Transcripts From the Head of *Drosophila melanogaster*.

A thesis submitted for the degree of
Doctor of Philosophy at the University of Glasgow.

by

Colin D. Milligan

Institute of Genetics,
University of Glasgow,
Church Street,
Glasgow, G11 5JS

August 1995.

The research reported within this thesis is my own work, except where otherwise stated and has not been submitted for any other degree.

For Dad and for Cal.

Table of Contents

List of Figures

Chapter One

Chapter Three

Chapter Four

Chapter Five

Chapter Six

Chapter Seven

Chapter Eight

Abbreviations.

Amp	- ampicillin
ATP	- adenosine triphosphate
b-NAD	- b-nicotamide adenine dinucleotide
bp	- base pairs
BSA	- bovine serum albumin
cDNA	- complementary DNA
CIAP	- calf intestinal alkaline phosphatase
cpm	- counts per minute
DAB	- diaminobenzidine
DEPC	- diethylpyrocarbonate
DNA	- deoxyribonucleic acid
dNTP	- deoxyribonuleotide triphosphate
DTT	- dithiothreitol
EDTA	- ethylendiamine tetra-acetic acid
EGTA	- ethylene glycol bis (b-amino ethyl ether) N-N-N'-N'-tetra-acetic acid.
EMS	- ethyl methyl sulfonate
EtBr	- ethidium bromide
GCG	- genetics computing group (University of Wisconsin)
GSP	- gene specific primer
HAP	- hydroxyapatite
HB	- homogenisation buffer
IPTG	- isopropyl-b-D-thiogalactopyranoside
kb	- kilobases/kilobasepairs
kD	- kilodaltons
L1, L2, L3	- 1st, 2nd, 3rd instar larvae
LMP	- low melting point
M	- molar
MOPS	- 3-n-morpholinopropane sulfonic acid
mRNA	- messenger ribonucleic acid
MuMLVRT	- murine moloney leukaemia virus reverse transcriptase

$NaPP_i$	- sodium pyrophosphate
NLB	- nuclear lysis buffer
nt	- nucleotide
OD	- optical density
PCP	- phosphate carrier protein
PCR	- polymerase chain reaction
PEV	- position effect variegation
pfu	- plaque forming units
PIPES	- piperazine - N,N' bis 2-ethane sulfonic acid
PVP	- polyvinylpyrrolidone
RF	- RNase free
RT	- reverse transcriptase
rpm	- revolutions per minute
SDS	- sodium dodecylsulphate
SSC	- saline sodium citrate
SSPE	- saline sodium phosphate EDTA
TEMED	- N, N, N' N', -tetramethylenediamine
TAE	- tris acetate EDTA
TBE	- tris borate EDTA
TE	- tris-EDTA
Tet	- tetracycline hydrochloride
Tm	- melting temperature
Tris	- tris (hydroxylmethyl) amino ethane
TSP	- transposon specific primer
u	- units
µCi	- microcuries
UV	- ultraviolet light
w:v	- weight : volume
v:v	- volume : volume
V	- Volts
WT	- wild type *Drosophila* strain
X-gal	- 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside

Abstract

Drosophila is an ideal organism in which to study the genetics of the nervous system. Reverse Genetics technologies, coupled with powerful traditional techniques make a speculative approach to *Drosophila* neurogenetics very attractive.

This study used differential screening procedures to isolate new genes expressed in the *Drosophila* nervous system; selecting against those whose expression is specific to the eye. This approach might be successful in identifying new genes which have hitherto been overlooked by traditional approaches.

In this study, clones representing four new *Drosophila* genes were characterised. A gene encoding an esterase has been isolated; corresponding to a previously recognised (but uncloned) biochemical variant in *Drosophila*. The gene encoding the *Drosophila* mitochondrial Phosphate Carrier Protein was isolated. Although distributed throughout the body, this gene was previously uncloned and therefore merited further characterisation. Two of the clones isolated in this study represent genes of unknown function. One gene is particularly interesting. Although restricted to the nervous system, its expression in this tissue is very high, making it somewhat surprising that this gene is not already known.

A reverse genetic approach, allowing efficient isolation of new retrotransposon induced mutations in selected genes was also developed. Strategies such as this may be used to elucidate the function of the unknown genes and further dissect the *in vivo* role of the two genes whose function is already known.

Chapter One

Introduction

1.1 Studying Neural Processes

The brain of any higher eukaryote is its most complex organ; controlling a myriad of processes from physiological constancy, through movement, olfaction, gustation, vision, hearing, and proprioception to the more complex processes underlying consciousness. For a biologist, to unravel its innermost workings is a great challenge. Using traditional investigative tools to examine anatomy and physiology, we have learned much about many basic neural processes. Such approaches have been augmented by behavioural studies like those examining social behaviour in the honeybee or simple reflex conditioning in the sea snail *Aplysia* (Abrams & Kandel, 1988). Together this range of approaches has taught us much about how animals sense and interact with their environment and with other members of their community.

Whilst the cellular basis for most neural processes yields to investigation, it is evident that the molecular and genetic basis of virtually all of these processes is still unknown; even in *Drosophila melanogaster*, the ideal organism in which to study genetics. Why is this the case? Genes controlling neural processes are often very complex in their organisation and control. Mutations associated with such genes may be pleiotropic or obscured by epistatic interactions. Alternatively, the phenotypes produced may prove relatively subtle; difficult to score simply by scanning a population, as one would do for a typical morphological mutation (for instance, the collection of I factor induced mutations at *white*, Sang *et al.*, 1984). Such intangible phenotypes reflect the organisational and functional complexity of the nervous system. Faced with this complexity, many behavioural scientists as well as geneticists and biochemists have chosen to treat the nervous system as a black box; with little regard for the mechanisms underlying the processing of inputs and generation of outputs. The brain is a complex organ, integrating the function of many independent subsystems, all with a unique genetic makeup. If we are to fully unravel the workings of organs such as the brain then it is their genetic basis that we must study. In this way, we may look at the control and specification of processes, as well as the molecules which participate in them. Studying molecular genetics gives us a framework in which to position our biochemical and physiological knowledge, allowing us to integrate all our findings into a complete picture. In addition molecular analysis of DNA and protein sequences

(Kimura, 1980) provides data which allows investigation of the evolution of processes and systems; helping us to a greater understanding of these more abstract problems.

1.2 Genetics: *Drosophila* as a Model

In *Drosophila*, the intractability of the nervous system to efficient genetic dissection has meant that neurogenetics as a discipline now lags some way behind the genetic studies of other systems like body plan development (Nusslein-Vollhard & Weischaus, 1980) or sex determination (Baker *et al.*, 1987), whose associated loci exhibit more obvious mutant phenotypes. Even so, the flexibility of the fruit fly as an experimental organism, its well-studied neurobiology and excellent genetics has made *Drosophila* a favourite organism of neurogeneticists for almost 30 years now. Studies started in the late 1960's by Seymour Benzer at CalTech have seen significant progress made in our understanding of the specification and functioning of the nervous system at the genetic level. Molecular studies are beginning to enhance our understanding of many processes. We can begin to open the black box and ask questions not just about the genes transcribed within a cell, the brain's physiology or an organism's behaviour, but about the relationship between them.

Benzer's ultimate goal was to investigate the relationship between the information contained within the genome and the production of a living, 'behaving' organism. All the information is contained within the genome, therefore we may study this relationship by characterisation of genes. Eventually this primary information can be integrated to help us understand this relationship.

Although traditional genetic approaches have managed to shed some light on the functioning of the nervous system (e.g. Hotta & Benzer, 1970), going from phenotype to gene is on the whole a relatively blind process with no real rewards until the whole process is elucidated (and the gene cloned). This inefficiency has tempted molecular geneticists to try and study these loci by relatively unconventional means: selecting genes for study not on the basis of the phenotype they exhibit when disrupted, but on their temporal or spatial pattern

of expression and on other features recognised initially at the molecular and not the organismal level. Some investigators have used sequence homologies to clone genes on the basis of the predicted biochemical role of the proteins they encode, or have used expression libraries to search for spatially or temporally restricted peptide motifs. In recent years such 'reverse genetic' approaches have been of great use in widening our knowledge of biological processes such as phototransduction (e.g. Shieh *et al.*, 1989; Fryxell & Meyerowitz, 1987). Together, these approaches provide a fully fledged strategy for studying the biology of a given organism.

The short generation time of *Drosophila* facilitates the collection of the vast quantities of tissue necessary for some of these molecular approaches. In addition, whilst ethical and practical considerations discourage the routine use of mammals as model systems, *Drosophila* has proven an adequate model system for many simple behaviours. Whilst the anatomy of the nervous system may have diverged significantly from mammals, its components and processes have been retained. Furthermore, evolutionary conservation of the proteins involved in these processes has been great. Amongst many multigene families, as much divergence is found within a single organism as between equivalent members in long diverged species such as *Drosophila* and mammals. For instance, a *kruppel* homolog is found in mice, which bears more resemblance to the *Drosophila kruppel* gene product than it does to other murine Zinc finger genes (Schuh *et al.*, 1986). In addition, Miller & Benzer (1983) showed that a large number of monoclonal antibodies raised against *Drosophila* nervous system tissue cross react with human neural tissue. This finding indicates that a significant amount of evolutionary conservation is present within the molecules of the central nervous system throughout eukaryotes.

1.3 Early *Drosophila* Neurogenetics

At the end of the 1960's, the elucidation of the genetic code widened the horizons of the genetic community. Although the tools were not yet available, it was already recognised that gene sequences might be determined and related to gene function and phenotype. An ability to correlate the information contained within the genome with genetic observations gave the study of genetics a new

impetus. With this in mind, biologists like Seymour Benzer began to turn their attention to phenomena and systems which, due to their apparent complexity had hitherto seemed completely refractile to genetic dissection. Benzer himself concentrated on studying behavioural processes (Benzer, 1967), working on the proviso that by inducing mutations in individual genes, the genetic basis of a whole range of processes might be opened up (Benzer, 1971). His chosen model - *Drosophila melanogaster* - reflected his appreciation of the importance of powerful and amenable genetics in these studies. The early work of Benzer and his colleagues was based upon the assumption that the processes of the nervous system would yield to genetic dissection. Complex processes are merely the result of the interaction of sets of single genes. Studying these genes would ultimately allow complete dissection of these pathways.

Initially, Benzer and his colleagues developed a variety of simple behavioural paradigms and assembled a collection of mutant strains based on variations in locomotor activity, courtship success, and response to stress (see Benzer, 1973 for a review). Having isolated mutants, they realised that the development of new genetic tools would be vital to the success of their studies. The first of these was mosaic analysis (Hotta & Benzer, 1970). A genetic cross was designed to generate gynandomorph individuals, whose somatic tissues were a mixture of male and female cells. Males carrying mutant alleles of X-linked loci were crossed to females bearing an abnormal X-chromosome. This X-chromosome (a ring-X) is inherently unstable and frequently lost during early development. In an adult female, clones of cells descended from cells which have lost the ring X-chromosome would display a male (and mutant) genotype. These male cells express recessive X-linked copies of the gene under study along with various X-chromosome markers. Genotypically female cells carry wild type alleles (on the ring-X chromosome) and are phenotypically normal. A population of flies would possess many different clones of mutant cells. By comparing the phenotypes induced to the clones of cells affected, the site of action of various X-linked loci could be localised. This technique has been of tremendous use over the last thirty years and controlling centres of circadian rhythm and male courtship behaviour have each been localised in this way (Konopka *et al.*, 1983; Hall, 1979).

Still however, the selection of mutants in many neural processes was impossible and throughout the early 1970's, various paradigms were devised which allowed more deep seated behaviours such as associative learning to be examined. In 1974, Quinn *et al.*, suggested that given the right tests, *Drosophila* could learn. Therefore, flies that failed to learn could be collected as learning mutants. The earliest (and most enduring) learning paradigm conditioned flies to associate a specific odour with an electric shock, thereby training them to migrate away from that odour. Quinn and his co-workers were careful to eliminate the influence of non-genetic factors like environmental conditions and odour preference. Within a few years (Quinn & Dudai, 1976), more information into how *Drosophila melanogaster* learned and remembered was gained and single gene defects in both learning (*dunce*; Dudai *et al.*, 1976) and 'remembering' (*amnesiac;* Quinn *et al.*, 1979) were isolated. Since then, the information gained about these two mutants illustrates well, the range of success that these traditional approaches have achieved and how a molecular understanding allied to genetic data can accelerate our knowledge of a particular process. Sixteen years after its identification, the *amnesiac* gene product has only just been cloned (Feany & Quinn, 1995) and we have learned little about the biochemistry of 'remembering'. In contrast, the *dunce* gene has been cloned for several years (by Chen *et al.*, 1986; though not because of its learning phenotype), characterised and sequenced, the product identified and the pathway into which it fits largely elucidated. The *dunce* gene product was cloned by chromosome walking and encodes a common phosphodiesterase, thought to be involved in intracellular signalling by modulating levels of cyclic AMP within the cell.

The cloning of the *dunce* locus confirmed earlier biochemical observations that had shown *dunce* flies to be lacking this type of enzymatic activity. Other enzymes and proteins biochemically linked to the *dunce* phosphodiesterase have now been found to be associated with the genetic lesions responsible for other 'learning' mutants such as *rutabaga* (Livingstone *et al.*, 1984; cloned by Levin *et al.*, 1992). The insight gained from our molecular studies of the *Drosophila* learning genes has allowed scientists to formulate coherent hypotheses as to the biochemistry underlying learning, thus accelerating our progress towards a full understanding of these higher processes within the fly brain. Other researchers have embarked upon mutational analyses of other components of

the second messenger pathway in which the *dunce* product participates. In the case of the type I regulatory subunit of Protein Kinase A (Goodwin *et al.*, submitted), mutations were induced with phenotypes similar, (though not identical) to the learning defects observed in *dunce* flies. These results further define the biochemical basis of memory and learning in the fruit fly. Before its cloning, *amnesiac* was thought to play a related role in the retrieval of information from memory. The gene was eventually cloned after a P-element insertion at the *amnesiac* locus was obtained. This P-element insertion was used as a molecular marker to obtain first genomic DNA and subsequently cDNA clones corresponding to the *amnesiac* locus. The transcript from this locus encoded a neuropeptide of 148 amino acids, related to the mammalian Pituitary Adenylate Cyclase Activating Peptide (PACAP; Hosoya *et al.*, 1992). The role of PACAP (and presumably the *amnesiac* product) is to activate adenylate cyclase and raise levels of cAMP in the cell. This finding further supports the hypothesis that cyclic AMP metabolism is central to the processes underlying *Drosophila* learning and memory.

1.4 Molecular Analysis

Due to its great complexity, much is still to be learned about the molecular basis of learning. Only the X-chromosome has been screened for learning genes in any great detail. We can only assume that there will be as many genes on the second and third chromosomes which have yet to be screened to the same extent as the X-chromosome. A few genes have been identified using P-element tagging (e.g. *latheo*, Boynton & Tully, 1992; *linotte*, Dura *et al.*, 1993). Even so, the number of genes characterised so far is certainly insufficient to specify and control the systems known to be present. Learning is just one of many behaviours exhibited by the adult fly. Molecular techniques will play a significant part in furthering our understanding of this and many other behaviours and processes.

In the field of *Drosophila* neurobiology, greater success has been achieved with more discrete phenomena such as the more severe morphological phenotypes (often embryonic lethal) associated with developmentally important neurogenic loci like *mastermind* and *Notch*. The neurogenic loci and their gene products specify differentiation of neural cells and organise neural development. Severe

and lethal phenotypes have facilitated the cloning of some of these genes (e.g. *Notch*, Artavanis-Tsakonas *et al.*, 1983) by chromosome walking from known chromosomal breakpoints. Cloned genes like *Notch* have been studied extensively and much has been learned about their action using molecular findings (gene structure and sequence, RNA localisation studies etc.) as a starting point. Correlating the genetic data with molecular data reinforces physiological and biochemical observations and prompts further detailed studies of the processes controlled by *Notch*. In *Notch* at least, the mass of genetic information available prior to cloning gave no real indication as to the precise function of the *Notch* gene product. Phenotypes associated with different alleles ranged from severe embryonic lethals to subtle defects in wing and eye morphology. Complex interactions between different groups of alleles further confused the overall picture. Once cloned and sequenced, it became immediately apparent that the *Notch* product encoded a putative transmembrane protein (a prediction borne out by immunochemical data; cited in Artavanis-Tsakonas, 1988) with both extracellular and intracellular domains. Comparison with other sequences revealed similarities in nucleotide and amino acid sequences to the mammalian Epidermal Growth Factor (EGF) and Low Density Lipoprotein Receptor (Wharton *et al.*, 1985) indicating that the *Notch* product may mediate cell-cell interactions.

All this has served to allow Artavanis-Tsakonas and his colleagues to put forward a rather coherent model of the function of the *Notch* product - based primarily on molecular data. Although the knowledge of *Notch* as a locus involved in neurogenesis was helpful, *in situ* hybridisation studies showing its spatial and temporal expression pattern (e.g. Hartley *et al.*, 1987) have implicated this anyway, and have further indicated that the effects of *Notch* are complex and far reaching: *Notch* product affects a great number of processes indirectly as opposed to a small number directly. Now, the studies of *Notch* concentrate on these wider questions; molecular approaches having answered many of the basic questions some time ago.

