su(f)* before *bobbed*, it is just as likely that the other

three are point mutations for *su(f)* and not deficiencies. The most interesting mutation in this respect is $su(f)^{l-L26}$, which is sph^{+}. DNA blotting evidence suggests that $su(f)^{l-L26}$ may be deficient at least for the interval covered by pBX64 (Chapter 4, section 4.4). If this is the case, then the transcription unit divergent to that of *su(f)* cannot be *sph*, as pBX64 covers the 5' region of its transcribed interval.

6.2 THE STRUCTURE OF THE *su(f)* GENE

With respect to the three RNA species found in the *su(f)* transcription unit, evidence has been presented to suggest that the 2.9kb RNA is functionally identical to the 2.6kb RNA; merely having a longer 3' untranslated region. However, little is known of the structure of the 1.3kb RNA, which may or may not play some role on *su(f)* function. The precise position of the Doc element insertion in $su(f)^{l-S2}$ apparently lying beyond the 3' end of the smaller RNA, would suggest that it does not. The cDNA clone cK13 is smaller than cK22, which represents the 2.6kb RNA; yet the two clones have 5' ends which are closely adjoining. Yet cK13 has a 3' poly-A tract, so an examination of the precise 3' end and intron/exon structure of this clone might prove informative as to the structure of a transcript from *su(f)* differing from that represented by cK22. Another approach to the relative contributions of the different transcripts would be to use P element mediated germ line transformation to test the ability of a chimaeric cDNA/genomic DNA construct to rescue *su(f)* mutant phenotypes. Several possibilities present themself, including replacing the 3.1kb SalI-BglII fragment of the genomic clone pBX64 with the corresponding fragment from cK22. This would produce a construct with the appropriate 5' and 3' flanking regions, but where all eight introns of the 2.6kb message had been removed. Similarly, replacing the genomic 2.1kb SalI-XhoI fragment with the corresponding cDNA fragment would remove the first five introns. The use of in vitro removal of introns to test the functional abilities of different transcripts from genes has been strikingly successful in the study of P

elements (Laski *et al.*, 1986).

Is the divergent transcription unit part of *su(f)*? This question cannot be unequivocally answered without the analysis of germ line transformants containing a defined interval of the *su(f)* region, but the available evidence suggests that it is not. Firstly, almost all the DNA lesions found associated with lethal *su(f)* alleles map within the transcription unit between coordinates +0.2 and -3.7. Secondly, the structural analysis of the transposable element insertion alleles $su(f)^{l\text{-}MS97}$, $su(f)^{l\text{-}MS252}$ and $su(f)^{l\text{-}S2}$, and the DNA blotting analysis of revertants of $su(f)^{l\text{-}MS252}$ demonstrated conclusively that this transcription unit is required for $su(f)^{+}$ lethal function. These results would suggest that the lethal fuction of *su(f)* is not associated with the divergent transcription unit. This applies also to the modifying function of *su(f)*, as typified by $su(f)^{1}$: none of the lethal alleles complement the mutant phenotype of $su(f)^{1}$; furthermore the temperature sensitive alleles of *su(f)* show a continuum of phenotypes according to the temperature at which they are grown, from wild type (18^{o}C), to $su(f)^{1}$-like (25^{o}C), to lethal (29^{o}C). No DNA lesions were detected in the *su(f)* region for $su(f)^{1}$ or the temperature sensitive alleles, suggesting that they are point mutations, which in turn suggests that all the mutant phenoytpes associated with *su(f)* alleles can be attributed to one transcription unit. Support for this conclusion is provided by the analysis of poly-A^{+} RNA transcripts from mutant alleles of *su(f)*. Qualitative differences in the pattern of transcripts from the "*su(f)*" transcription unit were detected both in a viable, modifying allele ($su(f)^{1}$) and in a lethal allele ($su(f)^{l\text{-}MS252}$) (Chapter 5, section 5.2.3). Closely adjoining transcription units are not unknown in *D.melanogaster*. In the cases of the genes *transformer* and $su(w^{a})$, P element mediated germ line transformation experiments showed that each was genetically separate from its surrounding transcription units (McKeown *et al.*, 1987; Zachar *et al.*, 1987a). More transcripts than identified genes appears to be a characteristic of *D.melanogaster*. In a

region on chromosome III, Hall *et al.* (1983) found 20 poly-A^+ transcripts in a region encompassing 12 lethal complementation groups.

Do the structures of the various mutant *su(f)* alleles provide an understanding of the intra-allelic complementation of $su(f)^{R-9-18}$ with other *su(f)* lethal alleles? At 25°C, the alleles $su(f)^{l-X1}$, $su(f)^{l-X2}$, $su(f)^{l-X3}$, $su(f)^{l-3DES}$ and $su(f)^{l-S2}$ all complement $su(f)^{R-9-18}$ for lethality. Yet other alleles show no complementation, such as $su(f)^{l-L26}$, $su(f)^{l-16.1.85}$, $su(f)^{l-D13}$ and $su(f)^{l-S1}$ (Table 1.2). On examining the DNA lesions associated with these alleles, a logical correlation between their structures and their interactions with $su(f)^{R-9-18}$ is lacking. Non-complementation with gross lesions such as the putative large deletion of $su(f)^{l-L26}$ and the inversion/large insertion of $su(f)^{l-16.1.85}$ is understandable. In the allele $su(f)^{l-S1}$ there is a ~250bp insertion or tandem duplication between coordinates 0.0 and -2.1; in the allele $su(f)^{l-S2}$ there is a Doc insertion near the 3' end of the transcription unit at coordinate -3.1. However, the latter complements $su(f)^{R-9-18}$, but the former does not. Neither does the position of the lesion within the transcription unit correlate with $su(f)^{R-9-18}$ complementation or non-complementation. Both $su(f)^{l-X1}$ and $su(f)^{l-X2}$ complement $su(f)^{R-9-18}$, yet the former is a ~150bp deletion towards the 3' end in the interval between coordinates -2.2 to -4.3, and the latter is a 1kb deletion removing the SalI site near the 5' end of the transcription unit. The molecular basis of the intra-allelic complementation with $su(f)^{R-9-18}$ remains unclear.

The alleles $su(f)^{l-MS97}$ and $su(f)^{l-MS252}$ are both associated with the insertion of identical defective P elements. The fact that the two P elements are identical may be due to the presence of least two elements of precisely this structure in π_2, the mutagenic P strain used to produce each allele (O'Hare and Rubin, 1983; S.McGrath and K.O'Hare, unpublished). In addition, the two independent insertions lie only 75bp apart, close to the 5' end of transcription unit. Two P element

insertions is a statistically meaningless number, yet the insertion into the 5' ends of other genes at close but separate sites has been documented before, notably at *singed* (Roiha *et al.*, 1988).

In spite of its chromosomal location, the *su(f)* transcription unit is not atypical of genes in *D.melanogaster*; it is merely surrounded by an unusual amount of repetitive DNA sequences. The 2.6kb RNA represented by cK22 is multiply spliced; of the eight introns, five are between 55bp and 65bp in size. Bingham *et al.* (1988) have made the suggestion that small introns of this type represent a "limit size" class of the minimum size that will function as a splicing substrate; evolutionary history having reduced what previously may have been larger introns to a minimum.

6.3 APPROACHES FOR FURTHER WORK

The isolation and characterisation of the *su(f)* locus described in this thesis provides an excellent starting point for further work in a number of interesting areas. Firstly, the clones available provide for continued examination of the molecular basis of suppression and enhancement in *D.melanogaster*. One obvious approach would be to use a cDNA clone from the *su(f)* transcription unit to make fusion proteins in order to begin characterisation of the putative protein product of *su(f)*. The SalI-PstI fragment from cK22 contains 96% of the open reading frame encoding the putative protein product of the 2.6kb RNA, lacking only 9 amino acids at the N-terminus and 22 amino acids at the C-terminus. Cloning of this convenient fragment into a β-galactosidase expression vector should enable the isolation of a chimaeric fusion protein. Once purified, such a protein could be used to make antibodies (both polyclonal sera and monoclonal antibodies) directed against epitopes on the putative *su(f)* protein. Such antibodies would be extremely useful in the analysis of the biological function of *su(f)*. At a general level, they could be used to reveal the distribution of *su(f)* protein in all tissues throughout

development, and its intracellular distribution. This might provide a better understanding of the cellular role of *su(f)*. Furthermore, they could reveal the presence of any cross-reacting proteins related to the predicted 84kD *su(f)*protein, possible products of alternative transcripts which might have a biological role. This approach could be extended to include the study of components which interact with a *su(f)* protein. If *su(f)* plays a role in the production of mature mRNAs, as appears to be the case for *su(w^a)*, these might include components of the splicing apparatus, or ribonucleoprotein particles.

With regard to a possible role for *su(f)* as a modifier gene acting directly on transposable elements, antibodies against a *su(f)* protein could be used to establish whether it is a DNA binding protein, as appears to be the case for the *su(Hw)* protein which contains multiple Zinc-fingers (S.Parkhurst and V.Corces, pers.comm.). Experiments using in vitro binding assays of cell extracts to defined regions of gypsy or copia DNA could be used to establish if the *su(f)* product is able to bind specifically to transposable element sequences in its role as a modifier of transposable element caused mutations. This approach has been extremely successful in studying the binding of the *zeste* protein to specific regulatory sequences in *white* and *Ubx* (Benson and Pirrotta, 1987; Biggin *et al.*, 1988).

Unique DNA clones from *su(f)* provide an excellent starting point for investigating the sequence organisation of the euchromatin/heterochromatin transition region. Firstly, they provide a starting point for further chromosomal walking within β-heterochromatin, possibly to other genes in the proximal X. This assumes that interspersed unique DNA sequences occur frequently enough for continued cloning steps, which may not be the case. Secondly, probes from *su(f)* allow an examination of the transition from DNA regions which are multiply replicated in larval polytene chromosomes, and those which are under-replicated. For example, is the *su(f)*

region itself fully replicated in polytene chromosomes? Simple experiments comparing the hybridisation of DNA from *su(f)* and from a euchromatic X-linked gene to genomic DNA isolated from polytene and diploid chromosomes might provide at least an approximate answer. The fact that unique DNA probes from *su(f)* give a detectable in situ hybridisation signal to polytene chromosomes would suggest that this chromosomal region is not wholly under-replicated (M.Yamamoto and G.Miklos, unpublished).

Another series of experiments might investigate the physical structure of the β-heterochromatic region of polytene chromosomes. If the *su(f)* region is itself not fully replicated, then one might predict that restriction fragments generated from digestion with rare-cutting enzymes such as NotI and SfiI would have a very curious structure in this transition region: perhaps a "christmas tree-like" structure. Very large restriction fragments containing *su(f)* might therefore have anomalous migration characteristics in pulse-field electrophoresis gels if polytene chromosomal DNA were used. Closely adjoining chromosomal regions which differ greatly in their replicative properties have previously been studied in the DNA regions surrounding *D.melanogaster* chorion genes (Spradling, 1981). Multiply forked DNA structures have been seen in electron microscopic analyses of amplifying chorion gene regions of follicle cell chromosomes (Osheim and Miller, 1983). In the case of chorion gene amplification, multiply replicated regions are separated from unreplicated regions by only a few tens of kb (Spradling, 1981). Is the transition in larval polytene chromosomes from multiply replicated euchromatic sequences to unreplicated heterochromatic sequences as sharply defined as this? The availability of DNA probes from *su(f)* provide a unique resource for addressing problems such as these.

REFERENCES

Ananiev, E.V., Barsky, V.E., Ilyin, Y.V. and Ryzic, M.V. (1984). The arrangement of transposable elements in the polytene chromosomes of *Drosophila melanogaster*. *Chromosoma (Berl.)* **90**, 366-377.

Appels, R. and Peacock, W.J. (1978). The arrangement and evolution of highly repeated (satellite) DNA sequences with special reference to *Drosophila*. *Int.Rev.Cytol.* **8**, suppl. 69-126.

Appleyard, R.K. (1954). Segregation of new lysogenic types during growth of a doubly lysogenic strain derived from *Escherischia coli* K12. *Genetics* **39**, 440-452.

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

Artavanis-Tsakonas, S., Muskavitch, M.A. and Yedvobnick, B. (1983). Molecular cloning of *Notch*, a locus affecting neurogenesis in *Drosophila melanogaster*. *Proc.Nat.Acad.Sci.USA* **80**, 1977-1981.

Ashburner, M. and Berendes, H.D. (1978). Puffing of polytene chromosomes. In: *The Genetics and Biology of Drosophila* (ed. Ashburner, M. and Wright, T.R.F), **2b**, Academic Press, New York, pp. 315-395.

Babu, P. and Bhat, S. (1980). Effect of *zeste* on *white* complementation. In: *Development and Neurobiology of Drosophila* (ed. Siddiqi, O., Babu, P., Hall, L.M. and Hall, J.C.) Plenum Press, New York, pp. 35-40.

Bankier, A.T. and Barrell, B.G. (1983). Shotgun DNA sequencing. In: *Techniques in Life sciences. Nucleic Acid Biochemistry* (ed. Flavell, R.A.) **B5**, Elsevier Scientific Publishers Ltd., Ireland, pp. 1-34.