The molecular characterisation of *Notch* is a good example of the wealth of knowledge that can be gained quickly if cloned material is available. In the complex nervous system, a molecular approach may speedily sort out early

problems, providing a language with which to address and answer biological questions. Primary data is of much greater value than observation when studying a whole system as what is present 'on the surface' often belies the complexity of the whole. Without molecular analysis, we can never hope to understand loci such as *amnesiac*, whose phenotypes are subtle and lend no direct clues as to the function or identity of any gene product. Purely genetic tools (for instance, those mutants which might exhibit interacting phenotypes) can't be used because we can not direct our research without having a good idea of what we expect the gene product to be.

1.5 Purely Molecular Approaches

For these reasons, selecting cloned material on the basis of spatial or developmental expression patterns provides a convenient if not perfect way of exploring those genetic loci that have not yielded to conventional approaches. As the molecular techniques available to geneticists increase, so a molecular approach becomes more attractive. Since the beginning of the 1980s, more and more researchers have begun to use these 'reverse genetic' approaches.

Over the past fifteen years, various molecular approaches have been used to isolate new genes from the *Drosophila* nervous system. Some of these are based upon DNA or protein sequence homology, whilst others are based upon protein expression patterns. Most (like the current study) are based on RNA transcription patterns. Recently, some studies have attempted to use marked transposable elements first to clone genes dependent upon the pattern of expression of a *lacZ* reporter construct, and then to study them using traditional genetic analysis alongside modern molecular techniques.

During studies in the early 1980s aimed at estimating the number of genes in the *Drosophila* genome, Levy and co-workers performed a number of experiments in which they used RNA-DNA reassociation kinetics to investigate the complexity of RNA populations from various stages of the life cycle of *Drosophila melanogaster*. Levy & Manning (1981) reported that approximately 16,000 distinct RNA species were present in the adult fruit fly, with around 70% (11,000) of these present in heads. It was further shown (in accordance with

earlier results (Levy & McCarthy, 1975)) that approximately one third of these sequences were polyadenylated. Addressing the question 'What portion of the RNA complement is shared and how much is unique to a particular tissue or stage?', Levy & Manning argued that the diversity of two tissues may reflect quantitative as well as qualitative changes in the RNA species present, as suggested by Beissmann (1981). This suggests that the adult head of *Drosophila* might contain a relatively small number of unique RNA species, gaining most of its individual characteristics from qualitative differences in RNA species which are not themselves restricted to the head. One must remember that the *Drosophila* head contains ~50% of the neural tissue in the fly in 10% of the total tissue (Demerec, 1950). This means that anything which makes a cell 'neural' will not be found restricted to the *Drosophila* head. Only those genes which specify or control greatly specialised (i.e. neither housekeeping nor neural) processes will have a pattern of transcription restricted to the head.

1.6 'Reverse Genetic' Strategies

In the last fifteen years, many groups of researchers have tried to use molecular techniques to clone and study transcripts expressed in the *Drosophila* nervous system. Below is a summary of the different 'reverse genetic' strategies which have been adopted, considering their success or failure and their relevance to the present study. First however, novel approaches which have not relied on RNA expression patterns will be briefly discussed.

1.6.1 Cloning by Homology

Of all the approaches used, the most directed are those which rely on homology to facilitate cloning of homologues to genes found initially in other species or to find new members of a gene family from within the same organism. Although this approach has been very successful in isolating members of large families like the 'homeobox' containing genes, it does not allow cloning on the basis of expression pattern. All members are available for cloning, not just those from the tissue under study. This strategy is only appropriate if you are looking for a specific gene (e.g. the *Drosophila* Cam Kinase II gene (Ohsako *et al.*, 1993)) or a specific type of gene (Zinc finger or homeobox domain containing genes). 'New' genes can't be cloned. In such studies, cDNA libraries may be screened at reduced stringency with the heterologous probe. Once isolated, 'positives' might

be grouped and studied according to the strength of signal they give, a measure of the degree of similarity between the probe and target sequences.

Alternatively, the Polymerase Chain Reaction (PCR, Saiki *et al.*, 1988) may be used; oligonucleotide primers may be made and first strand cDNA or genomic DNA may be used as a template in an attempt to amplify the sequence of interest (e.g. Rasmusson *et al.*, 1994). The amplification products can be characterised directly or used to screen a cDNA library at high stringency. This emerging technique has a number of advantages and has met with considerable success. Primers can be chosen when only protein sequence is known, utilising codon usage data to intelligently design optimal primer sequences. Theoretically, amplification can be achieved from a single copy of the appropriate template sequence. Therefore, this technique is efficient even when the amount of starting material is minimal. Furthermore, if the site of expression is known (or can be guessed) then that tissue can be used as an enriched template for PCR. The use of PCR as a means to identify new genes on the basis of homology to existing sequences was developed by Kamb *et al.*, (1989), who were searching for genes controlling pattern formation in the nematode *Caenorhabditis elegans* using previously cloned *Drosophila* homeobox sequences. This PCR approach is designed to avoid many of the problems associated with conventional low stringency screens where spurious hybridisation might hamper efficient progress towards cloning the desired gene(s) and where extensive restriction mapping and re-screening might be required before the gene of interest is finally cloned.

The cloning of the *Drosophila* Phospholipase C gene (PI-PLC21) by Pak and colleagues (Shortridge *et al.*, 1991) illustrates the success of a typical use of low stringency hybridisation to clone a new member of a protein family. Phosphatidyl-Inositol specific Phospholipase C is thought to be an important component for a range of signal transduction pathways throughout the nervous system, including phototransduction. A previously cloned PI-PLC, encoded by the *norpA* gene of *Drosophila* was shown by Bloomquist *et al.*, (1988) to be essential for phototransduction. Flies lacking a functional copy were blind and yet had no visible anatomical defects, signifying a biochemical rather than a physical defect. Using the *norpA* cDNA, a genomic library was screened at low

stringency, identifying seven classes of hybridising clones. One of these classes was chosen for further study because members hybridised independently to two conserved regions of the *norpA* cDNA. Using the genomic clone, a cDNA library was screened at high stringency. cDNA clones were isolated and (after sequencing) divided into two classes. Sequence analysis confirmed that these two related cDNAs did in fact encode a PI-PLC homologue, each sharing 32% identity to the *norpA* gene. *in situ* hybridisation indicated expression throughout adult and larval bodies. Northern analysis confirmed this data, additionally indicating the presence of two transcripts, one expressed throughout development and a second expressed only in adult heads. *in situ* hybridisation experiments also indicated that PI-PLC21 co-localises with a G-protein G_o, of previously unknown origin. The authors propose that these two proteins might work together in the same signalling pathway. As yet without mutants, these hypotheses are difficult to test.

Using an antisense bovine rhodopsin RNA, Zuker *et al.*, (1985) were able to isolate the RNA encoding the major *Drosophila* opsin (the product of the *ninaE* locus) found in the R1-R6 photoreceptors. They opted to isolate RNA-RNA hybrids because conventional low stringency hybridisation experiments (like the strategy used to isolate PI-PLC21) had been unsuccessful. RNA-RNA hybrids are more stable, and Zuker *et al.* found that by increasing the hybridisation stringency, conditions could be found in which only a single RNA species was identified. First strand cDNA made from the *Drosophila* RNA was then synthesised and used to screen a genomic library. This in turn led to the isolation of a cDNA, which when sequenced was found to encode a polypeptide which shared 22% identity with the original bovine polypeptide sequence.

1.6.2 Cloning Using Antibody Epitopes

Cloning by homology and cloning from protein sequences are obvious methods to use when a directed approach is possible. Of more interest here are the various speculative approaches which have been followed in recent years. Seymour Benzer's initial work in this field began in 1982 (Fujita *et al.*, 1982). Monoclonal antibodies were generated to homogenates of adult *Drosophila* brain tissue after its injection into mice. The distribution within the head of a portion was examined using immunohistochemistry. Fujita and colleagues found a

variety of expression patterns ranging from seemingly ubiquitous to 'neural' restricted. They also observed patterns of expression which confirmed that anatomically distinct regions of the fly brain are also antigenically (and therefore functionally?) distinct.

One of the monoclonal antibodies which Fujita *et al.* generated seemed to be restricted to photoreceptor neurons. Over the next three years, Benzer's group (Zipursky *et al.*, 1985) employed molecular techniques to progress from immunohistochemical phenotype to cloned gene. Initially, Zipursky *et al.* isolated enough protein to microsequence the amino terminus of this peptide. They were subsequently able to predict possible and probable DNA sequences for this short region (of 60 bases). Oligonucleotides were synthesised on the basis of these predictions and used to screen a bacteriophage λ genomic DNA library at low stringency (this was prior to the advent of PCR). A genomic lambda clone was isolated and partially sequenced. Once it was confirmed that the DNA sequence matched that predicted from the peptide sequence, the genomic clone was used as a probe to screen a cDNA library. Eventually, in 1988, Reinke *et al.* were able to present a cDNA sequence for this gene, along with analysis of its function and (in an accompanying paper (Van Vactor *et al.*, 1988)) molecular analysis of two mutants within this gene, *chaoptin*. The *chaoptin* gene encodes a large transmembrane protein which has homology to proteins previously cloned in yeast and humans. Comparative immunohistochemical studies of normal and mutant flies allowed hypotheses on the role of the *chaoptin* gene product to be formulated and confirmed. A major factor in the success of this research was the abundance of the initial protein in the *Drosophila* retina. Without such high levels of expression, it would have been difficult to obtain enough material for the microsequencing which got this research started. Despite the large amount of time and effort required (six years from antibody staining pattern to gene and function), the wealth of data eventually generated and the insight gained about the role of this protein in mediating cell adhesion during cellular morphogenesis in the development of the compound eye attests to the value of a molecular approach.

1.6.3 Cloning by Expression Pattern

The approaches described above have each contributed to our understanding of the mechanisms of the *Drosophila* nervous system. Most of the speculative molecular screens which have been carried out in the last fourteen years have investigated RNA expression patterns as a means of identifying new head specific or head elevated genes. Like Benzer's screen, they might be directed to a specific type of gene (e.g. one found only in photoreceptor cell types) but not restricted to genes (or types of genes) found previously elsewhere. Here is a brief summary of the approaches which have been taken, along with examples of the genes which they have successfully isolated and an evaluation of their relative merits. As we will see, the screens carried out in this way have tended to yield genes specific to or expressed predominantly within the eye. This has suited researchers like Zuker and Benzer who had chosen the compound eye as a convenient model for neurogenetic research. This is so for a number of reasons: there are many mutants available, the structure of the eye is relatively simple (allowing neurophysiological and anatomical investigation) and most importantly, the eye is an ideal interface between the fly brain and its environment.

Accompanying the work on RNA complexity in *Drosophila*, Levy generated a number of genomic clones containing 'genes' expressed only in the head (Levy *et al.*, 1982; Levy & Manning, 1982). These clones, selected by a differential screening procedure (carried out on a wild type genomic DNA library using wild type head and body mRNA as the source material for the cDNA probes) were further characterised in a number of laboratories. The results of these studies have been published in a number of papers (Fryxell & Meyerowitz, 1987, Ray & Ganguly 1992, Swanson & Ganguly, 1992 and Montell *et al.*, 1985). On investigation, many of the genes selected in this differential screen were found to be expressed wholly or predominantly in the eye. Fryxell & Meyerowitz (1987) took the seventeen independent bacteriophage clones generated by Levy *et al.* (1982) and performed a secondary differential screen using RNA prepared from the heads of *white* flies (with normal sized eyes) and RNA made from *Microcephalus/Transabdominal* (*Mc/Tab*) fly heads (a strain having greatly reduced eyes). They found that of the seventeen clones, nine showed a non-differential pattern of expression. One gave stronger expression with *Mc/*

Tab RNA, indicating a gene specific to the brain or perhaps some non-neuronal tissue like the mouthparts. The remaining seven clones gave a stronger signal with the probe made from *white* heads, signifying genes wholly or largely restricted to the eye. These seven clones were chosen for further analysis. An additional differential screen was performed to select for those clones which were restricted in their expression to photoreceptors. RNA from heads of the mutant *glass*^{3} (Pak *et al.*, 1969) which possesses reduced numbers of photoreceptor cells was used as a probe. Consequently, three of the seven clones were classified photoreceptor specific.

The fidelity of the screening strategy of Levy *et al.* (1982) was confirmed by the findings of Montell *et al.* (1985) which showed that one of these cloned genomic fragments complements a mutation *trp* (*transient receptor potential*) which specifically affects photoreceptor cells. As will be discussed later, demonstrating function (by complementing a mutant phenotype) is of great importance to the eventual success of a 'reverse genetic' approach. Fryxell & Meyerowitz fully characterised one of the photoreceptor specific genes. Carrying out a final differential screen utilising RNA made from a mutant which lacks the R7 photoreceptor cell (*sevenless*, *sev*^{LY3}; Banerjee *et al.*, 1987) they were able to identify one clone which seemed specific to this photoreceptor. cDNA sequence revealed an opsin with ~37% amino acid identity to each of the other cloned *Drosophila* opsins.

Zuker and colleagues (Shieh *et al.*, 1989) used a slightly different approach to look for phototransduction genes. 'Eye-specific' probes were prepared by hybridising wild type first strand head cDNA with a 20-fold excess of body mRNA. The single stranded fraction (recovered by Hydroxyapatite chromatography) was then subjected to a second round of subtraction with excess mRNA from heads of the eyeless mutant *eya* (*eyes absent*; Sved, 1986). Single stranded molecules (a population now enriched for molecules specific to the eye) were again recovered and labelled to high specific activity. This highly enriched probe was then used to screen a genomic DNA library. One bacteriophage clone was shown (by Northern analysis) to be absent from wild type bodies and *eya* heads, but present in the heads of wild type flies. A cDNA isolated subsequently was found to encode a homologue of Cyclophilin, a

protein which is thought to have a role in folding of Proline containing polypeptides. The authors hypothesised that this protein, the product of the *ninaA* locus might be involved in the light dependent activation and inactivation of the rhodopsin within photoreceptors. This screen also led to the isolation of an eye specific Protein Kinase C.

By far the largest screen of this type has been that carried out by Seymour Benzer, Elliot Meyerowitz and colleagues (Palazzolo *et al.*, 1989). This subtracted screen yielded 436 cDNA clones designated 'Head Not Embryo' (HNE). Palazzolo first made a cDNA library from mRNA isolated from wild type *Drosophila* heads. Sense RNA was transcribed from this library, and reverse transcribed into antisense cDNA. This template was hybridised with an excess of embryonic mRNA and the single stranded fraction recovered and subjected to a second round of subtraction. The HNE cDNA was made double stranded and cloned to generate a subtracted library. Preliminary characterisation of the library eliminated clones without inserts and clones which still seemed to represent genes expressed in early embryos. Finally, cross hybridisation analysis identified 436 unique clones. Of these clones isolated, 39 were expressed exclusively or predominantly in the adult visual system. 20 new cDNAs were chosen for further analysis (the other 19 being duplicates or artefacts). Hyde *et al.* (1990) report detailed characterisation of these twenty HNE, visual system cDNAs, tabulating their expression patterns as observed by Northern and *in situ* hybridisation analysis. In addition, Hyde *et al.* reported the sequence of one of these cDNAs which had been identified as the *Drosophila* homologue of the vertebrate arrestin, another protein involved in phototransduction. Partial sequences of the other 19 cDNA clones in this study were obtained and compared to sequences in the Genbank database. None corresponded to previously reported *Drosophila* cDNAs. It was encouraging that this screen did not merely re-isolate previously cloned genes. One disadvantage of this strategy was that the clones isolated were all very short (50-600bp). Although this does not bias the expression profile of the clones, preliminary characterisation is somewhat hampered by the fact that few of the clones actually contain coding sequences.

Any one molecular approach may not prove effective with every gene. The levels of expression of a gene like *amnesiac* for example, may be so low as to prohibit isolation using a differential screen. For a gene like this, having no candidate function or homologue, we can't clone by homology. A subtractive screen (for example, of wild type *versus* deficiency mutant) might lead to its isolation. The variety of approaches used reflects the differing needs of neurobiologists. This variety of approaches will in the long term lead to a richer understanding of the molecular basis of the nervous system. The real worth of a differential screen is speculative. Unless you screen a vast number of clones, the probability of finding a particular rare differentially expressed gene will be very small. However, the chance of finding one rare transcript which is a member of a large class of differentially expressed genes is reasonably high.

If we consider the success of the approaches described above, then it is obvious that the traditional differential screen as carried out by Levy *et al.* (1982) is hampered in its utility by insensitivity. Sixteen of the seventeen bacteriophage clones isolated in Levy's screen encoded transcripts which were shown subsequently to be expressed within the compound eye, seven of these being specific to eye tissue. The *Drosophila* compound eye is made up from 800 repeated elements and with its associated neural tissue accounts for 50% of the tissue in the brain. A conventional differential screen is really only sensitive enough to find genes from the eye because RNA transcribed within the repeated ommatidia dominate any sample of head RNA. Only one of the Levy bacteriophage clones characterised to date does not encode a transcript which is specific to the eye. (Ray & Ganguly, 1992). A differential screen which avoided such genes, whilst still being attractively simple to carry out might yield a larger proportion of novel genes.

1.7 Expectations and Considerations

As discussed previously, the *Drosophila* head is thought to express around 11,000 mRNA species, around 70% of the total number present in the adult fly (Levy & Manning, 1981). Only a fraction of these mRNAs will however be unique to the head, or indeed to the brain. The structural genes encoding mouthparts and antennae (though significantly not the eyes in this work) will all be

represented in a head cDNA library, despite having little or nothing to do with the brain. In addition all general structural genes and housekeeping genes will be represented even though they are not unique to the head. As detailed in Chapter Four, I chose to make a cDNA library using mRNA made from the heads of *eya* flies (Sved, 1986), mutants which lack the compound eye and much of the associated neural tissue. In this way, transcripts expressed solely or predominantly within the eye were avoided or under-represented, in the same way that using head, rather than whole fly mRNA under-represents or eliminates genes expressed wholly or predominantly outside the head. Admittedly this approach would still not yield clones representing transcripts of the rarity which a subtractive screen might, but at least it makes differential screening a worthwhile approach for those who (unlike Benzer or Zuker) wish to study parts of the brain other than the eye. It is important to recall that many general neural genes will not be specific to the brain because neural tissue is found in the body as well. To only choose specific molecules would be to bias this screen too much - we still need to know the basic workings of the Nervous System, even though it is the specific processes which are likely to be most exciting.