Barigozzi, C., Dolfini, S., Fraccaro, M., Rezzonico Raimondi, G. and Tiepolo, L. (1966). In vitro study of the DNA replication patterns of somatic chromosomes of *Drosophila melanogaster*. *Exp.Cell Res.* **43**, 231-234.

Beckendorf, S.K and Kafatos, F.C (1976). Differentiation in the salivary glands of *Drosophila melanogaster*: characterisation of the glue proteins and their developmental appearance. *Cell* **9**, 365-373.

Bender, W., Akam, M., Karch, F., Beachy, P.A., Peifer, M., Spierer, P., Lewis, E.B. and Hogness, D.S. (1983). Molecular genetics of the *bithorax* complex in *Drosophila melanogaster*. *Science* **221**, 23-29.

Benson, M. and Pirrotta, V. (1987). The product of the *Drosophila zeste* gene binds to specific DNA sequences in *white* and *Ubx*. *EMBO J.* **6**, 1387-1392.

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

Biessman, H. (1985). Molecular analysis of the *yellow* gene (*y*) region of *Drosophila melanogaster*. *Proc.Nat.Acad.Sci.USA* **82**, 7369-7373.

Biggin, M.D., Bickel, S., Benson, M., Pirrotta, V. and Tjian, R. (1988). *zeste* encodes a sequence-specific transcription factor that activates the *Ultrabithorax* promoter in vitro. *Cell* **53**, 713-722.

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

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

Bingham, P.M and Zachar, Z. (1985). Evidence that two mutations, w^{DZL} and z^{1}, affecting synapsis-dependent genetic behaviour of *white* are transcriptional regulatory mutants. *Cell* **40**, 819-825.

Bingham, P.M. and Chapman, C. (1986). Evidence that *white-blood* is a novel type of temperature-sensitive mutation resulting from temperature-dependent effects of a transposon insertion on formation of *white* transcripts. *EMBO J.* **5**, 3343-3351.

Bingham, P.M., Chou, T-B., Mims, I. and Zachar, Z. (1988). On/off regulation of gene expression at the level of splicing. *Trends in Genet.* **4**, 134-138.

Birnboim, H.C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmids. *Nucl.Acids Res.* **7**, 1513-1523.

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

Blochlinger, K., Bodmer, R., Jack, J., Yeh Jan, L. and Nung Jan, Y. (1988). Primary structure and expression of a product from *cut*, a locus involved in specifying sensory organ identity in *Drosophila*. *Nature* **333**, 629-635.

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

Bregliano, J-C. and Kidwell, M.G. (1983). Hybrid dysgenesis determinants. In: *Mobile Genetic Elements* (ed. Shapiro, J.A.) Academic Press, New York, pp. 363-410.

Brosseau, G.E., Nicoletti, B., Grell, E.H. and Lindsley, D.L. (1961). Production of altered Y chromosomes bearing specific sections of the X chromosome in *Drosophila*. *Genetics* **46**, 339-346.

Brutlag, D.L. (1980). Molecular arrangement and evolution of heterochromatic DNA. *Ann.Rev.Genet.* **14**, 121-144.

Burke, J.F. (1984). High-sensitivity S1 mapping with single-stranded [P^{32}] DNA probes synthesised from bacteriophage M13mp templates. *Gene* **30**, 63-68.

Burke, J.F. and Ish-Horowicz, D. (1981). Rapid and efficient cosmid cloning. *Nucl.Acids Res.* **9**, 2989-2998.

Busslinger, M., Portmann, R., Irminger, J.C. and Birnstiel, M.L. (1980). Ubiquitous and gene-specific regulatory 5' sequences in a sea urchin histone DNA clone coding for histone protein variants. *Nucl.Acids Res.* **8**, 957-977.

Campuzano, S., Carramolino, L., Cabrera, C.V., Ruiz-Gomez, M., Villares, R., Boronat, A. and Modolell, J. (1985). Molecular genetics of the *achaete-scute* gene complex of *D.melanogaster*. *Cell* **40**, 327-338.

Campuzano, S., Balcells, L., Villares, R., Carramolino, L., Garcia-Alonso, L. and Modolell, J. (1986). Excess function *Hairy-wing* mutations caused by gypsy and copia insertions within structural genes of the *achaete-scute* locus of *Drosophila*. *Cell* **44**, 303-312.

Carbonara, B.D. and Gehring, W.J. (1985). Excision of copia element in a revertant of the *white-apricot* mutation of *Drosophila melanogaster* leaves behind one long terminal repeat. *Mol.Gen.Genet.* **199**, 1-6.

Chang, D-Y., Wisely, B., Huang, S-M. and Voelker, R.A. (1986). Molecular cloning of *suppressor of sable*, a *Drosophila melanogaster* transposon-mediated suppressor. *Mol.Cell.Biol.* **6**, 1520-1528.

Charlesworth, B., Langley, C.H. and Stephan, W. (1986). The evolution of restricted recombination and the accumulation of repeated DNA sequences. *Genetics* **112**, 947-962.

Chia, W., Howes, G., Martin, M., Meng, Y.B., Moses, K. and Tsubota, S. (1986). Molecular analysis of the *yellow* locus of *Drosophila*. *EMBO J.* **5**, 3597-3605.

Chou, T-B., Zachar, Z. and Bingham, P.M. (1987). Developmental expression of a regulatory gene is programmed at the level of splicing. *EMBO J.* **6**, 4095-4104.

Clark-Adams, C. and Winston, F. (1987). The *SPT6* gene is essential for growth and is required for delta-mediated transcription in *Saccharomyces cerevisiae*. *Mol.Cell.Biol.* **7**, 679-686.

Copeland, N.G., Hutchison, K.W. and Jenkins, N.A. (1983). Excision of the DBA ecotropic provirus in *dilute* coat-color revertants of mice occurs by homologous recombination involving the viral LTRs. *Cell* 33, 379-387.

Craymer, L. (1980). Report of L.Craymer. *Dros.Inf.Serv.* 55, 197-200.

Cullen, B.R., Lomedico, P.T. and Ju, G. (1984). Transcriptional interference in avian retroviruses: implications for the promoter insertion model of leukaemogenesis. *Nature* 307, 241-245.

Dawid, I.B., Olong, E.O., Di Nocera, P.P. and Pardue, M.L. (1981). Ribosomal insertion-like elements in *Drosophila melanogaster* are interspersed with mobile sequences. *Cell* 25, 399-408.

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

Doolittle, W.F. and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature*, 284, 601-603.

Dover, G. (1982). Molecular drive: a cohesive mode of species evolution. *Nature* 299, 111-117.

Dudick, M.E., Wright, T.R.F. and Brothers, L-L. (1974). The developmental genetics of the temperature sensitive lethal allele of the *suppressor-of-forked*, $l(1)su(f)^{ts67g}$, in *Drosophila melanogaster*. *Genetics* 76, 487-510.