If we think of the insect nervous system as a collection of interacting cells, then the identity of these cells is defined by the genes expressed within them. Examining these genes, it is relatively easy to predict those whose expression is likely to be confined to the nervous system or one part of it. A major activity within the Nervous System is cell to cell communication - and genes whose products control and coordinate intercellular signalling (like gap junction proteins and cell adhesion molecules)are likely candidates for restricted expression within the Central Nervous System. We are likely to find genes encoding membrane localised or cytoplasmic proteins which mediate the uptake and release of neurotransmitters, 'hormones' and other messengers of the Nervous System. A subset of the kinases and phosphorylases involved in regulating these processes will also be brain specific. Intra-cellular signalling involves a vast number of genes from the receptor linked G-proteins at the cell surface through the messengers like cAMP phosphodiesterase, to the Protein Kinases which specifically alter the activity of the proteins controlling the behaviour of the cell. The anatomy and architecture of the nervous system is complex and the

genes controlling its specification and directing its assembly are likely to encode spatially restricted proteins. In turn, many of the structural proteins under their control will have spatially restricted patterns of expression. Much is now known of the biochemical basis of neurotransmitter synthesis. Cloning the genes regulating these biochemical pathways and the genes encoding the enzymes involved in neurotransmitter synthesis would facilitate a comprehensive analysis of the control of these pathways. Of great interest would be the genes encoding neurotransmitters and peptide hormones themselves. Although the latter are often simple molecules, the type of control exerted upon their transcription and translation is likely to be both complex and compelling. The proteins involved in this control (for instance transcription factors and peptidases) will also be worthy of investigation. There may also be neural housekeeping genes, encoding specialised forms of ubiquitous proteins like actins or basic metabolic proteins. Finally there will be many more genes, the nature of whose product or function cannot yet be predicted. When cloned, the gene disrupted in *eyes absent* mutant flies showed little or no homology to previously cloned genes in *Drosophila* or other eukaryotes (Bonini *et al.*, 1993), the same being true for the product of the l(1)optomotor blind locus (Pflugfelder *et al.*, 1992). In each case, the genes were cloned after walking from chromosomal breakpoints. The amino acid and nucleotide sequence for both these genes are now in GenBank and if any related genes are cloned, function may be inferred from their common features.

A class of genes which might not appear to be differentially expressed are those where more than one transcript is derived from the same locus by differential splicing or utilisation of a different transcription initiation or polyadenylation site. These genes would not appear specifically expressed in a differential screen though their expression pattern might indicate elevated expression (presence of a second transcript) in only one tissue.

If the methods for differential screening were faultless, then one would expect to pick up all these classes of gene, but the clones obtained will most likely correspond to those whose level of expression is highest. Even so, expression at very high levels in a small number of cells may not be enough to ensure that a gene will be cloned. There is however, no reason to presume that highly expressed genes will not be as interesting.

The nervous system is often thought of as a black box. Sensory stimuli from an animal's environment invoke internal processing and a response is produced. There is an input and an output, but within the black box, the Central Nervous System is a complex network of interactions. Using molecular biology we can begin to study the Central Nervous System as a whole (not as a series of inputs and outputs). As individual transmission pathways are elucidated they can be integrated, treated together and not in isolation. *Drosophila*, a relatively simple organism (when compared to mammals) provides an ideal opportunity to study the Nervous System as a unit. In recent years, studies of other more complex behaviours have found it difficult to draw real conclusions because so many diverse processes are involved. To illustrate, courtship involves movement, visual, auditory, olfactory and tactile cues and is greatly dependent upon experience (learning). To look at courtship behaviour without considering the importance of experience would be pointless, and yet studying learning (the gaining of experience) is a full time study in itself. At least, the presence of some framework would allow investigators to perceive just how closely related and how interlinked specific processes are, allowing us to ask the right questions.

A main obstacle in the study of neurogenetics over the past decade or so has been the attitude that real conclusions cannot be drawn unless all possible explanations have been considered - an effectively impossible task. Consider *dunce*: as discussed earlier, the *dunce* product is known and its biochemical role is clear. Other genes from the second messenger pathway like *dunce* have been correlated with mutants which also show 'learning' defects (see Davis, 1993). *dunce*, exhibits a somewhat pleiotropic phenotype. Mutant flies have aberrant cAMP phosphodiesterase levels, various internal structural abnormalities and there are effects on female fertility (Bellen & Kiger, 1988). Many researchers argue that to say that 'cAMP phosphodiesterase is an enzyme important in the learning processes of *Drosophila melanogaster*' cannot be proven. The learning defects of *dunce* flies may be due to one or a combination of the other phenotypic abnormalities. For example, if a *dunce* fly was unable to smell it would not learn, and this would not be due to a 'learning', but to a sensory defect. Of course, controls may be used to clarify such issues. Ultimately, only by knowing the whole story can one satisfy these critics and only using molecular approaches can we ever hope to find the whole story. Molecular biology provides this key

to many phenomena in neurobiology. Ultimately, 'reverse genetic' approaches must give way to biological studies which prove (or disprove) hypotheses formulated on the basis of molecular data.

1.8 New Tools

One group of methods now available holds great promise for molecular studies of neurobiology. The use of transposons (and in particular the P-element) in the past five years has been a great driving force behind these studies' recent advances. Suddenly, genes cloned on the basis of expression pattern or homology can be mutated almost 'on demand', allowing structure and function to be related and alleviating the major disadvantages of this 'reverse genetic' approach. Approaches based on the Polymerase Chain Reaction (like those developed by Ballinger & Benzer, (1989), Kaiser & Goodwin, (1990) and Milligan & Kaiser (1993) - discussed in greater detail in Chapter Three) allow the selection of new transposon induced mutant flies without reference to a visible phenotype. This method, 'Site-Selected' Mutagenesis (SSM) has been used successfully to disrupt a number of genes, including the gene encoding synaptotagmin, a protein which modulates neurotransmitter release at synapses. Littleton *et al.*, (1993) were able to generate a lethal insertion within this gene (*syt*). Subsequently, they remobilised this element, generating imprecise excisions and flies with less viable mutant phenotypes. Physiological studies were then initiated on these lines in an attempt to determine the exact mode of action of the *synaptotagmin* gene product. Clark *et al.*, 1994 used wild populations and screened for naturally occurring P-elements close to accessory gland protein encoding genes cloned by a reverse genetic approach (diBenedetto *et al.*, 1987). Although not phenotypic insertions, the P-elements in these lines could be remobilised to generate imprecise excisions which may themselves significantly disrupt the accessory gland protein genes. Strains containing marked P-elements throughout the genome are available from stock centres and a fly line harbouring a P-element in the proximity of the gene under investigation might already be available. This leaves the investigator with the relatively simple task of remobilising the P-element and creating new mutant alleles at this locus.

With the development of P-element mediated germ line transformation (Rubin & Spradling, 1982), mutant phenotypes can be rescued with wild type copies of genes, allowing researchers to test hypotheses on gene function and answer the calls of the critics cited above. The rescue of the phototransduction mutation *trp (transient receptor potential)*, by Montell *et al.*, (1985) is an ideal illustration of the power of such a technique. Flies which contained no functional copy of the *trp* gene were transformed with P-element DNA, harbouring wild type *trp* sequences. Flies in which the P-element-*trp* transgene had integrated into the genome were 'cured' of their abnormal electroretinogram phenotype, providing conclusive proof that the cloned *trp* sequence represented the necessary information for correct specification of this element of the fly visual system. As an alternative to differential screens, 'enhancer trap' screens can be undertaken to look for new P-element insertions in the vicinity of genes which are under the control of tissue specific regulatory elements (O'Kane & Gehring, 1987). Such screens utilise engineered P-elements which carry a copy of the *E.coli* β-galactosidase gene (*lacZ*) under the control of a weak promoter. If the P-element construct inserts close to a tissue specific enhancer, then high level expression of the *lacZ* gene is initiated, in the same tissue specific pattern. This pattern can be easily visualised by staining tissue *in situ* with the chromogenic substrate X-gal. P-elements engineered to include a plasmid origin of replication (in addition to the *lacZ* gene) and inserted near to a gene of interest (with localised *lacZ* expression - the equivalent of a positively hybridising clone in a differential screen) can be rescued (Bier *et al.*, 1989; Hamilton *et al.*, 1991a), along with a fragment of the flanking genomic DNA. Cloning DNA by plasmid rescue gives us a molecular foothold in the vicinity of the gene. Initial enhancer trap P-element insertions are rarely phenotypic. Once cloned, new (more severe) mutations can be generated by remobilising the P-element to create small deletions. In this way, a collection of mutant alleles can be efficiently collected, allowing more sensitive study.

Recently this technology has been developed to increase the utility of the enhancer trap approach. Giniger *et al.*, (1993) developed a P-element vector containing the *lacZ* gene fused to part of the kinesin coding sequence. Normally, kinesin is involved in vesicular transport up and down axons. In the engineered kinesin-*lacZ* construct, the kinesin derived region transports the β-Galactosidase

away from the nucleus down the axon. X-gal staining of enhancer trap lines bearing this construct traces the organisation of axons within the brain, allowing precise characterisation of neural circuits. The disadvantage of such an approach is that a new enhancer trap screen must be set up for each new construct devised. To obtain the construct of choice next to the gene of choice might require a great deal of work.

To alleviate these disadvantages, second generation enhancer trap elements have now been developed (Brand & Perrimon, 1993). Instead of *lacZ*, expression of the yeast transcription factor *GAL4* (Fischer *et al.*, 1988) is directed by tissue specific enhancers in the vicinity of the inserted P-element. The *GAL4* protein will subsequently direct expression of any gene placed downstream of its DNA binding site UAS_G. After an initial mutagenesis, a collection of lines is established with this *GAL4* element inserted randomly throughout the genome. A second P-element construct, with any one of a number of genes under the control of a UAS_G element can be subsequently crossed into any of these original lines. Figure 1.1 illustrates this technique. In theory, any gene may be put under the control of the UAS_G element. For instance, one could introduce a UAS_G linked antisense *transformer* cDNA, which has the effect of masculinising tissue in which it is expressed (compare this with the gynandomorph mosaic mapping of Hotta & Benzer, 1970). Together these techniques can facilitate detailed *in vivo* examination of the function of specific genes. Constructs carrying more powerful transposon based tools are being developed constantly, ever widening our prospects.

Recent molecular studies have provided information about many of the processes and genes of the nervous system. It is clear that no single approach will answer all the questions. It is certain however, that together the molecular approaches described here will continue to aid our dissection of the functioning of the *Drosophila* brain.

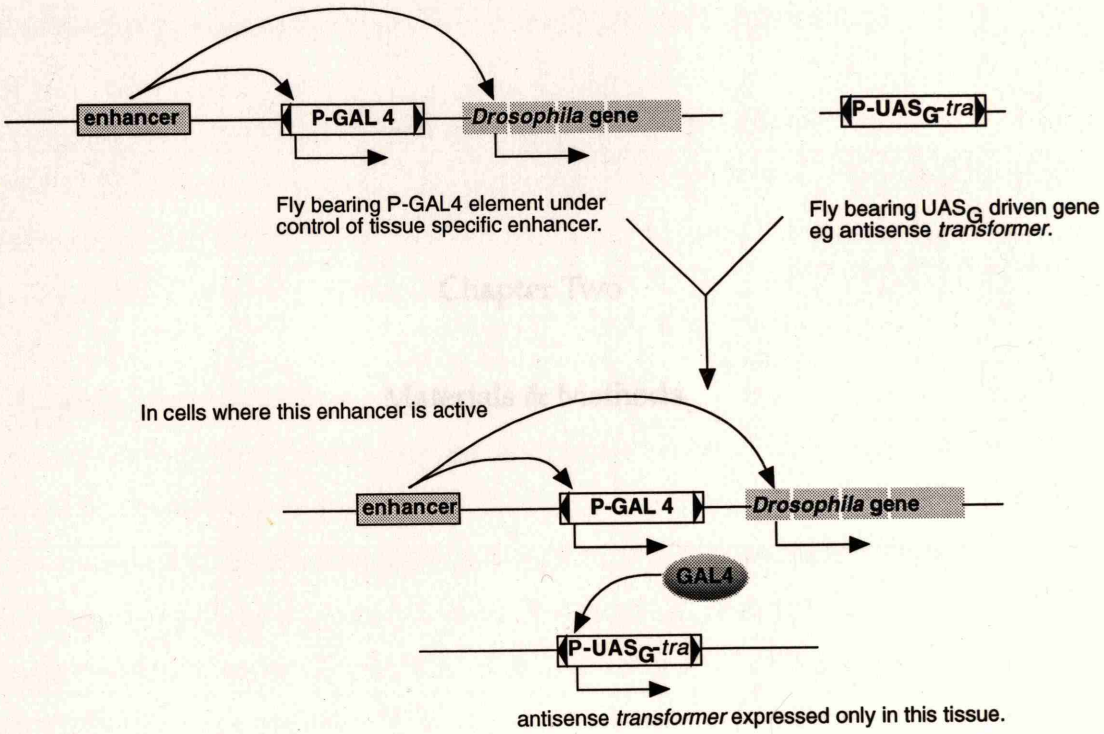


Figure 1.1 *Second Generation Enhancer Trap Technology*

A binary crossing scheme as used to generate flies expressing a Gal4 responsive gene in a tissue of interest. In theory, any gene may be put under the control of the UASG promoter, and be made responsive to Gal4. Examples of GAL4 fused genes available at present include antisense *transformer*, *kinesin lacZ* and *even skipped*.

Chapter Two

Materials & Methods

2.1 Growth and Storage of Bacteriophage λ and *E.coli*

Media for the growth of bacteriophage λ and *E.coli* were prepared as described in Sambrook *et al.*, 1989.

2.1.1 Culture

For propagation of bacteriophage λ, cultures were supplemented with 10mM $MgSO_4$; its presence aids bacteriophage adsorption. A less nutritious medium, BBL Agar was used for titration of Bacteriophage libraries and for secondary screens as detailed in Section 2.12. Phage plaques are larger but contain less bacteriophage when grown on BBL rather than on L-Agar plates.

All media were sterilised by autoclaving prior to use. *E.coli strains* were grown at 37°C in L-broth or L-agar. Liquid cultures were shaken vigorously during incubation. 2xYT broth was used for the propagation of single stranded phage in the λZAP II excision protocol.

2.1.2 Antibiotic Selection

The antibiotics Tetracycline hydrochloride and Ampicillin (Sigma) were used to facilitate selective growth of *E.coli* harbouring plasmids carrying the appropriate antibiotic resistance genes. Ampicillin was added to cultures and plates at a working concentration of 50-60μg/ml. Tetracycline was used at a final concentration of 12.5μg/ml. 100x stocks of each of these antibiotics were stored for short periods at -20°C.

2.1.3 Storage

Bacteriophage λ were stored as plugs or plate lysates at 4°C in Phage Buffer (10mM Tris-Cl pH 7.5, 10mM $MgSO_4$) or SM buffer (20mM Tris-Cl pH 7.4, 100mM NaCl, 10mM $MgSO_4$).

Bacterial strains were stored at room temperature as stabs grown on L-Agar or at -70°C as glycerol stocks; prepared by adding 1ml of an overnight culture to 1ml 40% (v:v)glycerol/2% (w:v) peptone.

2.2 Vectors and Hosts

2.2.1 Bacteriophage λ Vectors

The cDNA library used in the initial differential screen was constructed in the vector λZAPII (Stratagene; Short *et al.*, 1988). The genotype of λZAPII is λ*sbhI*λ1° *chi*A131 <T *amp Col*E1 *ori lac*Z T3 promoter-polycloning site-T7 promoter I> *srI*λ3° *cIts*857 *srI*λ4° *nin*5 *srI*λ5°. Subsequently, cDNA clones were isolated from a library made in the vector λNM1149 (Murray, 1983) by Steven Russell. This λNM1149 library was directional and constructed according to a protocol reported by Dorssers & Postmes, (1987). The genotype of λNM1149 is λ*b538* *srI*λ3° *imm434* *srI*λ4° *shnd*IIIλ6° *srI*λ5°.

The genomic library screened for the λD913 genomic clones was constructed in this laboratory by Simon Tomlinson. Size selected *Sau*3A1 digested wild type genomic DNA was used to produce this genomic library which utilised the Promega vector λGEM-11 (Promega). This vector is derived from the EMBL3 vector (Frischauf *et al.*, 1983).

2.2.2 Plasmid Vectors

Unless otherwise stated, plasmid subcloning utilised the SK⁻ form of the plasmid pBlueScriptII (Stratagene, USA). This plasmid is identical to the plasmid obtained when non-recombinant λZAPII bacteriophage are excised. Cloning of the I factor amplification products described in Chapter 3 utilised the KS⁻ form of pBluescriptII.

2.2.3 Single Stranded Helper Phage

The modified f1 bacteriophage R408 (Stratagene, Russel *et al.*, 1986) was used as a helper phage for excision of λZAPII cDNA inserts as single stranded phagemids, prior to their growth as newly generated plasmids.