Duncan, I. (1987). The Bithorax complex. *Ann.Rev.Genet.* 21, 285-319.

Emerman. M and Temin, H.M. (1984). Genes with promoters in retrovirus vectors can be independently suppressed by an epigenetic mechanism. *Cell* 39, 459-467.

Emori, Y., Shiba, T., Kanaya, S., Inouye, S., Yuki, S. and Saigo, K. (1985). The nucleotide sequences of copia and copia-related RNA in *Drosophila* virus-like particles. *Nature* 315, 773-776.

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

Engels, W.R. (1985). A set of P cytotype balancer stocks. *Dros.Inf.Serv.* 61, 70-71.

Engels, W.R. (1988). P elements in *Drosophila*. In: *Mobile DNA* (ed. Berg, D. and Howe, M.) ASM Publications.

Engels, W.R. and Preston, C.R. (1979). Hybrid dysgenesis in *Drosophila melanogaster*: the biology of male and female sterility. *Genetics* 92, 161-175.

Engels, W.R. and Preston, C.R. (1981). Identifying P factors in *Drosophila* by means of chromosome breakage hotspots. *Cell* **26**, 421-428.

Engels, W.R. and Preston, C.R. (1984). Formation of chromosome rearrangements by P factors in *Drosophila*. *Genetics* **107**, 657-678.

Fassler, J.S. and Winston, F. (1988). Isolation and analysis of a novel class of suppressor of Ty insertion mutations in *Saccharomyces cerevisiae*. *Genetics* **118**, 203-212.

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

Finnegan, D.J. and Fawcett, D.H. (1986). Transposable elements in *Drosophila melanogaster*. *Oxford Surveys on Eukaryotic Genes* **3**, 1-62.

Flavell, A.J. (1984). Role of reverse transcription in the generation of extrachromosomal copia mobile genetic elements. *Nature* **310**, 514-516.

Flavell, A.J., Ruby, S.W., Toole, J.J., Roberts, B.E. and Rubin, G.M. (1980). Translation and developmental regulation of RNA encoded by the eukaryotic transposable element copia. *Proc.Nat.Acad.Sci.USA* **77**, 7107-7111.

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

Frischauf, A-M., Lehrach, H., Poustka, A. and Murray, N. (1983). Lambda replacement vectors carrying polylinker sequences. *J.Mol.Biol.* **170**, 827-842.

Gall, J.G. (1973). Repetitive DNA in *Drosophila*. In: *Molecular Cytogenetics*, New York, Plenum Press, pp. 59-74.

Gall, J.G., Cohen, E.H. and Polan, M.L. (1971). Repetitive DNA sequences in *Drosophila*. *Chromosoma (Berl.)* **33**, 319-344.

Gans, M. (1953). Etude genetique et physiologique du mutant *z* de *Drosophila melanogaster*. *Bull.Biol.France Belg.* (Suppl.) **38U**, 1-19.

Garcia-Bellido, A. (1979). Genetic analysis of the *achaete-scute* system of *Drosophila melanogaster*. *Genetics* **91**, 491-520.

Garen, A., Kauvar, L. and Lepesant, J. (1977). Roles of ecdysone in *Drosophila* development. *Proc.Nat.Acad.Sci.USA* **74**, 5099-5103.

Gelbart, W.M. and Wu, C-T. (1982). Interactions of *zeste* mutations with loci exhibiting transvection effects in *Drosophila melanogaster*. *Genetics* **102**, 179-189.

Gergen, J.P., Stern, R.H. and Wensink, P.C. (1979). Filter replicas and permanent collections of recombinant DNA plasmids. *Nucl.Acids Res.* 7, 2115-2136.

Geyer, G., Spana, C. and Corces, V.G, (1986). On the molecular mechanism of gypsy-induced mutations at the *yellow* locus of *Drosophila melanogaster*. *EMBO J.* **5**, 2657-2662.

Goldberg, D.A., Posakony, J.W. and Maniatis, T. (1983). Correct developmental expression of a cloned alcohol dehydrogenase gene transduced into the *Drosophila* germ line. *Cell* **34**, 59-73.

Green, M.M. (1955). Phenotypic variations and psudoalleleism at the *forked* locus in *Drosophila melanogaster*. *Proc.Nat.Acad.Sci.USA* **41**, 375-379.

Green, M.M. (1956). An unstable gene in *Drosophila melanogaster*. *Genetics* **41**, 651.

Green, M.M. (1959a). Spatial and functional properties of pseudoalleles at the *white* locus in *Drosophila melanogaster*. *Heredity* **13**, 303-315.

Green, M.M. (1959b). Putative non-reciprocal crossing over in *Drosophila melanogaster*. *Zeitschrift fur Vererbungslehre* **90**, 375-384.

Gritz, L. and Davies, J. (1983). Plasmid-encoded hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*. *Gene* **25**, 179-188.

Hall, L.M.C., Mason, P.J. and Spierer, P. (1983). Transcripts, genes and bands in 315,000 base pairs of *Drosophila* DNA. *J.Mol.Biol.* **169**, 83-96.

Hansson, L., Lineruth, K. and Lambertsson, A. (1981). Effect of the *l(1)su(f)*^{ts67g} mutation of *Drosophila melanogaster* on glue protein synthesis. *Wilhelm Roux's Arch.* **190**, 308-312.

Hansson, L. and Lambertsson, A. (1983). The role of *su(f)* gene function and ecdysterone in transcription of glue polypeptide mRNAs in *Drosophila melanogaster*. *Mol.Gen.Genet.* **192**, 395-401.

Hansson, L. and Lambertsson, A. (1984). Ecdysterone responsive functions in the mutant *l(1)su(f)*^{ts67g} of *Drosophila melanogaster*. *Wilhelm Roux's Arch.* **193**, 48-51.

Heitz, E. (1928). Das Heterochromatin der Moose. *Jb.wiss Bot.* **69**, 762-818.

Heitz, E. (1929). Heterochromatin, Chromocentren, Chromomeren. *Ber.dtsch.Bot.Ges.* **47**, 274-284.

Heitz, E. (1934). Uber alpha und beta-Heterochromatin sowie Konstanz und Bau der chromomeren bei *Drosophila*. *Biol.Zbl.* **54**, 588-609.

Hoffman, J.A., Lagueux, M., Hetru, C., Charlet, M. and Goltzene, F. (1980). Ecdysone in reproductively competent female adults and embryos of insects. In: Progress in Ecdysone Research (ed. Hoffman, J.A.), *Developments in Endocrinology* **7**. Elsevier/North Holland Biomedical Press, Amsterdam, pp. 431-465.

Hohn, B. and Murray, K. (1977). Packaging recombinant DNA molecules in bacteriophage particles in vitro. *Proc.Nat.Acad.Sci.USA* **74**, 3259-3263.

Holmgren, R. (1984). Cloning sequences from the *hairy* gene of *Drosophila*. *EMBO J.* **3**, 569-573.

Hoyt, M.A., Knight, D.M., Das, A., Miller, H.I. and Echols, H. (1982). Control of phage λ development by stability and synthesis of *cII* protein: role of the viral *cIII* and host *hflA*, *himA* and *himD* genes. *Cell* **31**, 565-573.

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

Ish-Horowicz, D., Howard, K.R., Pinchin, S.M. and Ingham, P.W. (1985). Molecular and genetic analysis of the *hairy* locus in *Drosophila*. *Cold Spring Harbor Symp.Quant.Biol.* **50**, 135-144.

Jack, J.W. (1985). Molecular organisation of the *cut* locus of *Drosophila melanogaster*. *Cell* **42**, 869-876.

Jack, J.W. and Judd, B.H. (1979). Allelic pairing and gene regulation: A model for the *zeste-white* interaction in *Drosophila melanogaster*. *Proc.Nat.Acad.Sci.USA* **76**, 1368-1372.

Jacobson, K.B., Yim, J.J., Grell, E.H. and Wobbe, C.R. (1982). Mechanism of suppression in *Drosophila*: evidence for a macromolecule produced by the $su(s)^+$ locus that inhibits sepiapterin synthase. *Cell* **30**, 817-823.

Jenkins, N.A., Copeland, N.G., Taylor, B.A. and Lee, B.K. (1981). *Dilute* (*d*) coat color mutation of DBA/2J mice is associated with the site of integration of an ecotropic MuLV genome. *Nature* **293**, 370-374.

Jurgens, G. and Gateff, E. (1979). Pattern specification in imaginal discs of *Drosophila melanogaster*. *Wilhelm Roux's Archives* **186**, 1-25.

Kadonaga, J.T., Carner, K.R., Masiarz, F.R. and Tjian, R. (1987). Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* **51**, 1079-1090.

Karess, R. and Rubin, G.M. (1982). A small tandem duplication is responsible for the unstable *white-ivory* mutation in *Drosophila*. *Cell* **30**, 63-69.

Karess, R. and Rubin, G.M. (1984). Analysis of P transposable elements function in *Drosophila*. *Cell* **38**, 135-140.

Karn, J., Brenner, S., Barnett, L. and Cesareni, G. (1980). A novel bacteriophage cloning vector. *Proc.Nat.Acad.Sci.USA* **77**, 5172-5176.

Kaufman, T.C., Tasaka, S.E. and Suzuki, D.T (1973). The interaction of two complex loci, *zeste* and *bithorax* in *Drosophila melanogaster*. *Genetics* **75**, 299-321.

Kidd, S. and Young, M.W. (1986). Transposon-dependent mutant phenotypes at the *Notch* locus of *Drosophila*. *Nature* **323**, 89-91.

Kidwell, M.G. (1985). Hybrid dysgenesis in *Drosophila melanogaster*: nature and inheritance of P element regulation. *Genetics* **111**, 337-350.

Kidwell, M.G. and Novy, J.B. (1979). Hybrid dysgenesis in *Drosophila melanogaster*: sterility resulting from gonadal dysgenesis in the P-M system. *Genetics* **92**, 1127-1140.

Kieny, M.P., Lathe, R. and Lecocq, J.P. (1983). New versatile cloning and sequencing vectors based on bacteriophage M13. *Gene* **26**, 91-99.

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

Klose, W., Gateff, E., Emmerich, H. and Beikirch, H. (1980). Developmental studies on two ecdysone deficient mutants of *Drosophila melanogaster*. *Wilhelm Roux's Arch.* **189**, 57-67.

Klug, A. and Rhodes, D. (1987). "Zinc fingers": a novel protein motif for nucleic acid recognition. *Trends in Biochem.Sci.* **12**, 464-469.

Korge, G. (1977). Larval saliva in *Drosophila melanogaster*: production, composition and relationship to chromosome puffs. *Dev.Biol.* **58**, 339-355.

Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucl.Acids Res.* **12**, 857-872.

Kubli, E. (1986). Molecular mechanisms of suppression in *Drosophila*. *Trends Genet.* **2**, 204-209.

Lambertsson, A. and Fekete, E. (1981). A temperature sensitive suppressor mutation affecting the synthesis of imaginal disc proteins of *Drosophila melanogaster*. *Hereditas* **94**, 135-137.

Laskey, R.A. and Mills, R.D. (1977). Enhanced autoradiographic detection of ^{32}P and ^{125}I using intensifying screens and hypersensitized film. *FEBS Letters* **82**, 314-316.

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

Lees, A.D. and Waddington, C.H. (1942). The development of the bristles in normal and some mutant types of *Drosophila melanogaster*. *Proc.Roy.Soc.(London) Ser.B* **131**, 87-110.

Lefevre, G., Jr. (1974). The relationship between genes and polytene chromosome bands. *Ann.Rev.Genet.* **8**, 51-62.

Lefevre, G., Jr. (1976). A photographic representation and interpretation of the polytene chromosomes of *Drosophila melanogaster* salivary glands. In: *The genetics and Biology of Drosophila* (ed. Ashburner, M. and Novitski, E.) Academic Press, London, pp. 31-66.

Levis, R. and Penman,S. (1978). 5'-terminal structures of poly(A)$^+$ cytoplasmic messenger RNA and of poly(A)$^+$ and poly(A)$^-$ heterogenous nuclear RNA of cells of the Dipteran *Drosophila melanogaster*. *J.Mol.Biol.* **120**, 487-515.

Levis, R. and Rubin, G.M. (1982). The unstable w^{DZL} mutation of *Drosophila* is caused by a 13 kilobase insertion that is imprecisely excised in phenotypic revertants. *Cell* **30**, 543-550.

Levis, R., Bingham, P.M. and Rubin, G.M. (1982). Physical map of the *white* locus of *Drosophila melanogaster*. *Proc.Natl.Acad.Sci.USA* **79**, 564-568.

Levis, R., O'Hare, K. and Rubin G.M. (1984). Effects of transposable element insertions on RNA encoded by the *white* gene of *Drosophila*. *Cell* **38**, 471-481.

Lewis, E.B. (1949). Report of E.B. Lewis. *Dros.Inf.Serv.* **23**, 59-60.

Lewis, E.B. (1956). An unstable gene in *Drosophila melanogaster*. *Genetics* **41**, 651.

Lewis, E.B. (1981). Developmental genetics of the *bithorax* complex in *Drosophila*. In: *Developmental Biology using purified genes*. (ed. Brown, D.D. and Fox, C.F.) Academic Press, New York, pp. 189-208.

Lindsley, D.L. and Grell, E.H. (1968). Genetic variations of *Drosophila melanogaster*. Carnegie Institution of Washington, Publication No.627.