2.2.4 *Escherichia coli* Host Strains

Four host strains were utilised in this work:

For most bacteriophage work, and for cloning procedures which utilised blue-white selection, the *recA* strain XL1-Blue (Stratagene; Bullock *et al.*, 1987) was used. The genotype of this host is as follows: *recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac,* [F′ *proAB, lacI*^{q}*ZΔM15,* Tn*10* (tet^{r})].

Initial propagation of the λZAP cDNA library was carried out in the host strain PLKF' (Stratagene), which is deficient for the *mcrA* and *mcrB* restriction systems. Its genotype is as follows: *re14 (mcrA), mcrB1, recA, lac, hsdR2,* ($r_{k-}m_{k+}$), *supE44, galK2, galT22, metB1,* [F' *proAB, lacI^{q}ZΔM15,* Tn*10* (tet^r)].

The host strain, NM621 (Whittaker *et al.*, 1988) was used for some bacteriophage work. Its genotype is: *hsdR, mcrA, mcrB, lac, supE44, recD 1009.*

The strain JM109 (Yannish-Perron *et al.*, 1985) was used in the preparation of some templates for sequencing toward the end of this study after a paper by Taylor *et al* (1993) indicated that *endo*A$^-$ strains yielded better quality templates for double stranded sequencing. Its genotype is as follows: *endA1 recA1 gyrA96 thi hsdR17* ($r_{k-}m_{k+}$)*relA1 supE44* D(*lac proAB)* [F'*proAB, lacI^{q}ZΔM15,*].

2.3 *Drosophila melanogaster* Media and Culture.

Flies were generally kept at 16-18°C on a standard fly media containing 10g Agar, 15g Sucrose, 30g Glucose, 35g Dried Yeast, 15g Maizemeal, 10g Wheatgerm 30g Treacle, and 15g of Soya flour in a final volume of 1 litre with the addition of 5ml Propionic acid and 10ml 10%(w:v in ethanol) Nipagen (methyl *p*-hydroxybenzoate; BDH) as anti-fungal agents.

During amplification and for large scale crosses and mutageneses, flies were kept in population cages at 25°C. In these cages, the flies were fed on plates containing Grape Juice Agar (52g Glucose, 26g Sucrose, 7g Yeast, 20g Agar, 58.8ml Grape Juice and 6ml 10% (w:v) Nipagen per litre), smeared with live liquid yeast to encourage egg-laying. The plates were changed daily and eggs transferred to a rich larval medium (100g Yeast, 20g Agar, 100g Glucose and 6ml 10% (w:v) Nipagen per litre) for 4 days before transferring the crawling third instar larvae to standard fly media for the final five days of their development prior to eclosion.

2.4 *Drosophila melanogaster* stocks used.

Oregon R; Kevin O'Hare; Imperial College London, UK. A wild type strain used as a source of DNA and RNA and also for *in situ* hybridisation experiments.

eya^{1}; *eyes absent*. Sved, 1986; Bonini *et al.*, 1993
a mutant strain, adult flies lack the compound eye and associated neural tissues. Used as a source of RNA for the cDNA library.

w^{IR1}; Bucheton *et al.*, 1984

w^{IR5}; Sang *et al.*, 1984

Strains bearing I factors inserted at known positions within the *white* locus.

Luminy; David Finnegan, University of Edinburgh, UK.

*cha*RC^{+}; David Finnegan, University of Edinburgh, UK.

Wild type Inducer (I factor bearing) strains.

charolles; David Finnegan, University of Edinburgh, UK.
A wild type Reactive (I factor lacking) strain.

2.5 Oligonucleotides.

The sequences of oligonucleotides utilised in this work are given below, along with their name, and a description of their source and use.

TSP-IL `5' GAG GCA CGA CTT ATC TCT TCG GAG G 3'`

Runs from position 39 to position 15 of the I-factor sequence published by Fawcett *et al.*, 1986. This primer was used as a Transposon Specific Primer (TSP) in I factor site-selected mutagenesis.

TSP-IR `5' TAG CTG TAA GCC CCG TAG CTA ATG C 3'`

Runs from position 5289 to position 5313 of the I-factor sequence (see above) This primer was also used as a TSP in I factor site-selected mutagenesis.

GSP-1 5' GCA GGA ATG GTA TGA TAA CCG GCG G 3'

Runs from position -1793 to -1769 of the published *white* locus sequence (O'Hare *et al.*, 1984). This primer was used as a Gene Specific Primer (GSP) in site-selected mutagenesis of the *white* locus.

GSP-5 5' GGC TTC TTC TTG AAC TCG GGC TCG G 3'

Runs from position -1802 to -1826 of the published *white* locus sequence. This primer was also used as a GSP in site-selected mutagenesis of the *white* locus.

pT3 5' TAT GAC CAT GAT TAC GCC AAG C 3'

pT7 5' CGA CTC ACT ATA GGG CGA ATT G 3'

pT3 and pT7 were primers derived from the sequence of pBluescriptII and were used to allow amplification of cDNA inserts by the Polymerase Chain Reaction (PCR).

nm1149hin 5' AAC CTT CAG CCA GAA TCC ATT GCC 3'

gt10rev 5' GGC TTA TGA GTA TTT CTT CCA GGG TA 3'

These two oligonucleotides were derived from sequences flanking the cloning sites in λNM1149; nm1149hin lies just upstream of the *Hind*III site and runs from base 267 to base 290 of the published sequence (Nikolnikov *et al.*, 1984) whereas gt10rev runs from base 640 to base 615 of the same sequence and lies just downstream of the *Eco*RI site in the vector. These oligonucleotides were used for amplification of cDNA inserts, for sequencing of these PCR products and for screening the λNM1149 library. The gt10rev oligonucleotide was kindly provided by Douglas Crompton, University of Glasgow.

The following oligonucleotides were used for sequencing and PCR. Their origin will be detailed at the appropriate point in the text.

682KP1 5' GGG CTG GCC ATC AAA GG 3'

682KP2 5' TGT GTT AGA GAA ATA CG 3'

682KP3 5' CAA TTG GAC TAG CCC AC 3'

682KP4 5' GGC GCT AGC TCC CTT GG 3'

682KP5 5' ATG GTG TAT GGG ATC TG 3'

682KP6 5' CAG ATA CAA GGA GGT GC 3'

682KP7 5' CAG ATT CTT GTA CTT GGC CTG G 3'

974KP1 5' GCT GGT CGT AAT GCC TAT AGT GTG C 3'

913-3-1 5' TGC CAA ATT GTA AGT TG 3'

913-3-2 5' ACA GCA ATA ACA CAA AC 3'

913-3-3 5' GTC CGA TAC AGT ACC GAC 3'

913-7-1 5' AGA TAA GGC CGT GCA GG 3'

913-7-2 5' TAC CCA CTT GGT GCT TTG GG 3'

913-7-3 5' AAG CCA GAG GGA AAG AGC G 3'

The above oligonucleotides were synthesised on an Applied Biosystems oligonucleotide synthesizer using chemicals from Applied Biosystems and Cruachem.

T3 (Promega) 5' ATT AAC CCT CAC TAA AGG GA 3'

T7(Promega) 5' TAA TAC GAC TCA CTA TAG GG 3'

These oligonucleotides were bought commercially from Promega, and used for DNA sequencing.

A modified oligo dT primer was used in the construction of the cDNA library. This primer was supplied by Stratagene and had the following sequence:

5' GAGAGAGAGAGAGAGAGAGAACTAGTCTCGAGTTTTTTTTTTTTTTTTTTT 3'

GA repeat *XhoI* site 18xT

2.6 Plasmids and Phage (whose generation is not described elsewhere)

Various cloned DNA fragments with known patterns of expression were used as controls in this study:

α-tubulin-α1 (Kalfayan & Wensink, 1982); a component of the cell cytoskeleton which is expressed throughout development and through all tissues, was used as a control to ensure equal loading of Northern blots.

Actin 5C (Fyrberg *et al.*, 1981). An 8.7 kb *Eco*RI fragment of the actin gene resident at 5C in the X-chromosome, cloned into pBR322. This plasmid, representing a gene of known abundance was used to check the quality of the cDNA library.

rp49; *Drosophila melanogaster* ribosomal protein (O'Connell & Rosbash, 1984) used as a loading control for Northerns and as a control on Reverse Northerns.

λ41, pST41; a bacteriophage λ clone and plasmid subclone isolated by Simon Tomlinson and which correspond to the major opsin (Zuker *et al.*, 1985 O'Tousa *et al.*, 1985) in the *Drosophila* eye; this clone was used as a head specific control for Reverse Northerns.

pST170; a cDNA clone related to my clone pC13 isolated by Simon Tomlinson.
pmsfK10; a male specific clone provided by Diane Harbison (Harbison, 1995)

2.7 Buffers and Reagents

All buffers and reagents used were Analar, Biochemica or Molecular Biology Grade. Commonly used solutions were made as described in Sambrook *et al.*, (1989) and autoclaved for 20 minutes at 120°C.

2.8 Isolation of DNA

2.8.1 Bacteriophage λ DNA

Bacteriophage λ DNA was prepared as described in Chisholm, 1989.

2.8.2 Plasmid DNA.

Plasmid DNA was prepared as described in Sambrook *et al.*, 1989 by a procedure modified from the original Alkaline Lysis method published by Birnboim & Doly (1979). When large quantities of plasmid were required, DNA was purified by centrifugation in Caesium Chloride, again as described in Sambrook *et al.*,

1989. Plasmid DNA prepared specifically for double stranded sequencing was purified by extraction sequentially with Phenol, Phenol/Chloroform and Chloroform, before reprecipitation.

2.8.3 *Drosophila melanogaster* Genomic DNA

Drosophila melanogaster genomic DNA was made in one of three ways depending upon the amount and type of starting material and the quality of DNA required.

2.8.3.1 DNA From Eggs Laid by up to 100 Flies Over a 24 Hour Period

Eggs were washed into an eppendorf microfuge tube with tap water and then homogenised with a micropestle (eppendorf) in 50-100µl of ice cold homogenisation buffer (100mM Tris-Cl pH 8.0, 60mM NaCl, 10mM EDTA pH 8.0, 15mM Spermine, 15mM Spermidine and 0.5% (v:v) Triton X-100). The volume of the homogenate was then adjusted to approximately 450µl with more Homogenisation buffer. 10% (w:v) Sodium Dodecyl Sulphate (SDS) was then added to a final concentration of 1% (w:v) and 10mg/ml Proteinase K was added to a final concentration of 200µg/ml. After 2-3 hours incubation at 37°C, the homogenate was extracted twice with Phenol:Chloroform (1:1), once with chloroform and precipitated by adding NaOAc pH 5.2 (to a final concentration of 0.3M) and 2 volumes of Ethanol. DNA was recovered by centrifugation at 4°C. After removing the supernatant, the DNA pellet was washed with 70% Ethanol and resuspended in 200µl TE with RNAseA (prepared as described in Sambrook *et al.*, 1989 and stored as stock at 20µg/ml). After incubation for 1 hour at 37°C, the DNA solution was extracted once with Phenol: Chloroform (1:1) and twice with Chloroform prior to precipitation as before. The DNA was finally resuspended in TE (10mM Tris-Cl, 1mM EDTA; pH8.0) without RNAse A and was stored at 4°C.

2.8.3.2 Preparation of DNA From Single Flies

DNA from single flies was made according to a protocol from Ashburner, 1989.

2.8.3.3 Large Scale Preparation of DNA From Around 1,000 Flies

Flies were ground in liquid Nitrogen in a mortar and pestle and the powder transferred to a 30ml Wheaton Homogeniser with a pre-cooled spatula. 9ml Homogenisation Buffer (HB: 100mM NaCl, 30mM Tris-Cl pH 8.0, 100mM EDTA,

10mM β-Mercaptoethanol) was added along with 50µl Triton X-100. Once thoroughly homogenised, the homogenate was strained through Nylon gauze into a 30ml COREX tube and the nuclei pelleted by centrifugation at 10,000rpm in a Beckman centrifuge for 10 minutes at 4°C. The supernatant was discarded and the pellet resuspended in 1ml HB. Once in solution, the nuclei were lysed by the addition of 1ml Nuclear Lysis Buffer (NLB: 100mM Tris-Cl pH 8.0, 100mM EDTA, 100mM NaCl, 1% (w:v) Sarcosyl and 0.5mg/ml Proteinase K). The lysate was then incubated for 2-3 hours at 56°C, recentrifuged to pellet the debris and the supernatant transferred to a new 15ml Falcon Tube. 1.25g of Caesium Chloride was added for every ml of lysate and the resulting mixture was loaded into a Beckman ti70 ultracentrifuge tube and spun at 45,000 rpm for 24 hours. Aliquots of the spun DNA were collected and assayed on Ethidium Bromide plates (1% (w:v) water Agarose, with EtBr at a concentration of 0.1µg/ml. The most concentrated aliquots of DNA were dialysed against TE (3 changes of 5 litres over a 24 hour period) and the DNA checked on a gel before storage at 4°C. Accurate assessment of DNA concentration was achieved by assaying the Optical Density at 260nm of the DNA solution in a spectrophotometer.

2.9 Preparation of RNA

RNA was prepared from fractionated tissue homogenised in Guanidinium Thiocyanate (BDH) using a Polytron motorised homogeniser. For RNA work, solutions were made RNase Free (RF) with the addition of 0.1% DEPC.

2.9.1 Fractionation of Head and Body Tissue *en masse*

To facilitate isolation of large quantities of pure head and body RNA for construction of the cDNA library and for use in expresion studies, it was necessary to seive frozen flies *en masse*, as described below.

Flies were sorted and placed in 50ml Falcon tubes on ice in 5-10g batches (1/3 volume of Falcon tube). These flies were then frozen by addition of liquid Nitrogen and vortexed for ~1 minute to separate heads from bodies and appendages. After vortexing, the tissue was kept under liquid Nitrogen and passed through sieves (Endecotts, from BDH) of various gauge to fractionate heads from bodies. Typically, most bodies were retained by a 710µM sieve and

the remainder by a 600μM sieve. Wild type heads were retained by a 425μM sieve, whilst those of the *eyes absent* strain were retained by a 300μM sieve. The quality of fractionation could be assessed easily by removing a sample from the sieve and counting the ratio of heads to bodies under a binocular microscope. Effectively uncontaminated preparations (<1:400 body:head contamination) were used to make RNA. It is important to remember that a contamination rate of 1 body/10 heads gives an effective RNA population of 50:50 head:body RNA.

2.9.2 Purification by Centrifugation Through a Caesium Cushion

Tissue was homogenised in 5M Guanidinium Thiocyanate, 0.1M Tris-Cl pH8, 0.1M β-Mercaptoethanol for 60 seconds at full speed and the homogenate strained through glass wool before centrifugation at 10,000rpm for 15 minutes. After spinning, the supernatant was removed and layered onto a bed of Caesium in a beckman SW28 tube. Two types of cushion were used, 5.7M Caesium Chloride (10ml/tube) was used for most work, whilst 6.1M Caesium TriFluoroAcetate (CsTFA; 19ml/tube, Pharmacia) was used during preparation of the material used in construction of the cDNA library. Centrifugation through CsTFA is considered to produce RNA which is less contaminated with DNA. After layering of the homogenate into the centrifuge tubes, the buckets were balanced and centrifuged at 25,000rpm for 25 hours at 14°C. After centrifugation, the RNA pellets were resuspended in 5% phenol, 0.2M NaOAc pH5.2 and reprecipitated with 2.5 volumes of ethanol. This pellet was then resuspended in 7.5M Guanidinium Hydrochloride, 75mM NaOAc pH5.2 and 0.5% (v:v) Sarkosyl, phenol/chloroform extracted and stored as an ethanol precipitate at -70°C.

2.9.3 Preparation of RNA by Guanidinium-Phenol-Extraction

Tissue was homogenised as above, but in a solution of 4M Guanidinium Thiocyanate, 0.1M Tris-Cl pH 8.0 and 0.1M β-Mercaptoethanol. After homogenisation, RNA was isolated as described in Chomczynski & Sacchi, (1987), but incorporating an extra precipitation with 4M Lithium Chloride (BDH) as recommended by Puissant & Houdebine (1990). This step is proposed to reduce the quantity of contaminating polysaccharides in the final RNA sample and was found to make the RNA pellet more soluble in subsequent manipulations.

2.9.4 Isolation of Poly A+ RNA

Poly A+ RNA was isolated from total RNA using oligo dT cellulose (type 7; Pharmacia). Oligo dT cellulose was prepared by swelling in RF dH_2O and equilibrating with 1x Binding Buffer (20mM Tris-Cl pH 7.5, 500mM NaCl, 1mM EDTA, 0.1% Sarkosyl). Up to 500mg total RNA was combined with an equal volume of 2x Binding Buffer, giving a final solution with the ion concentration of 1x Binding Buffer. This was added to pre-swelled oligo dT cellulose (typically 0.1g) in a falcon tube and the mixture rocked gently at room temperature for 20 minutes. Once the poly A+ RNA was bound, the solution was spun briefly at 2000rpm to pellet the RNA bound to the oligo dT cellulose. Unbound RNA was carefully removed and the pellet re-suspended in a small volume of RF H_2O, mixed and pelleted. This time, the supernatant was removed and stored and the process repeated two times. The eluted RNA was kept on ice whilst the oligo dT cellulose was stripped with 0.1N NaOH and re-equilibrated with 1x Binding buffer as before. The eluted RNA was then re-added to the oligo dT and the whole process repeated twice. A fraction of the eluted RNA was kept aside at each stage for analysis. After 3 purification steps, the poly A+ RNA was precipitated by addition of 1/10 volume of 3M NaOAc pH5.2 and 2.5 volumes of ice cold ethanol. Poly A+ RNA was stored at -70°C in this form and was recovered prior to use by centrifugation at 4°C in a Beckman SW28 rotor for 24 hours at 25,000 rpm. Yields were typically 2% of starting material.

2.10 λZAP II Excision

in vivo excision of recombinant pBSII SK- plasmids from the λZAP cDNA library was performed essentially as described in the Stratagene ZAP cDNA synthesis manual. The following brief protocol was developed to facilitate excision of large numbers of plasmids at once.

Bacteriophage plaques were picked into 500µl phage buffer and left for one hour. 200µl of phage were removed and mixed (in a Falcon tube) with an equal volume of a 200:1 mix of XL1-Blue:R408 helper phage (a fresh overnight of XL1-Blue, R408 at $1x10^6$pfu/ml) and incubated at 37°C. After 15 minutes, the mixture was vortexed and 5ml 2xYT added, before incubation for further 3 hours at 37°C. After this time, 1ml was removed to a microfuge tube and incubated at 70°C for 30 minutes. Cell debris was spun down and 200µl supernatant removed to a new tube with 200µl XL1-Blue. After a further 15

minutes at 37°C, half the mixture was plated out on L-Amp-Tet plates and incubated overnight at 37°C. Single colonies were picked into 20ml L-broth, grown overnight and plasmid DNA isolated as in Section 2.8.2.

2.11 Polymerase Chain Reaction

Conventional PCR (Saiki *et al.*, 1988) was used in a number of experiments either as a diagnostic tool, or as a means of generating DNA for further use in cloning or sequencing. PCR reactions were carried out in a final volume of 20µl. Up to 100ng of DNA was used as template in the presence of oligonucleotide primers at a concentration of 0.33µM and dNTPs (Pharmacia) at a final concentration of 200µM for each base. All PCR reactions were performed in 1x PCR buffer (supplied with the enzyme) and with 1 unit of *Taq* polymerase (purchased from Promega or Boehringer Mannheim). Samples were overlayed with mineral oil (Sigma) before incubation in air or water cooled thermal cyclers purchased from various manufacturers (Techne, Cambio and Hybaid). A typical cycling profile would be; incubation at 94°C for 1 minute followed by 30 cycles of incubation at x°C for 1 minute; 72°C for 3 minutes and 94°C for 1 minute. x was the empirically derived annealing temperature for each pair of oligonucleotides which was judged to generate least background. If optimisation was not possible, a temperature of 5°C below the T_m of the primer was used.

2.12 cDNA and Genomic Library Screening

For screening, cDNA and genomic libraries were plated on 10x10cm square plates. Plating cells were prepared by diluting 1ml of an overnight culture of NM621 into 100ml L-broth supplemented with 10mM $MgSO_4$. This culture was grown for 3-4 hours until the cells had reached an OD_{650} of 0.3-0.4. Cells were then chilled and centrifuged at 4,000 rpm at 4°C for 15 minutes and gently resuspended in 0.4 volumes of 10mM $MgSO_4$ to give a final OD_{650} of 1.0; corresponding to $1x10^9$ cells per ml. Plating cells were stored at 4°C and kept for up to 3 days. For each plate, 10,000 pfu were combined with $5x10^8$ cells and incubated at room temperature for 20 minutes to allow phage to adsorb to bacteria. 6ml molten (42-45°C) L-Top Agar was added to the phage/bacteria mixture and immediately poured over the surface of the L-Agar plate. Once set, plates containing phage were incubated for 8-12

hours (depending on the vector used) until plaques were visible, but still well isolated. Nylon filters were lifted according to manufacturers protocols (Amersham) and hybridised as described in Section 2.18. Secondary screens were carried out using 9cm diameter round plates. An appropriate dilution (~1/1,000) of the primary phage plug (which had been picked into 500µl SM buffer) was plated onto these plates (with 3ml molten Agar and $2x10^8$ cells). This dilution would normally give 150-200 pfu/plate, allowing single plaques to be picked at this stage.

2.13 Restriction Digestion and Subcloning

2.13.1 Restriction Digestion

Genomic, bacteriophage and plasmid DNA was digested at 37°C for one to three hours in the appropriate incubation buffer with a three-fold unit excess of enzyme to ensure complete digestion. All restriction enzymes were obtained from BRL or Promega.

2.13.2 Subcloning

DNA fragments for subcloning were purified from TAE Agarose gel slices by centrifugation through glass wool, phenol extraction and ethanol precipitation. Occasionally, DNA fragments were prescipitated directly from a restriction digest or PCR reaction. Vector DNA was cut in large (10µg) quantities, and precipitated *en masse* after an aliquot had been fractionated on an agarose gel to check that digestion was complete. If vector (or insert) was to be treated with Calf Intestinal Alkaline Phosphatase (CIAP; to prevent re-circulation) then DNA was reprecipitated to remove the enzyme CIAP which may interfere with subsequent manipulation.

PCR fragments were cloned into *Sma*I cut pBS KS^-. Insert (I) and vector (V) were mixed at a molar ratio of 3:1 (I:V) and ligated in a final volume of 10µl using 1u of T4 DNA ligase (BRL). After ligation at 14°C for 24 hours, ligation products were transformed into competent XL1-Blue as described earlier and plated on media containing Ampicillin and Tetracycline, along with IPTG (β-D-Isoproyl-thiogalactopyranoside) and X-gal (5,-Bromo,4-Chromo,3-inodyl-β-galactopyranoside) to allow colour selection of recombinant clones - as described in Sambrook *et al.*, (1989).

cDNA inserts were cloned into pBS SK$^-$ cut with appropriate enzymes. All products of the bacteriophage restriction digest were precipitated and ligated to cut vector at a 30-fold molar excess of insert to vector (as typically only ~1/15 of the bacteriophage clone DNA was insert). After ligation, transformation and plating as before, Recombinant clones were analysed by restriction digestion and/or hybridisation.

2.14 Transformation of *E.coli* Hosts with Plasmid DNA

Competent XL1-Blue (the most versatile host available) or JM109 were transformed with plasmid DNA. Competent cells were prepared by a standard $CaCl_2$ procedure. Mid log phase cells(OD_{650}=0.4-0.5) were pelleted at 4,000 rpm in a Beckman centrifuge for 10 minutes at 4°C before being resuspended in one half volume of ice cold 50mM $CaCl_2$ and incubated on ice for 30 minutes. After this time, the cells were pelleted as before and resuspended in 1/20 the original volume of $CaCl_2$. Glycerol was added to a final concentration of 15% and the cells were snap frozen in liquid Nitrogen prior to storage at -70°C for up to 2 months.

For transformation, 100µl competent cells were thawed on ice and incubated (on ice) with 50-100ng plasmid DNA for 30 minutes. The cells were then 'heat shocked' at 42°C for 90 seconds and replaced on ice. 500µl L-Broth was added to the cells and the culture shaken at 37°C for 1-2 hours to allow genes conferring antibiotic resistance to be expressed. After this period, 200µl cells were plated out on media containing the appropriate antibiotic and incubated overnight at 37°C.

2.15 Sequencing, Polyacrylamide gels and Exonuclease III deletions

All sequencing reactions were carried out on double stranded DNA utilising either plasmid DNA or PCR product as template. Sequencing reactions were performed in accordance with standard protocols outlined in the USB Sequenase 2.0 sequencing kit. In all cases, α^{35}S dATP (Amersham, UK) was utilised as a radioactive label for sequencing reactions.

Sequencing reactions were run on 6-7% Acrylamide (29:1 Acrylamide:BIS Acrylamide), 7M Urea, 1x TBE gels; polymerised by the addition of 25µl TEMED and 1ml of 10% (w:v) Ammonium Persulphate just before pouring. The gels were cast in a Bio-Rad Sequigen gel kit and run in 1x TBE buffer at ~2200V (to keep a constant temperature of 55°C) for 2.5 to 6 hours. Gels were dried under vacuum (without fixing) and sequences visualised by autoradiography.

Unidirectional deletions were created in plasmids using a technique developed by Henikoff (1984) and modified by Promega. The enzyme, exonuclease III used was purchased from Promega or BRL. Deletions were performed using protocols derived from the Promega '*Protocols and Applications Guide*' (1992) using buffers originally prepared by Simon Tomlinson.

2.16 Sequence Analysis

DNA and Peptide sequences were collected, assembled and analysed on computer using the following packages. On Macintosh computers, sequences were assembled using AssemblyLign (IBI) and analysed using MacPattern (Fuchs, 1991) and MacVector (IBI). On PCs, the programs CLUSTALV (Higgins & Sharpe, 1988) and BOXSHADE (Hoffman, 1993) were utilised. For database searching, Unix and VAX based mainframe computing services linked to the GenBank Database were used. Data was manipulated and analysed using the GCG suite of programs developed by the University of Wisconsin (Devereux *et al.*, 1984).

2.17 Gel Electrophoresis of Nucleic Acids

2.17.1 Electrophoresis

Standard agarose gel electrophoresis was carried out on 0.4% - 1.5% (w:v) gels made with Ultrapure Agaose (BRL). Gel Loading Buffer (10x; 0.25% (w:v) Orange G, 0.25% (w:v) Bromophenol Blue, 0.25% (w:v) Xylene Cyanol, 25% (w:v) Ficoll) was added to samples for electrophoresis prior to running. Gels were run in 1x TBE (90mM Tris, 90mM Borate, 2mM EDTA pH8.3) buffer unless DNA was to be recovered in which case 1x TAE (40mM Tris, 40mM Acetate, 1mM EDTA pH7.6) was used as a buffer. DNA was visualised by staining in

0.5µg/ml EtBr and viewing the gel under UV light. When very high resolution of DNA fragments was called for, Vertical 0.7% (w:v) Agarose; 3% (w:v) Acrylamide gels were run. To pour such a gel, 0.7g Agarose was melted in 83ml H_2O and 10 ml of 10x TBE. Once cooled to around 60°C, 6.6ml of 40% (w:v) Acrylamide was added and the gel mix polymerised by the addition of 20µl TEMED and 1ml of 10% (w:v) Ammonium Persulphate before pouring. After electrophoresis, these gels could be stained with Ethidium Bromide and viewed as before.

Single stranded DNA (e.g. first strand cDNA) was fractionated on alkaline agarose gels, run in a buffer of 300mM NaOH and 20mM EDTA. In all cases, DNA was radiolabelled with α^{32}P dCTP or dATP, and visualised by drying down gels and autoradiography of the gel itself.

RNA was fractionated on Formaldehyde-MOPS denaturing gels as described in Sambrook *et al.*, 1989. EtBr (at a final concentration of 50ng/ml was added to the denatured RNA mixture prior to loading. This allowed visualisation of the RNA after electrophoresis and transfer to Nitrocellulose.

2.17.2 Recovery of DNA

DNA was recovered from gels either by electroelution into dialysis tubing (as in Sambrook *et al.*, 1989) or by spinning through glass wool. Recovered DNA was always phenol, phenol/chloroform extracted and precipitated before further use.

2.17.3 Size Markers

For DNA, three standards were used to allow determination of size of fragments on Agarose gels: λ DNA cut with *Hind*III and *Eco*RI (fragment sizes 21227, 5148, 4973, 4268, 3530, 2027, 1904, 1584, 1375, 974, 831, 564, 125bases) was prepared in the lab, whereas 1kb ladder (12 fragments corresponding to multiples of 1016bp along with fragments of 1636, 517, 506, 396, 344, 298, 220, 201, 154, 134, and 75bp) and 123bp ladder (multiples of 123bp) were purchased from BRL. For RNA, the molecular weight standard was also bought from BRL and possessed bands of 9490, 7460, 4400, 2370, 1350 and 240 bases.

2.18 Hybridisation and Labelling of DNA

2.18.1 Transfer of Nucleic Acids to Membranes

DNA was transferred to Nitrocellulose (Schleicher & Schuell) or Nylon (Amersham) either by capillary or vacuum blotting after pretreatment of gels according to the membrane manufacturers recommendations. Bacteriophage plaques were similarly transferred to Nylon using protocols from Amersham. RNA was transferred to Nitrocellulose as described in Sambrook *et al.*, 1989. Nucleic acids were fixed to Nylon membranes by automatic UV crosslinking in a Stratalinker™ and to Nitrocellulose by baking at 80°C for 2 hours.

2.18.2 Nucleic Acid Labelling

DNA was labelled with α^{32}P dATP or dCTP (650Ci/mMol and 3,000Ci/mMol from ICN; 800Ci/mMol from NEN Inc.) in one of a number of ways. Plasmid DNA was labelled to a high specific activity by Nick Translation using *E.coli* DNA polymerase I (BRL) according to the method of Rigby *et al.*, 1977. Plasmid DNA, restriction fragments and PCR products were labelled to a high specific activity by Random Priming (Feinberg & Vogelstein, 1982) using the Klenow fragment of DNA polymerase I (Promega). DNA markers for gels were labelled to low specific activity by 'filling in' with the Klenow fragment as described in the Promega '*Protocols and Applications Guide*'. In all cases, labelled probe was separated from unincorporated nucleotide by chromatography through Sephadex G-50 columns (Sambrook *et al.*, 1989) prepared in disposable 1ml syringes. Oligonucleotides were labelled with ^{32}P γ-ATP (3000Ci/mmol, NEN) using the enzyme Polynucleotide Kinase from bacteriophage T4 (BRL) as described in Sambrook *et al.*, (1989). These probes were not further purified before use.

First strand cDNA probes of very high specific activity were prepared as follows using polyA$^+$ RNA as template. 1μg RNA was annealed at 70°C for 5 minutes with 100ng oligo $dT_{(14\text{-}18)}$ (Pharmacia) and chilled on ice. This template was then incubated with 70μCi α^{32}P dCTP (800Ci/mMol); 500μM each dATP, dGTP and dTTP; 10mM DTT; 50mM Tris-Cl pH 8.3; 75mM KCl; 3mM $MgCl_2$ and 200 units of Superscript Reverse Transcriptase (BRL) for 90 minutes at 45°C. After this period, the reaction was 'chased' with cold dCTP (to produce longer probes) at a final concentration of 500μM and a further 80 units of enzyme for 45 minutes

at 45°C. After this step, EDTA was added to a final concentration of 20mM and NaOH was added to a final concentration of 600mM and the reaction incubated at 68°C for 45 minutes to hydrolyse the RNA template. Finally, labelled single stranded DNA was separated from unincorporated nucleotide by chromatography through Sepahadex G-50 (Pharmacia) as described before. A fraction of the probe was removed at this stage and incorporation assayed by Cerenkov Scintillation. From the measure of incorporation, it was possible to estimate the specific activity of the probe which was usually >1x 10^9cpm/µg. Further purification of the single stranded fraction of this probe was achieved by chromatography through Hydroxyapatite (HAP; Bio-Rad). A 1ml disposable syringe was kept at 68°C with a water jacket and a bed of HAP resin in 40mM $NaPO_4$ was poured. The probe was heated to 70°C and loaded onto the column. After collection of the unbound probe, the column was washed six times with 40mM $NaPO_4$, six times with 120mM $NaPO_4$ and six times with 400mM $NaPO_4$, collecting the eluate at each step. Single stranded DNA is eluted from the column by120mM $NaPO_4$ whereas double stranded DNA stays bound until the $NaPO_4$ concentration is raised to 400mM. Assaying the activity of the pooled 120mM fractions and comparing this with the activity of the 400mM fractions allowed an estimation of the proportion of single stranded DNA in the probe. This proportion was typically 90-95%. After assaying, double stranded fractions were discarded and single stranded fractions were combined to be used as the purified single stranded probe.

Typically, filters were prehybridised for >2 hours at 65°C before addition of probe. The 'prehyb' solution used for routine DNA hybridisations was as follows:

5x SSPE pH 7.4 (0.9M NaCL, 50mM NaPO4, 500µM EDTA)
10x Denhardts (0.2% (w:v) Ficoll, BSA and PVP)
0.5% (w:v) SDS
100µg/ml denatured sheared salmon sperm DNA

Filters to be hybridised with oligonucleotide probes were prehybridised as above, but at a lower temperature to reflect the different hybridisation kinetics of an oligonucleotide probe. A 17-mer oligonucleotide probe would be prehybridised and hybridised at 37°C.

Filters probed with first strand cDNA probes were prehybridised in:

1%SDS (w:v)
6xSSC pH7.2. (0.9M NaCL, 90mM NaCitrate)
5x Denhardts (0.1% (w:v) Ficoll, BSA and PVP)
50mM $NaPO_4$ pH 6.8 (1:1 NaH_2PO_4 : $Na_2HPO_{4)}$
100µg/ml denatured sheared salmon sperm DNA
0.005% (w:v) $NaPP_i$
1mM EDTA
30µg/ml PolyAdenylic Acid (Pharmacia).

Northern blots were prehybridised for >2 hours at 42°C in:

50% Formamide (v:v)
5x SSPE pH 7.4
2x Denhardts (0.04% (w:v) Ficoll, BSA and PVP)
0.1% (w:v) SDS

2.18.3 Hybridisation

Hybridisations were carried out for at least 16 hours in a Techne rotisserie oven or in sealed plastic bags in a waterbath. Probes in ~0.5ml TE were added to <10ml prehybridisation solution after boiling for 5 minutes and quenching on ice to denature the probe. There was no need for single stranded probes to be boiled before use.

2.18.4 Washing.

Stringency of hybridisation depends on the final salt concentration and the temperature of washing. Lower salt concentration and higher temperature being the more stringent.