Lindsley, D.L. and Zimm, G. (1985). The Genome of *Drosophila melanogaster*. Part 1: Genes A-K. *Dros.Inf.Serv.* **62**.

Lindsley, D.L. and Zimm, G. (1986). The Genome of *Drosophila melanogaster*. Part 2: Lethals, cytogenetic map. *Dros.Inf.Serv.* **64**.

Lindsley, D.L. and Zimm, G. (1987). The Genome of *Drosophila melanogaster*. Part 3: Rearrangements. *Dros.Inf.Serv.* **65**.

Lifschytz, E. and Falk, R. (1968). Fine structure analysis of a chromosome segment in *Drosophila melanogaster*. Analysis of X-ray-induced lethals. *Mut.Res.* **6**, 235-244.

Lifschytz, E. and Falk, R. (1969). Fine structure analysis of a chromosome segment in *Drosophila melanogaster*. Analysis of ethyl methanesulphonate-induced lethals. *Mut.Res.* **8**, 147-155.

Lifschytz, E. and Green, M.M. (1979). Genetic identification of dominant overproducing mutations: the *Beadex* gene. *Mol.Gen.Genet.* **171**, 153-159.

Lifschytz, E. and Green, M.M. (1984). The *zeste-white* interaction: induction and genetic analysis of a novel class of *zeste* alleles. *EMBO J.* **3**, 999-1002.

Lineruth, K. and Lambertsson, A. (1986). Effect of $l(1)su(f)^{ts67g}$ mutation of *Drosophila melanogaster* on ovarian development. *Hereditas* **104**, 103-109.

McKeown, M., Belote, J.M. and Baker, B.S. (1987). A molecular analysis of *transformer*, a gene in *Drosophila melanogaster* that controls female sexual differentiation. *Cell* **48**, 489-499.

McLachlan, A. (1986). *Drosophila forked* locus. *Mol.Cell.Biol.* **6**, 1-6.

McLachlan, A.D., Staden, R. and Boswell, D.R. (1984). A method for measuring the non-random bias of a codon usage table. *Nucl.Acids Res.* **12**, 9567-9575.

Mandel, M. and Higa, A. (1970). Calcium dependent bacteriophage DNA infection. *J.Mol.Biol.* **53**, 159-162.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). *Molecular cloning: A laboratory manual*. (Cold Spring Harbour, New York. Cold Spring Harbour laboratory).

Mariani, C, Pirrotta, V. and Manet, E. (1985). Isolation and characterisation of the *zeste* locus of *Drosophila*. *EMBO J.* **4**, 2045-2052.

Marlor, R.L., Parkhurst, S.M. and Corces, V.G. (1986). The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. *Mol.Cell.Biol.* **6**, 1129-1134.

Maruyama, T., Gojobori, T., Aota, S. and Ikemura, T. (1986). Codon usage tabulated from the GenBank genetic sequence data. *Nucl.Acids Res.* **14**, r151-r197.

Mattox, W.W. and Davidson, N. (1984). Isolation and characterisation of the *Beadex* locus of *Drosophila melanogaster*: a putative cis-acting negative regulatory element for the *heldup-a* gene. *Mol.Cell.Biol.* **4**, 1343-1353.

Messing, J.H., Crea, R. and Seeburg, P. H. (1981). A system for shotgun DNA sequencing. *Nucl.Acids Res.* **9**, 309-318.

Messing, J. and Vieira, J. (1982). A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* **19**, 269-276.

Meyerowitz, E.M. and Hogness, D.S. (1982). Molecular organisation of a *Drosophila* puff site that responds to ecdysone. *Cell* **28**, 165-176.

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

Miller, J., McLachlan, A.D. and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor III A from *Xenopus* oocytes. *EMBO J.* **4**, 1609-1614.

Modolell, J., Bender, W. and Meselson, M. (1983). *Drosophila melanogaster* mutations suppressible by the *suppressor of Hairy-wing* are insertions of a 7.3-kilobase mobile element. *Proc.Nat.Acad.Sci.USA* **80**, 1678-1682.

Mount, S.M. (1982). A catalogue of splice sequences. *Nucl.Acids Res.* **10**, 459-472.

Mount, S.M., Green, M.M. and Rubin, G.M. (1988). Partial revertants of the transposable element-associated suppressible allele *white-apricot* in *Drosophila melanogaster*: structures and responsiveness to genetic modifiers. *Genetics* **118**, 221-234.

Mozer, B., Marlor, R., Parkhurst, S. and Corces, V. (1984). Characterisation and developmental expression of a *Drosophila ras* oncogene. *Mol.Cell.Biol.* **5**, 885-889.

Murgola, E.J. (1985). Transfer RNA, suppression and the code. *Ann.Rev.Genet.* **19**, 57-80.

Muskavitch, M.A.T. and Hogness, D.S. (1980). Molecular analysis of a gene in a developmentally regulated puff of *Drosophila melanogaster*. *Proc.Nat.Acad.Sci.USA* 77, 7362-7366.

Nash, W.G. and Yarkin, R.J. (1974). Genetic regulation and pattern formation: a study of the *yellow* locus in *Drosophila melanogaster*. *Genet.Res.* 24, 19-26.

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

O'Hare, K., Murphy, C., Levis, R. and Rubin, G.M. (1984). DNA sequence of the *white* locus of *Drosophila melanogaster*. *J.Mol.Biol.* 180, 437-455.

Orgel, L.E. and Crick, F.H.C. (1980). Selfish DNA: the ultimate parasite. *Nature* 284, 604-607.

Osheim, Y.N. and Miller, O.L. (1983). Novel amplification and transcriptional activity of chorion genes in *Drosophila melanogaster* follicle cells. *Cell* 33, 543-553.

Parkhurst, S.M. and Corces, V.G. (1985). *forked*, gypsys, and suppressors in *Drosophila*. *Cell* 41, 429-437.

Parkhurst, S.M. and Corces, V.G. (1986.a). Interactions among the gypsy transposable element and the *yellow* and the *suppressor of Hairy-wing* loci in *Drosophila melanogaster*. *Mol.Cell.Biol.* 6, 47-53.

Parkhurst, S.M. and Corces, V.G. (1986.b). Mutations at the *suppressor of forked* locus increase the accumulation of gypsy-encoded transcripts in *Drosophila melanogaster*. *Mol.Cell.Biol.* 6, 2271-2274.

Parkhurst, S.M. and Corces, V.G. (1986.c). Retroviral elements and suppressor genes in *Drosophila*. *Bioessays* 5, 52-57.

Parkhurst, S.M. and Corces, V.G. (1987). Developmental expression of *Drosophila melanogaster* retrovirus-like transposable elements. *EMBO J.* 6, 419-424.