DNA blots were washed
1x briefly in 2xSSC, 0.1% SDS (w:v) at 65°C.
1x 15 min in 1xSSC, 0.1% SDS (w:v) at 65°C
2x 15 min in 0.2xSSC, 0.1% SDS (w:v) at 65°C
1x 15 min in 0.1xSSC, 0.1% SDS (w:v) at 65°C

For low stringency hybridisation, DNA blots were washed to 0.5xSSC.

Oligonucleotide blots were washed for 2 hours in 2 x SSC at 37°C for 2 hours.

Northern blots were washed
1x briefly in 2xSSC, 0.1% SDS (w:v) at 42°C
1x 15min in 1xSSC, 0.1% SDS (w:v) at 42°C
1x 15min in 0.5xSSC, 0.1% SDS (w:v) at 65°C
1x 15min in 0.2xSSC, 0.1% SDS (w:v) at 65°C

Filters were exposed to autoradiography film with intensifying screens at -70°C.

Filters were wrapped in Saran Wrap to keep them damp in case they needed further washing or were to be stripped for re-use. If filters were to be stripped, they were immersed in boiling 0.1% SDS (w:v) and left for ~45 minutes while the solution cooled.

2.19 *in situ* Hybridisation to Polytene Chromosomes

Chromosomal localisation of the cDNA clones isolated was determined by hybridisation to third instar salivary gland polytene chromosomes as described by Pardue, 1986. Plasmid DNA to be used for hybridisation probes was labelled to low specific activity with Biotin-dUTP (Boehringer) using the Nick Translation technique described by Rigby *et al.*, 1977. Hybridising sequences were detected using the Vectastain Kit from Vector Laboratories and visualised using DAB and conventional Giemsa staining procedures. Some of the *in situ* hybridisation work presented in this thesis was carried out by Zong Sheng Wang.

2.20 *in situ* Hybridisation to Embryos

Embryos were prepared for hybridisation by a Paraformaldehyde fixation procedure derived from Tautz & Pfeiffle (1989). Briefly, embryos were dechorionated in Bleach and fixed by agitation in an emulsion of 1:1 4% (w:v) Paraformaldehyde, 100mM PIPES, 1mM EGTA, 2mM $MgSO_4$ pH6.8:Heptane for 90 minutes. The embryos were then collected in the organic phase by the addition of 9:1 Methanol:0.5M EGTA. Recovered embryos were rinsed and rehydrated through 7:3, 5:5, 3:7, Methanol/EGTA: Paraformaldehyde/PIPES. Hybridisation, washing and signal detection was carried out essentially as

described in Tautz, (1992). Hybridisation to fixed embryos was carried out using plasmid DNA labelled with DIG-dUTP (digoxygenin-dUTP). Labelling was achieved either by the method recommended in the Boehringer Mannheim DIG labelling kit (a modification of the Random Priming method of Feinberg & Vogelstein, 1982) or by a primer extension procedure developed by Patel & Goodman, (1992). This method utilised the T3 and T7 primer sites flanking the polycloning site of pBluescriptII plasmids. Using *Taq* polymerase and an oligonucleotide corresponding to just one of these sites, single stranded probes could be produced corresponding to sense or antisense sequences of the RNA species under investigation.

2.21 *in situ* Hybridisation to Adult Head Cryostat Tissue Sections

in situ hybridisation to sectioned head tissue was carried out according to a protocol from the Davis Lab at CSHL, USA. Briefly, serial frontal sections (10μM) were cut on an Anglia Scientific Cryotome (by Mingyao Yang) and fixed in 2% Paraformaldehyde; 10mM Sodium Periodate; 75mM Lysine in PBS. Sections were then washed in PBS and denatured in 0.2N HCl before treating with Pronase E (350μg/ml in 50mM Tris pH8; 5mM EDTA pH8). Pronase treatment was stopped by washing twice in Glycine (2mg/ml in PBS). Subsequently, sections were refixed in 4% (w:v) Paraformaldehyde in PBS, rewashed and acetylated in Acetic Anhydride (0.000625% (w:v)); Triethanolamine (1.3mM). Prehybridisation was carried out at 42°C in a humid box in the following Prehyb. solution; 5xSSPE: 50% (w:v) Formamide: 5% (w:v) Dextran Sulphate: 1xDenhardts (w:v): 500μg/ml Salmon Sperm DNA and 250μg/ml Yeast tRNA. Hybridisation was carried out in the same solution containing a small amount of probe (made as described in Section 2.20 for embryo *in situ* hybridisation). After hybridisation, the sections were washed in 2xSSPE (3 times (for 15 minutes) at room temperature): 1xSSPE (2 times at room temperature): 0.5xSSPE (once at room temperature): 2mM $NaPP_i$ (once at 42°C). Detection was performed essentially as described in Tautz (1992), except that the method was adapted for use on slides.

Chapter Three

'Site-Selected' Mutagenesis Using the I Factor Retrotransposon

3.1 Introduction

For ninety years now, *Drosophila* has been a favoured organism in which to study genetics. Since the discovery of DNA and the elucidation of the genetic code the fruit fly has become the premier model system in which to examine the relationship between genotype and phenotype. This eminence is due mainly to the amenability of *Drosophila* as an experimental organism (particularly its short generation time), the wide bank of variation available in stock centres throughout the world, and the relative ease with which new mutations can be induced and recognised.

Historically, a reliance on variation has ensured that those genes which have come to be studied in detail are those for which mutants have been available. From mutation, the journey to cloned gene is often long and arduous. Genes which when mutated give rise to severe defects are often the easiest to clone because they can be localised quickly by recombination whilst the availability of chromosomal duplications may be utilised to rescue and dissect lethal defects. Not every gene gives rise to obvious phenotypes when mutated. However in the last decade it has become apparent that to obtain a complete picture of how the fly works (or even how a small set of cells develop and function), other methods of study must be employed which do not initially rely upon recognition of an altered phenotype as a means of obtaining information.

As discussed in Chapter One, increasing numbers of genes are now cloned on the basis of a specific spatial or temporal pattern of expression, or by virtue of shared homology with a gene from another organism. Often, there are no known mutations at these loci. Much can be done to study these genes at the molecular level; the cDNA or genomic DNA may be sequenced, the peptide they encode predicted, their genomic organisation determined and their expression pattern described, but function can not be unequivocally attributed to a piece of cloned DNA unless *in vivo* function can be demonstrated. Ultimately, unless a mutant fly is available, in which an observable defect or an altered molecular phenotype is seen, then the whole process of molecular characterisation is of limited worth. There is therefore a pressing need for a system whereby one can progress from cloned gene to mutant phenotype with relative ease and efficiency; a system of targeted mutagenesis.

By 1990, such a system had been developed both in Glasgow (Kaiser & Goodwin, 1990) and in Pasadena (Ballinger & Benzer, 1989). These two groups used the Polymerase Chain Reaction (PCR; Saiki *et al.*, 1988) to facilitate a method of screening mutagenised flies for new transposon insertions at a particular locus. This method (called 'Site-Selected' Mutagenesis - SSM) allows efficient screening of large numbers of flies. Screening is performed on heterozygotes and isolation of new insertions is therefore not prejudiced by the severity of the mutation induced or the phenotype exhibited by the mutant flies. Initially, these experiments utilised the P-element, a transposable element which can be induced to 'jump' at high frequency in the progeny of crosses between specific fly strains. This technique has been used successfully to mutagenise a number of genes (e.g. Goodwin *et al.*, submitted; Littleton *et al.*, 1993).

For P-element SSM, a Transposon Specific Primer (TSP) is designed pointing outwards from the ends of the P-element (Figure 3.1a; the inverted repeat of 31 base pairs makes an ideal universal screening primer). In addition, an oligonucleotide primer is chosen from each gene to be disrupted. Evidence suggests that P-elements prefer to insert at the 5' end of genes (though this may merely reflect phenotypic insertions), and routinely, this Gene Specific Primer (GSP) is designed pointing upstream from around 300 base pairs downstream of the transcription start site to detect these events (see Figure 3.1b). Under normal circumstances, (e.g. on genomic DNA made from unmutagenised flies) there should be no template for PCR. In a mutagenised fly however (Figure 3.1c), the insertion of a P-element near the 5' end of a gene under study should generate a DNA template allowing PCR between the GSP and TSP, giving an amplification product which should be visible as a band on an agarose gel. If genomic DNA is made from the eggs laid by 100 mutagenised flies then any new insertions in that population can be screened in a single PCR reaction. An amplification product would denote the presence, in that population, of a female bearing a new P-element insertion at the locus under scrutiny. Subdividing the population and repeating the PCR screen on DNA made from eggs laid by these sub-populations would narrow down the search to 10 flies from which lines may be established before further characterisation (Figure 3.1d).

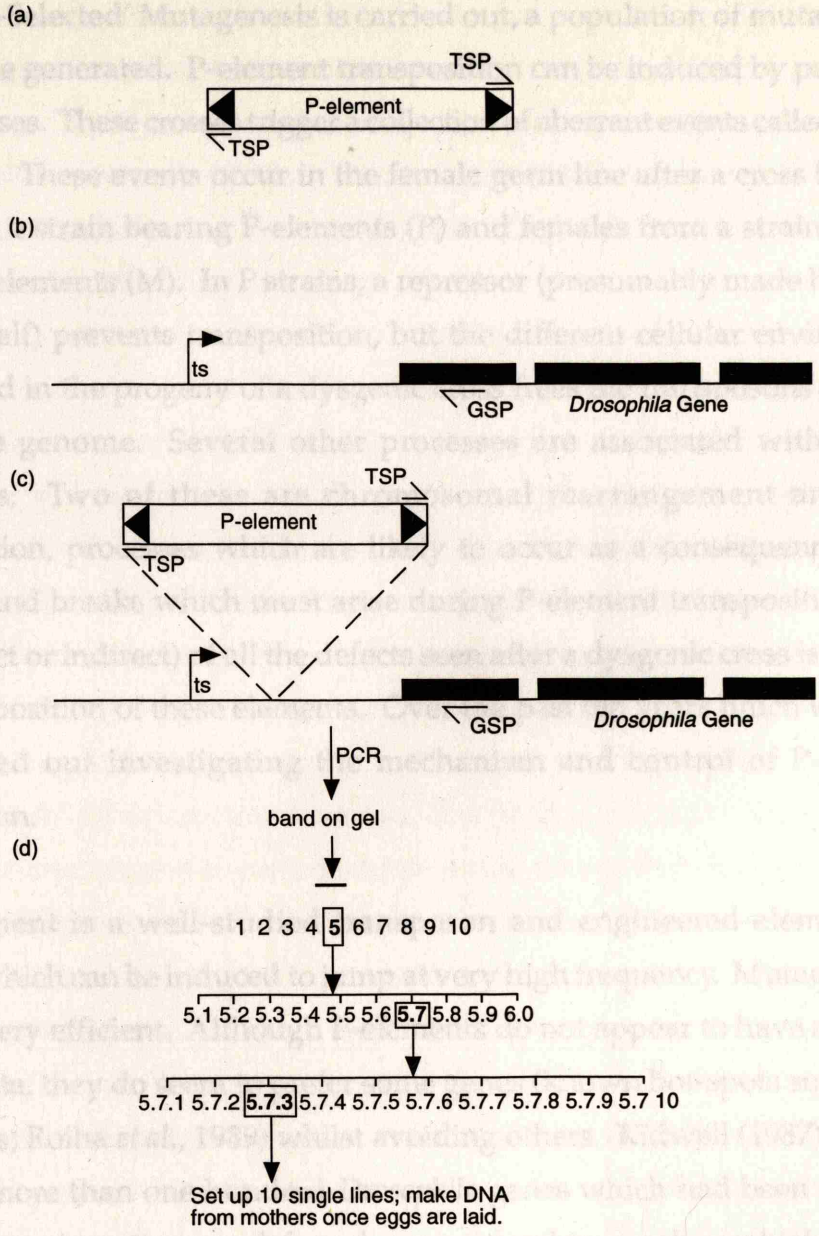


***Figure 3.1** 'Site-Selected' Mutagenesis*

(a) A wild type P-element has 4 ORF's and is 2907bp long, though many ammunition elements are shorter and unable to move unless supplied with transposase from a mutator element such as Δ2-3. All mobile P-elements must possess the 31bp terminal inverted repeats (black triangles). The TSP shown here will allow amplification of P-elements inserted in either orientation with respect to a GSP. TSP's designed from sequences just internal to the inverted repeats are unique and may be used to determine the orientation of new insertions. (b) GSP's are routinely designed from within the gene, pointing upstream towards the transcription start site (ts). (c) Transposition of a P-element to a site upstream of the GSP provides a unique template for PCR, generating a band on a gel or autoradiograph. (d) Sib Selection. PCR and hybridisation is carried out on 10 sub-populations of 1000 flies. If one population (5 here) is positive (i.e. a band is found) then this population is sub-divided into 10 populations and the process repeated on these groups of flies. Once the positive is narrowed down to a group of 10 flies, single female vials are set up and the flies allowed to lay for a few days before DNA is made from them and amplified as before.

Before 'Site-Selected' Mutagenesis is carried out, a population of mutagenised flies must be generated. P-element transposition can be induced by particular genetic crosses. These crosses trigger a collection of aberrant events called hybrid dysgenesis. These events occur in the female germ line after a cross between males from a strain bearing P-elements (P) and females from a strain lacking functional elements (M). In P strains, a repressor (presumably made by the P-element itself) prevents transposition, but the different cellular environment encountered in the progeny of a dysgenic cross frees the transposons to move around the genome. Several other processes are associated with hybrid dysgenesis. Two of these are chromosomal rearrangement and male recombination, processes which are likely to occur as a consequence of the double strand breaks which must arise during P-element transposition. The cause (direct or indirect) of all the defects seen after a dysgenic cross is thought to be transposition of these elements. Over the past ten years much work has been carried out investigating the mechanism and control of P-element transposition.

The P-element is a well-studied transposon and engineered elements are available which can be induced to jump at very high frequency. Mutagenesis is therefore very efficient. Although P-elements do not appear to have a specific insertion site, they do seem to prefer some genes (known hot-spots such as the *singed* locus; Roiha *et al.*, 1989) whilst avoiding others. Kidwell (1987) collated details of more than one hundred *Drosophila* genes which had been screened for P-element insertions and found a considerable number which seemed refractile to disruption by this method (e.g. *Nap*, *Adh*). It is plausible to suggest that a different transposable element with an independent mechanism of transposition may have a different set of 'cold spots' and therefore may be effective in obtaining insertions into genes which had previously proved refractile to targeted transposon mutagenesis.

This project set out to develop a SSM system based upon the utilisation of the I factor; another transposable element known to cause hybrid dysgenesis which had been studied extensively in *Drosophila* (for a review, see Finnegan, 1989).

Before its identification as a transposable element, the I factor (Picard, 1976) was investigated as an inducer of female sterility. Studies of inheritance of the I factor showed that it could be linked to any of the three major chromosomes of *Drosophila* and would produce markedly elevated female sterility in the progeny of crosses between males carrying this I factor and females contributing a more elusive R factor. In the years since this work, various studies have combined to give us a more complete picture of the basis of I factor induced (I-R) hybrid dysgenesis. In 1984, Sang *et al* identified the molecular lesions associated with several mutations at the *white* locus of *Drosophila melanogaster* arising after an I factor induced dysgenic cross. Four of these lesions were associated with insertions of 5.4kb within the *white* locus and once cloned (Bucheton *et al.*, 1984), it became apparent that this 5.4kb segment of DNA represented a functional I factor. The complete element was present only in Inducer strains and never in Reactive strains. Sequencing of the I factor (Fawcett *et al.*, 1986), suggested that it was able to encode a polypeptide with strong homology to a viral reverse transcriptase. This finding indicated that the I factor was a retrotransposon and gave credence to the theory that the I factor was a transposable element and that hybrid dysgenesis associated mutations were a consequence of its transposition. In recent years, evidence for transposition of I factors has been found (Chaboissier *et al.*, 1990) and proof that the mode of transposition is *via* an RNA intermediate has come from studies using I factors with engineered changes (Pelisson *et al.*, 1991; Jensen & Heidmann, 1991). The elusive R factor postulated by Picard is merely the cytoplasmic environment contributed by females of a reactive strain. Some inhibitor of transposition (perhaps made by the I factor itself - the nucleic acid binding product of the first open reading frame has been implicated by Chaboissier *et al.*, 1990) must accumulate in the cytoplasm of inducer flies. In reactive flies, this inhibitor does not accumulate and any introduced I factors (in a fertilising sperm) are freed from repression because the tiny amount of inhibitor in sperm cytoplasm is diluted by egg cytoplasm. This would explain why the reciprocal cross (with an inducer egg and reactive sperm) does not lead to significant levels of transposition; because the cytoplasmic inhibitor present in the egg is not sufficiently diluted.

3.2 I factor Site Selected Mutagenesis

3.2.1 The Model System

The aim of this study was to develop a model system utilising previously well characterised strains bearing I factors at the *white* locus. Two *Drosophila* strains w^{IR1} and w^{IR5} were obtained from David Finnegan (University of Edinburgh, UK). Each of these strains carried an I factor at a known position within the *white* locus (Figure 3.2; Bucheton *et al.*, 1984, Sang *et al.*, 1984). As can be seen from Figure 3.2, the *white* locus ORF extends over around 6kb of the *Drosophila* genome. The brown eye colour conferring mutations w^{IR1} and w^{IR5} are I factor insertions at the 3' end of the *white* locus. The insertion responsible for the w^{IR1} phenotype is within the 3' untranslated region of the locus whilst the w^{IR5} mutation is caused by an insertion within the last intron. Thus both these insertions, though phenotypic, are outside protein coding sequence. Two other I factor insertions within *white* (w^{IR3} and w^{IR4}) disrupt the gene at the same nucleotide as the w^{IR1} insertion. This region at the 3′ end of the *white* locus may be a hot spot for I factor insertion in the same way as the small region of the *singed* locus identified by Roiha *et al.*, 1988 is thought to be a hot spot for P-element insertion. Such a concentration of insertions at the 3' end of this gene may be further evidence that I factors have a different insertion site preference to P-elements.

The region between the two insertion sites (nucleotides -1435 to -2125 using the co-ordinate system of O'Hare *et al.*, 1984) was examined and a pair of 25 base oligonucleotides were chosen to act as gene specific primers. GSP-1 (nt -1793 to -1769) points toward the site of insertion in w^{IR1} and would detect any new insertions upstream (toward the 5' end of the gene). Meanwhile GSP-5 (nt -1802 to -1826) points toward the site of insertion in w^{IR5} and will detect new insertions at the 3' end of the gene. These primers were designed to have similar GC-content and similar melting temperatures allowing them to be used in multi-primer PCR reactions.

In parallel, sequences at the ends of the I factor were examined and two suitable PCR primers (again with matching T_m's and GC content) were designed 'pointing out' from the left and right ends of the retrotransposon. In theory, this would allow detection of new I factor insertions in either orientation with

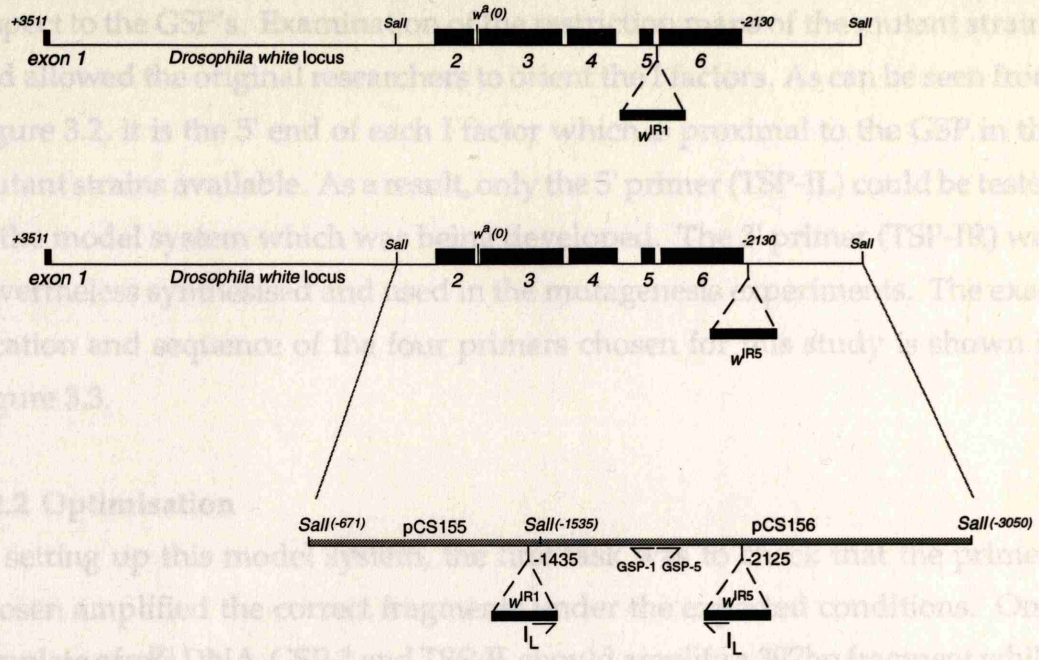
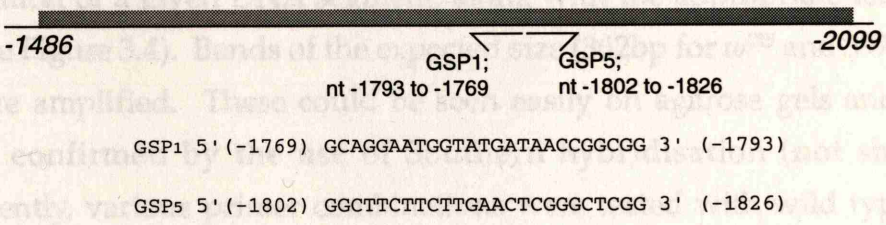


Figure 3.2 *Location of I Factors resident at the white Locus in Strains w^{IR1} and w^{IR5}*
Precise location and relative orientation of insertions was determined by Fawcett *et al.*, 1986. During this study the *SalI* fragment pCS156 was used as a probe for Southern hybridisation. This fragment stretches from nt -1535 to -3050 of the *white* locus sequence (O'Hare *et al.*, 1984). Approximate locations of the GSP's used in this study are shown.

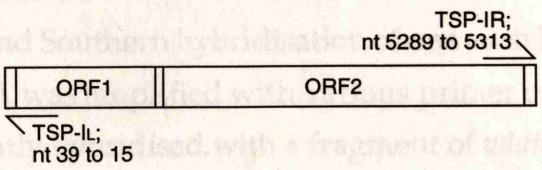
(a)



GSP1 5'(-1769) GCAGGAATGGTATGATAACCGGCGG 3' (-1793)

GSP5 5'(-1802) GGCTTCTTCTTGAACTCGGGCTCGG 3' (-1826)

(b)



TSP-IL 5'(39) GAGGCACGACTTATCTCTTCGGAGG 3' (15)

TSP-IR 5'(5289) TAGCTGTAAGCCCCGTAGCTAATGC 3' (5313)

Figure 3.3 *Gene Specific and Transposon Specific Primers Used in This Study*
(a) Sequence of the Gene Specific Primers used in this study, along with relative locations of the I factors resident in the strains w^{IR1} and w^{IR5}. The sequences shown were derived from the *white* locus published by O'Hare *et al.*, 1984. (b) Sequence of the transposon specific primers chosen for this study. The complete I Factor is 5371bp in length and contains 2 ORF's. The sequences shown here were derived from the published sequence of the I Factor (Fawcett *et al.*, 1986).

respect to the GSP's. Examination of the restriction maps of the mutant strains had allowed the original researchers to orient the I factors. As can be seen from Figure 3.2, it is the 5' end of each I factor which is proximal to the GSP in the mutant strains available. As a result, only the 5' primer (TSP-IL) could be tested in the model system which was being developed. The 3' primer (TSP-IR) was nevertheless synthesised and used in the mutagenesis experiments. The exact location and sequence of the four primers chosen for this study is shown in Figure 3.3.

3.2.2 Optimisation

In setting up this model system, the first task was to check that the primers chosen amplified the correct fragments under the expected conditions. On a template of w^{IR1} DNA, GSP-1 and TSP-IL should amplify a 397bp fragment whilst the primer combination of GSP-5 and TSP-IL should generate a 362bp fragment when used with a w^{IR5} DNA template. In theory, no other DNA/primer combination should yield any band.

Initially, PCR was carried out using just the pair of primers needed for amplification of a given DNA segment, along with the appropriate template DNA (see Figure 3.4). Bands of the expected size (362bp for w^{IR5} and 397bp for w^{IR1}) were amplified. These could be seen easily on agarose gels and their identity confirmed by the use of Southern hybridisation (not shown). Subsequently, various primer combinations were tested with wild type and mutant DNA to examine the source of spurious amplification products which appear when all four primers were used in the same PCR reaction. Figure 3.5 shows an agarose gel and Southern hybridisation of one such PCR experiment. 100 nanograms of DNA was amplified with various primer combinations (see legend) and subsequently hybridised with a fragment of *white* locus DNA (for details of the probe see Figure 3.2 and Section 2.6) Clearly, although extra bands of various sizes and intensities are present, none of these extra bands hybridise and should not therefore present a problem of interpretation during a mutagenic screen. Extra bands can be associated with specific primers or primer pairs but should be present in all (or almost all) templates in contrast to bands corresponding to new insertions which shall be specific to one of a number of templates under test. It is important to consider the source of the spurious bands associated with I factor primers (TSP-IL alone, TSP-IR alone and TSP-IL,

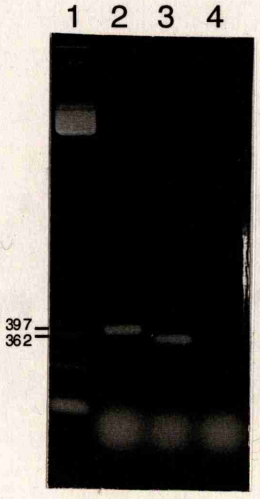


Figure 3.4 *Initial Amplification of DNA from Mutant Strains.*
1.5% Ethidium Bromide stained TBE Agarose gel showing amplification products obtained using mutant fly DNA as a template. Lane 1: 123bp Ladder. Lane 2: 100ng w^{IR1} amplified with the primers GSP-1 and TSP-IL. Lane 3: 100ng w^{IR5} amplified with the primers GSP-5 and TSP-IL. Lane 4: *Luminy* (wild type) DNA amplified with GSP-5 and TSP-IL.

(a)

Lane	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
TSP-IL	✓	✓	✓	✓				✓	✓	✓	✓	✓	✓	✓			
TSP-IR	✓	✓	✓		✓				✓			✓			✓		
GSP1	✓	✓	✓			✓				✓	✓	✓	✓			✓	
GSP5	✓	✓	✓				✓	✓	✓	✓			✓				✓
DNA	-	Lu	wIR5								wIR1						

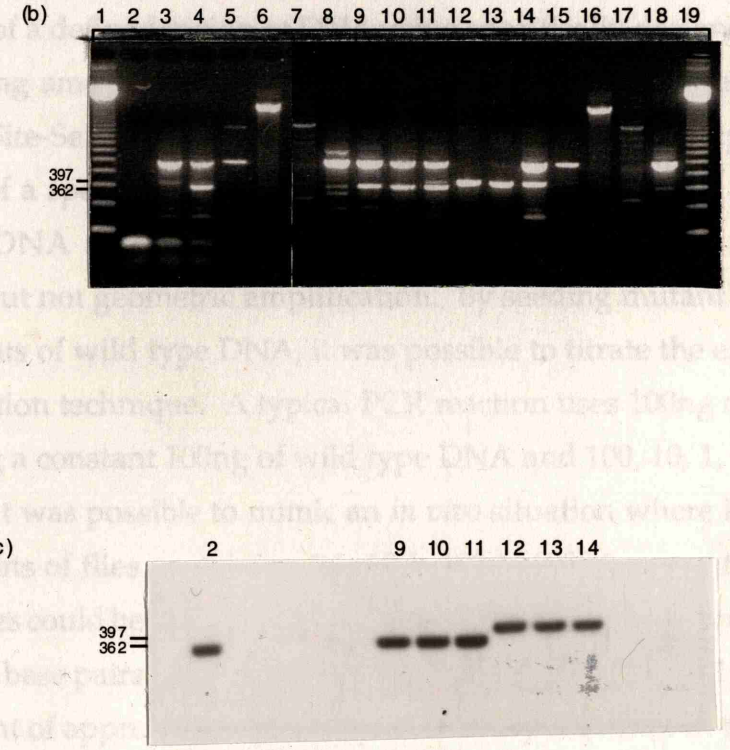


Figure 3.5 *Exploration of the Origin of Spurious Bands.*

Various primer-DNA combinations were amplified in standard PCR reactions as tabulated in (a). Lane 1,19: 123bp ladder. Lane 2: All 4 primers, no DNA. Lane 3: All 4 primers, wild type DNA (*Luminy*). Lane 4: All 4 primers, w^{IR5} DNA. Lane 5-8: Each primer individually, w^{IR5} DNA. Lane 9: Primers GSP-5, TSP-IL, w^{IR5} DNA. Lane 10: Primers GSP-5, TSP-IL, TSP-IR, w^{IR5} DNA. Lane 11: Primers GSP-1, GSP-5, TSP-IL, w^{IR5} DNA. Lane 12: Primers GSP-1, TSP-IL, wIR1 DNA. Lane 13: Primers GSP-1, TSP-IL, TSP-IR, w^{IR1} DNA. Lane 14: Primers GSP-1, GSP-5, TSP-IL, w^{IR1} DNA. Lane 15-18: Each primer individually, w^{IR1} DNA. (b) Ethidium Bromide stained 1.5% TBE Agarose gel. (c) Autoradiograph of a Southern blot of this gel hybridised with pCS156, labelled with ^{32}P dCTP.

TSP-IR together). In an inducer strain, 10-15 intact I factors are present, scattered throughout the genome. An amplification product from a PCR reaction containing only I factor primers can be generated from a template consisting of two I factors situated within several hundred base pairs of each other. This situation is important during a mutagenesis because, assuming random transposition, for every new I factor insertion within range of a GSP, there will be 10-15 within range of a resident I factor. Whilst likely to be visible on an agarose gel, these bands should not hybridise to a gene specific probe and should not impair our ability to interpret results.

3.2.3 Sensitivity of Detection.

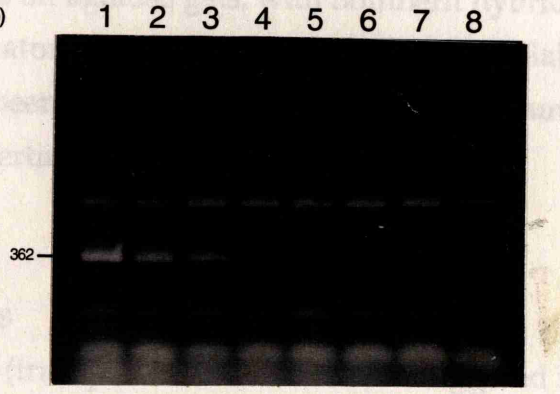
Amplification of a defined region of DNA using specifically designed primers and non-limiting amounts of template DNA is a routine procedure, but the technique of 'Site-Selected' Mutagenesis depends upon an ability to detect amplification of a specific template sequence within a population containing an excess of DNA sequences which can serve as templates for linear amplification but not geometric amplification. By seeding mutant DNA with varying amounts of wild type DNA, it was possible to titrate the efficiency of this PCR detection technique. A typical PCR reaction uses 100ng of template DNA. By using a constant 100ng of wild type DNA and 100, 10, 1, 0.1 ... ng of mutant DNA, it was possible to mimic an *in vivo* situation where DNA made from populations of flies containing 1:1, 10:1, 100:1, 1000:1, etc. ratios of wild type:mutant flies could be tested. The haploid genome of *Drosophila melanogaster* is ~165,000,000 base pairs and weighs approximately 0.4pg. Thus 1pg of DNA is the equivalent of approximately 2.5 copies of every sequence in the genome and 100ng corresponds to 250,000 copies of each sequence (see the table in Figure 3.6).

Figure 3.6 shows an Ethidium Bromide stained agarose gel and corresponding Southern hybridisation from an experiment where w^{IR5} DNA was diluted into a constant 100ng of wild type DNA (from the *Luminy* inducer strain). Amidst a smear of bands, the 362bp fragment corresponding to the w^{IR5} template can be seen on the agarose gel in lanes 1, 2, 3 and 4 - indicating visual detection of amplification down to the level of 1/1,000 flies. Using Southern hybridisation, this sensitivity can be extended to at least 1/10,000. This shows that Southern

(a)

Lane	Luminy DNA	Genome Equivalents	wIR5 DNA	Genome Equivalents	Ratio
1	100ng	250,000	100ng	250,000	1:1
2	100ng	250,000	10ng	25,000	10:1
3	100ng	250,000	1ng	2,500	100:1
4	100ng	250,000	100pg	250	1,000:1
5	100ng	250,000	10pg	25	10,000:1
6	100ng	250,000	1pg	2.5	100,000:1
7	100ng	250,000	0.1pg	0.25	1,000,000:1
8	100ng	250,000	NONE	-	-

(b)



(c)

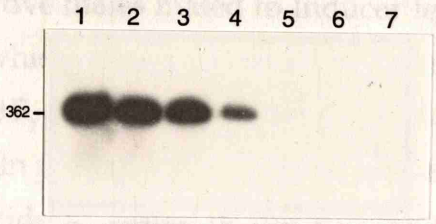


Figure 3.6 *Sensitivity of Detection*

A constant 100ng of wild type (*Luminy*) DNA was mixed with varying amounts of w^{IR5} DNA as shown in (a). These mixes were then amplified in the presence of all four primers in a standard PCR reaction. The amplification products were separated on a 1.5% Agarose gel and stained with Ethidium Bromide (b). Products were then transferred to Nylon membrane and hybridised with the pCS156 DNA fragment as described before. (c) shows an autoradiograph exposed for four hours. For loading order, see (a).

hybridisation, though more labour intensive and less immediate, is a more sensitive system of detection and would be the method of choice during a mutagenesis. In addition, Southern hybridisation allows the identification of non-spurious bands from within a population of amplification products. As shall be seen later, this feature is essential for the correct interpretation of results during a mutagenesis. The choice of primers unique to each end of the I factor allows an additional means of confirmation of any new insertion. After the first round of PCR and hybridisation, the size of positively hybridising bands may be estimated. Subsequently, repeating the PCR reaction using specific primer pairs should confirm that a new insertion is dependent upon one TSP and one GSP only. Thereafter in the screening process, the use of just two primers will minimise the appearance of spurious bands and possibly allow detection of diagnostic bands on agarose gels, with Southern hybridisation being used only as a confirmatory tool. Once sensitive and reliable conditions for amplification had been determined, a mutagenesis was undertaken to search for new I factor insertions at *white*.