Peifer, M. and Bender, W. (1986). The *anterobithorax* and *bithorax* mutations of the *bithorax* complex. *EMBO J.* 5, 2293-2303.

Perrimon, N. and Gans, M. (1983). Clonal analysis of the tissue specificity of recessive female-sterile mutations of *Drosophila melanogaster* using a dominant female-sterile mutation, *Fs(1)K1237*. *Dev.Biol.* 100, 365-373.

Pimpinelli, S., Bonaccorsi, S., Gatti, M. and Sandler, L. (1986). The peculiar genetic organisation of *Drosophila* heterochromatin. *Trends in Genet.* 2, 17-20.

Pirrotta, V. and Brockl, C. (1984). Transcription of the *Drosophila white* locus and some of its mutants. *EMBO J.* **3**, 563-568.

Poole, S.J., Kauvar, L.M., Drees, B. and Kornberg, T. (1985). The *engrailed* locus of *Drosophila*: structural analysis of an embryonic transcript. *Cell* **40**, 37-43.

Proudfoot, N.J. and Brownlee, G.G. (1976). 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* **263**, 211-214.

Queen, C. and Korn, L.J. (1984). A comprehensive sequence analysis package for the IBM personal computer. *Nucl.Acids Res.* **12**, 581-600.

Rigby, P.W.J., Diekmann, M., Rhodes, C. and Berg, P. (1977). Labelling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J.Mol.Biol.* **113**, 237-251.

Rimm, D.L., Horness, D., Kucera, J. and Blattner, F. (1980). Construction of coliphage lambda Charon vectors with BamHI cloning sites. *Gene* **12**, 301-309.

Roeder, G.S. and Fink, G.R. (1983). Transposable elements in yeast. In: *Mobile Genetic Elements* (ed. Shapiro, J.A.) Academic Press, New York, pp. 299-328.

Roeder, G.S., Beard, C., Smith, M. and Keranen, S. (1985). Isolation and characterisation of the *SPT2* gene, a negative regulator of Ty-controlled yeast gene expression. *Mol.Cell.Biol.* **5**, 1543-1553.

Rogers, J.E. (1985). The origin and evolution of retroposons. *Int.Rev.Cytol.* **93**, 188-280.

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

Rosenberg, U.B., Schroder, C., Preiss, A., Kienlin, A., Cote, S., Riede, I. and Jackle, H. (1986). Structural homology of the product of the *Drosophila Kruppel* gene with *Xenopus* transcription factor IIIA. *Nature* **319**, 336-339.

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

Rubin, G.M. and Spradling, A.C. (1983). Vectors for P-element mediated gene transfer in *Drosophila*. *Nucl.Acids Res.* **11**, 6341-6351.

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

Rudkin, G.T. (1969). Nonreplicating DNA in *Drosophila. Genetics* (suppl.) **61**, 227-238.

Russell, M.A. (1974). Pattern formation in the imaginal discs of a temperature-sensitive cell-lethal mutant of *Drosophila melanogaster. Dev.Biol.* **40**, 24-39.

Russell, M.A., Girton, J.R. and Morgan, K. (1977). Pattern formation in a ts-cell-lethal mutant of *Drosophila*: the range of phenotypes induced by larval heat treatments. *Wilhelm Roux's Arch.* **183**, 41-59.

Rutledge, B.J., Mortin, M.A., Schwarz, E., Thierry-Mieg, D. and Meselson, M. (1988). Genetic interactions of modifier genes and modifiable alleles in *D.melanogaster. Genetics*, in press.

Saigo, K., Kugimiya, W., Matsuo, Y. Inouye, S., Yoshioka, K., and Yuki, S. (1984). Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster. Nature* **312**, 659-661.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. *Proc.Nat.Acad.Sci.USA.* **74**, 5463-5476.

Scalenghe, F., Turco, E., Edstrom, J.E., Pirrotta, V. and Melli, M.L. (1981). Microdissection and cloning of DNA from a specific region of *Drosophila melanogaster* polytene chromosomes. *Chromosoma (Berl.)* **82**, 205-216.

Schalet, A. (1968). Report of A. Schalet. *Dros.Inf.Serv.* **43**, 64.

Schalet, A. (1972). Report of A. Schalet. *Dros.Inf.Serv.* **49**, 36-37.

Schalet, A.P. (1986). The distribution of and complementation relationships between spontaneous X-linked recessive lethal mutations recovered from crossing long-term laboratory stocks of *Drosophila melanogaster. Mut.Res.* **163**, 115-144.

Schalet, A. and Singer, K. (1971). A revised map of genes in the proximal region of the X-chromosome of *Drosophila melanogaster. Dros.Inf.Serv.* **46**, 131-132.

Schalet, A. and Lefevre, G. (1973). The localisation of "ordinary" sex-linked genes in Section 20 of the polytene X chromosome of *Drosophila melanogaster. Chromosoma (Berl.)* **44**, 183-202.

Schalet, A. and Lefevre, G. (1976). The proximal region of the X chromosome. In: *The Genetics and Biology of Drosophila* (ed. Ashburner, M. and Novitski, E.), **1b**, Academic Press, New York, pp. 847-902.

Scherer, G., Tschudi, C., Perera, J., Delius, H. and Pirrotta, V. (1982). B104, a new dispersed repeated gene family in *Drosophila melanogaster* and its analogies with retroviruses. *J.Mol.Biol.* **157**, 435-452.

Schneuwly, S., Kuroiwa, A. and Gehring, W.J. (1987). Molecular analysis of the dominant homoeotic *Antennapedia* phenotype. *EMBO J.* **6**, 201-206.

Scholnick, S., Morgan, B. and Hirsch, J. (1983). The cloned dopa decarboxylase gene is developmentally regulated when reintegrated into the *Drosophila* genome. *Cell* **34**, 37-45.

Schwartz, H.E, Lockett, T.J. and Young, M.W. (1982). Analysis of transcripts from two families of nomadic DNA. *J.Mol.Biol.* **157**, 49-58.

Searles, L.L. and Voelker, R.A. (1986). Molecular characterisation of the *Drosophila vermilion* locus and its suppressible alleles. *Proc.Nat.Acad.Sci.USA* **83**, 404-408.

Searles, L.L., Jokerst, R.S., Bingham, P.M.,Voelker, R.A. and Greenleaf, A.L. (1983). Molecular cloning of sequences from a *Drosophila* RNA polymerase locus by P element transposon tagging. *Cell* **31**, 585-592.

Shiba, T. and Saigo, K. (1983). Retrovirus-like particles containing RNA homologous to the transposable element copia in *Drosophila melanogaster. Nature* **302**, 119-124.

Simmons, M.J., Raymond, J.D., Johnson, N.A. and Fahey, T.M. (1984). A comparison of mutation rates for specific loci and chromosome regions in dysgenic hybrid males of *Drosophila melanogaster. Genetics* **106**, 85-94.

Singh, L., Purdom, I.F. and Jones, K.W. (1981). Conserved sex-chromosome-associated nucleotide sequences in eukaryotes. *Cold Spring Harb.Symp.Quant.Biol.* **45**, 805-813.

Singh, L., Phillips, C. and Jones, K.W. (1984). The conserved nucleotide sequences of Bkm, which define *Sxr* in the mouse, are transcribed. *Cell* **36**, 111-120.

Snyder, R.D. and Smith, P.D. (1976). The *suppressor-of-forked* mutation in *Drosophila melanogaster*: interactions with the *lozenge* gene. *Biochem.Genet.* **14**, 611-617.

Soberon, X., Covarrubias, L., and Bolivar, F. (1980). Construction and characterisation of new cloning vehicles. IV. Deletion derivatives of pBR322 and pBR325. *Gene* **9**, 287-305.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J.Mol.Biol.* **98**, 503-517.

Spradling, A.C. (1981). The organisation and amplification of two chromosomal domains containing *Drosophila* chorion genes. *Cell* **27**, 193-201.

Spradling, A.C. and Rubin, G.M. (1981). *Drosophila* genome organisation: conserved and dynamic aspects. *Ann.Rev.Genet.* **15**, 219-264.

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

Spradling, A.C. and Rubin, G.M. (1983). The effect of chromosomal position on the expression of the *Drosophila* xanthine dehydrogenase gene. *Cell* **34**, 47-57.

Spradling, A.C. (1986). P element mediated transformation. In *Drosophila - a practical approach* (ed. Roberts, D.B.) IRL press, Oxford, pp. 175-198.

Staden, R. (1982). An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucl.Acids Res.* **10**, 2951-2961.

Staden, R. (1984). Computer methods to locate signals in nucleic acids. *Nucl.Acids Res.* **12**, 505-520.

Staden, R. (1984). Measurements of the efffects that coding for a protein has on a DNA sequence and their use for finding genes. *Nucl.Acids Res.* **12**, 551-567.

Suzuki, D.T., Kaufman, T., Falk, D. and the U.B.C. Drosophila Research Group (1976). Conditionally expressed mutations in *Drosophila melanogaster*. In: *The Genetics and Biology of Drosophila* (ed. Ashburner, M. and Novitski, E.) **1a**, pp. 207-263.

Sweet, H.O. (1983). *Dilute suppressor*, a new suppressor gene in the house mouse. *J.Hered.* 74, 304-306.

Tartof, K.D. (1973). Regulation of ribosomal RNA gene multiplicity in *Drosophila melanogaster*. *Genetics* 73, 57-71.

Tautz, D., Lehmann, R., Schnurch, H., Schuh, R., Seifert, E., Kienlin, A., Jones, K. and Jackle, H. (1987). Finger protein genes of novel structure encoded by *hunchback*, a second member of the gap class of *Drosophila* segmentation genes. *Nature* 327, 383-389.

Temin, H.M. (1980). Origin of retroviruses from cellular moveable genetic elements. *Cell* 21, 599-600.

Temin, H.M. and Engels, W. (1984). Moveable genetic elements and evolution. In: *Evolutionary theory - paths into the future*, (ed. Pollard, J.W.) John Wiley and Sons, pp. 173-201.

Varmus, H.E. (1983). Retroviruses. In: *Mobile Genetic Elements* (ed. Shapiro, J.A.). Academic Press, Orlando, pp. 411-503.

Vincent, A., Colot, H.V. and Rosbash, M. (1985). Sequence and structure of the *serendipity* locus of *Drosophila melanogaster*: a densely transcribed region including a blastoderm-specific gene. *J.Mol.Biol.* **186**, 149-166.

Voss, R. and Falk, R. (1973). The nature of reverse mutations in *Drosophila melanogaster*. *Mut.Res.* **20**, 221-234.

Walker, A., Howells, A.J. and Tearle, R.G. (1986). Cloning and characterisation of the *vermilion* gene of *Drosophila melanogaster*. *Mol.Gen.Genet.* **202**, 102-107.

Watson, M.E.E. (1984). Compilation of published signal sequences. *Nucl.Acids Res.* **12**, 5145-5164.

Whittinghill, M. (1937). New Mutants: Report of M. Whittinghill. *Dros.Inf.Serv.* **8**, 11.

Whittinghill, M. (1938). The induction of oogonial crossing over in *Drosophila melanogaster*. *Genetics* **23**, 300-306.

Wilson, T.G. (1980). Studies on the female-sterile phenotype of $l(1)su(f)^{ts76a}$, a temperature-sensitive allele of the *suppressor-of-forked* mutation in *Drosophila melanogaster*. *J.Embryol.Exp.Morph.* **55**, 247-256.

Winston, F., Chaleff, D.T., Valent, B. and Fink, G.R. (1984a). Mutations affecting Ty-mediated expression of the *HIS4* gene of *Saccharomyces cerevisiae*. *Genetics* **107**, 179-197.

Winston, F., Durbin, K.J. and Fink, G.R. (1984b). The *SPT3* gene is required for normal transcription of Ty elements in *S.cerevisiae*. *Cell* **39**, 675-682.

Winston, F. and Minehart, P.L. (1986). Analysis of the yeast *SPT3* gene and identification of its product, a positive regulator of Ty transcription. *Nucl.Acids Res.* **14**, 6885-6900.

Wood, W.B. (1966). Host specificity of DNA produced by *Escherischia coli*: bacterial mutations affecting the restriction and modification of DNA. *J.Mol.Biol.*, **16**, 118-133.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-119.

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

Young, M.W. (1979). Middle repetitive DNA: a fluid component of the *Drosophila* genome. *Proc.Nat.Acad.Sci USA* 76, 6274-6278.

Zachar, Z. and Bingham, P.M. (1982). Regulation of *white* locus expression: the structure of mutant alleles at the *white* locus of *Drosophila melanogaster*. *Cell* 30, 529-541.

Zachar, Z., Davison, D., Garza, D. and Bingham, P.M. (1985). A detailed developmental and structural study of the transcriptional effects of insertion of the copia transposon into the *white* locus of *Drosophila melanogaster*. *Genetics* 111, 495-515.

Zachar, Z., Garza, D., Chou, T-B, Goland, J. and Bingham, P.M. (1987a). Molecular cloning and genetic analysis of the *suppressor-of-white-apricot* locus from *Drosophila melanogaster*. *Mol.Cell.Biol.* 7, 2498-2505.

Zachar, Z., Chou, T-B. and Bingham, P.M. (1987b). Evidence that a regulatory gene autoregulates splicing of its transcript. *EMBO J.* 6, 4105-4111.