3.3 Mutagenesis

Hybrid dysgenesis (transposition of I factors) is triggered in the germ line of female offspring (SF Females) from a cross between males from an Inducer strain (e.g. *cha*RC$^+$) and females from a Reactive strain (such as *charolles; cha*). The reciprocal cross (Reactive males mated to Inducer females) yields female offspring (RSF Females) which show only minimal signs of dysgenesis. In a strategy to provide ~10,000 F_2 females over a ten day period the crosses outlined in Figure 3.7 were set up. In parallel to the dysgenic cross, a control reciprocal cross was set up to provide a gauge of the success of the mutagenesis. Hatchability of eggs laid by SF and RSF females was assessed by counting the proportion of eggs that had hatched 48 hours after laying. Of 500 eggs laid by SF Females, just 19 (3.8%) had hatched whilst in RSF females, 372 eggs out of 500 (74.5%) hatched in the first 48 hours. Low hatchability is a sign of hybrid dysgenesis and therefore these results (though admittedly a restricted sample) would indicate that the mutagenic cross was successful and that a significant level of transposition was occurring.

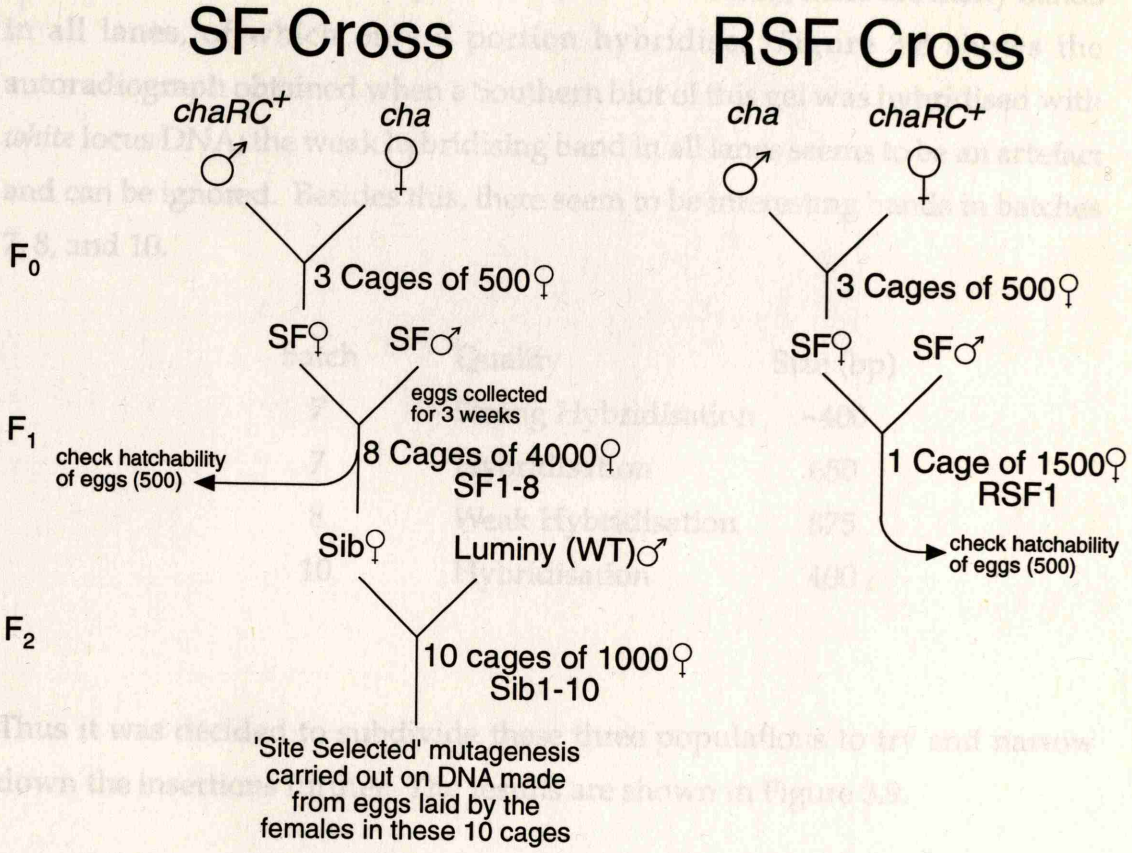


Figure 3.7 *I-R Mutagenesis: Crossing Strategy*

These crosses were designed to generate >10,000 virgin females heterozygous for new insertions over a short period. A vast excess of the F_1 cross (with dysgenic (SF) females) was set up because so few (<5%) eggs hatch from these females. Crosses were carried out in population cages at 25ºC as described in Section 2.2.

Once hatchability estimations were complete, egg collections were made from each of 10 populations of 1,000 F_2 females and 250 wild type males. From these eggs, DNA was made and used as a template for PCR. Figure 3.8 shows the results of these 10 PCR's. As predicted (in Section 3.2.2), there are many bands in all lanes, of which only a portion hybridise. Figure 3.8 shows the autoradiograph obtained when a Southern blot of this gel was hybridised with *white* locus DNA; the weak hybridising band in all lanes seems to be an artefact and can be ignored. Besides this, there seem to be interesting bands in batches 7, 8, and 10.

Batch	Quality	Size (bp)
7	Strong Hybridisation	~400
7	Hybridisation	650
8	Weak Hybridisation	875
10	Hybridisation	400

Thus it was decided to subdivide these three populations to try and narrow down the insertions further. The results are shown in Figure 3.9.

Again although interpretation is hampered by the non-specific hybridisation, an attempt was made to choose positives and subdivide. The conclusions are summarised below.

Batch 7: Subdivide 7.7
Batch 8: Lost?, subdivide batch 8.3
Batch 10: Candidate batches are 10.2 and 10.8

Subdivision of these 4 batches of flies was carried out. The result shown in Figure 3.10 represents subdivision of one batch, 7.7. After hybridisation, none of the expected bands appeared and the experiment was stalled whilst reactions were repeated and solutions re-prepared. Nothing could be salvaged however

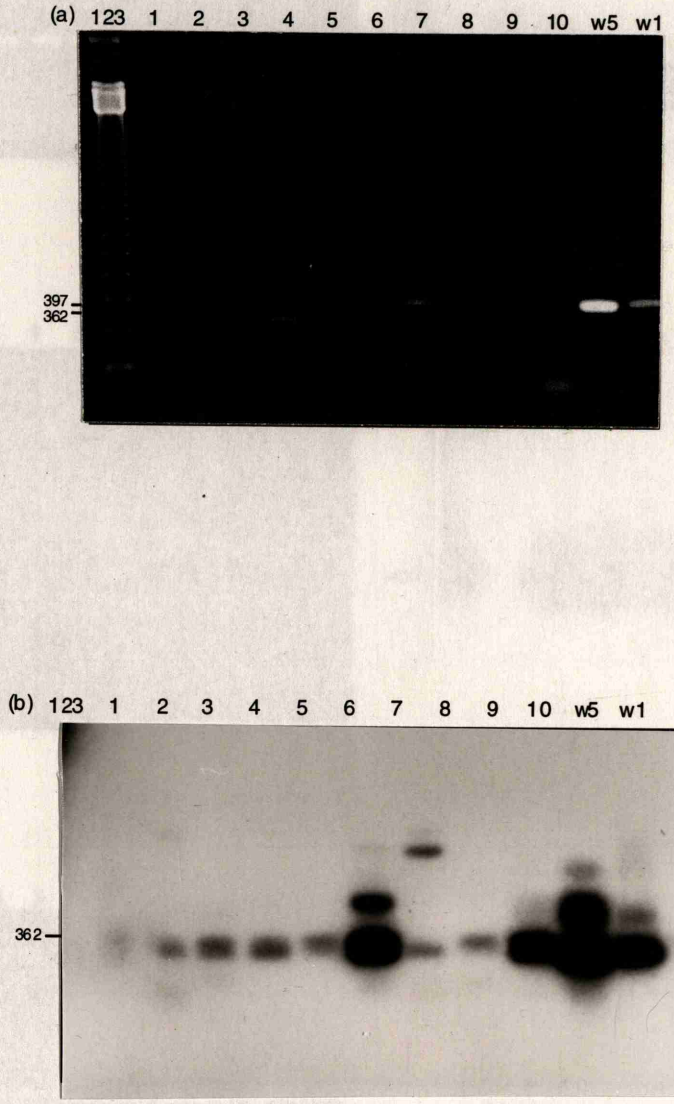


Figure 3.8 *Initial Sib-Selection PCR*

100ng of each DNA made from eggs laid by flies in cages 1-10 was amplified in a standard PCR reaction with all four primers and the products run on a 1.5% Agarose gel as before (a). Southern blotting and hybridisation with pCS156 (as before) gave the result seen in the autoradiograph (b). The weakly hybridising band seen in all lanes was first seen in these PCR's and may be connected with the strains used. DNA from the strains used in the original dysgenic cross had not previously been amplified with these primers. Loading order: 123; 123bp ladder, 1-10;mutagenised DNA's 1-10, w5; w^{IR5} control; w1 w^{IR1} control.

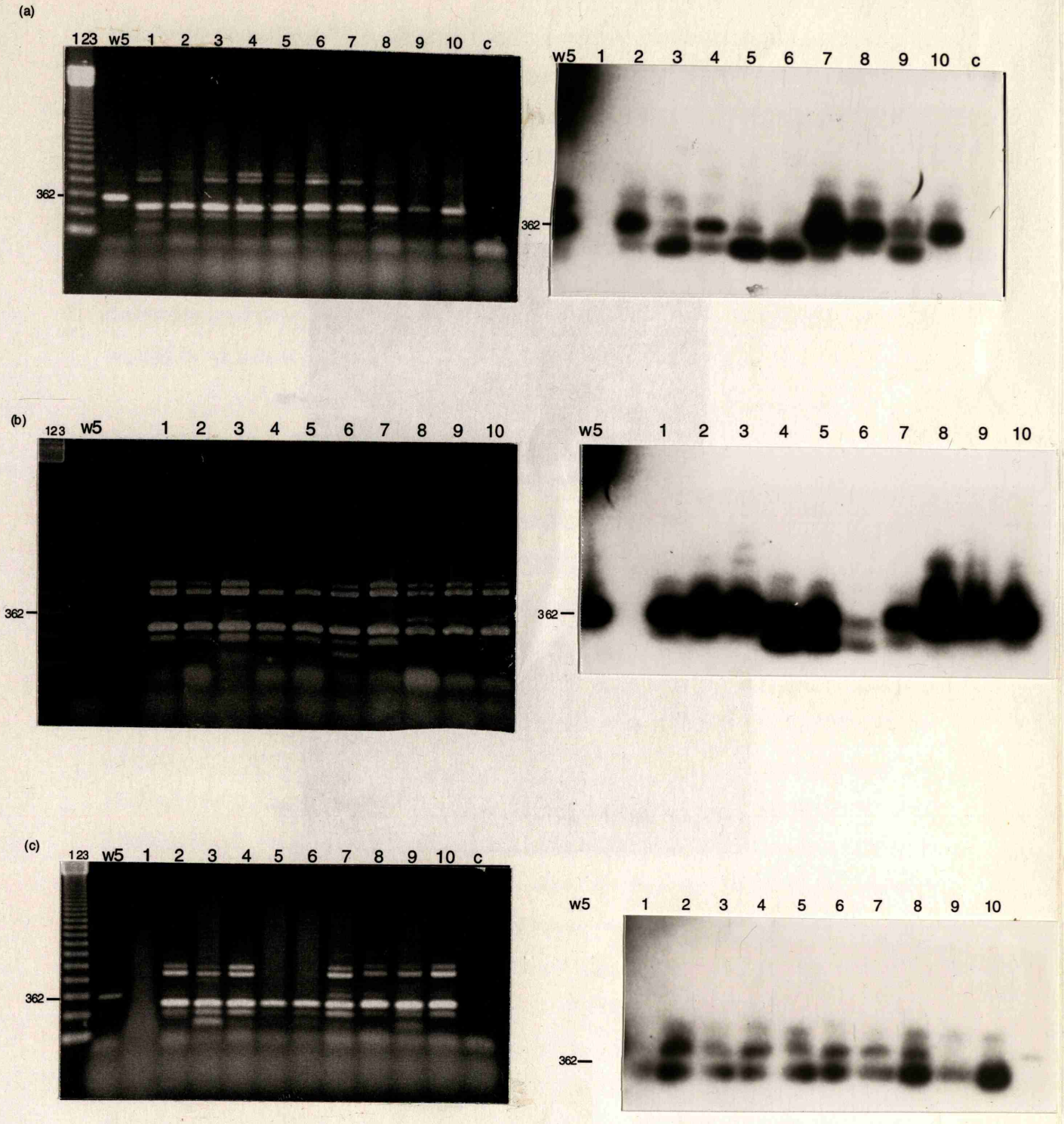


Figure 3.9 *Sib Selection of 'Positive' Batches*

Females from 'positive' cages were subdivided into groups of approximately 100 and allowed to lay for 1 day. DNA was made from the eggs collected in these batches and PCR carried out on the 30 samples, along with control DNA. Shown above are Ethidium Bromide stained gels and autoradiographs showing the results of these amplification reactions. Loading order (a) 123, 123bp ladder; *w5*, w^{IR5} control, 1-10; mutagenised DNA's 7.1-7.10; c, No DNA control. (b) 123, 123bp ladder; *w5*, w^{IR5} control; empty lane; 1-10, mutagenised DNA's 8.1-8.10. (c) 123, 123bp ladder; *w5*, w^{IR5} control; 1-10, mutagenised DNA's 10.1-10.10; c, No DNA control.

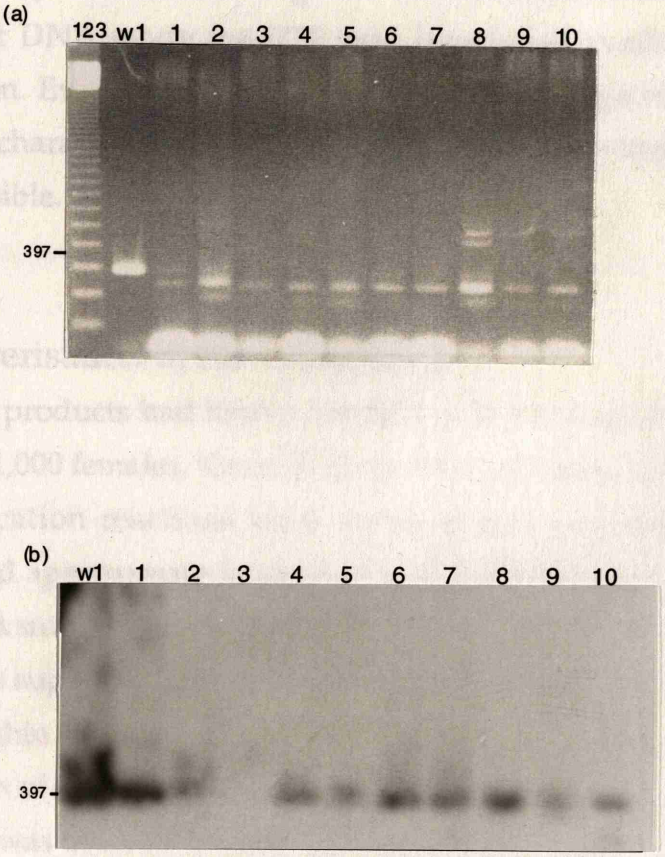


Figure 3.10 *Further Subdivision*

After the identification of positive batches of 100 flies (Figure 3.9), the 4 batches chosen for subdivision were treated as before and DNA made from the eggs laid by groups of ~10 females. (a) shows a 1.5% Agarose gel of amplification products of one such batch: 123, 123bp ladder; *w*1, w^{R1} control; 1-10, mutagenised DNA's 7.7.1-7.7.10. (b) shows an autoradiogaph of the same gel. No significant hybridisation is seen aside from the positive control and the weakly hybridising band seen in all lanes. On DNA made from batches of 10 flies, a single positive would be expected to give a band of almost equal intensity to that given by the positive control.

and even repeating PCR on previously positive batches yielded no results. Although not known at the time, it later became apparent that the PCR machine used during this mutagenesis had broken down around this time and inconsistencies in the results could in part have been due to the malfunctioning of the machinery. No flies bearing new I factor insertions were recovered therefore, but DNA made for PCR was intact and available for further characterisation. Even without a surviving fly line showing a visible phenotype, detection and characterisation of new I factor insertions in these pools of DNA would be possible.

3.4 Characterisation of New Insertions

Amplification products had been generated from DNA made from eggs laid by batches of 1,000 females. Once PCR machine problems had been resolved, these amplification reactions were repeated and extended to determine orientation and approximate location of new I factor insertions within these populations. A summary of these reactions are shown in Figure 3.11. Analysis of these results suggested that there had been four detectable insertions at the *white* locus within these mutagenised populations. The approximate location and orientation of these insertions are shown in Figure 3.11b. To confirm these predictions, it was decided to clone these four amplification products into pBS KS$^-$. Plasmid DNA was restricted with *Sma*I to generate blunt ended vector. Amplification products were run on LMP agarose, bands cut out and the DNA purified as described in Chapter 2 (Section 2.13). Vector and insert were mixed at a molar ratio of >3:1 (50ng insert:100ng vector) and ligated overnight at 14°C. Competent XL1-Blue were transformed with ligation products and incubated overnight at 37°C on L-Agar containing X-gal and IPTG. DNA was prepared from several recombinant clones and plasmids restricted with *Kpn*I and *Sst*I; enzymes which would remove the whole insert. Recombinant clones with inserts of the same size as the PCR products were chosen for further analysis; as shown in Figure 3.12.

The amplification products which have been cloned should have GSP-1 (7w1a and 7w1) or GSP-5 (10w5 and 7w5) sequences at one end and I factor sequences at the other. In addition, one of the two breakpoints of insertion will have been

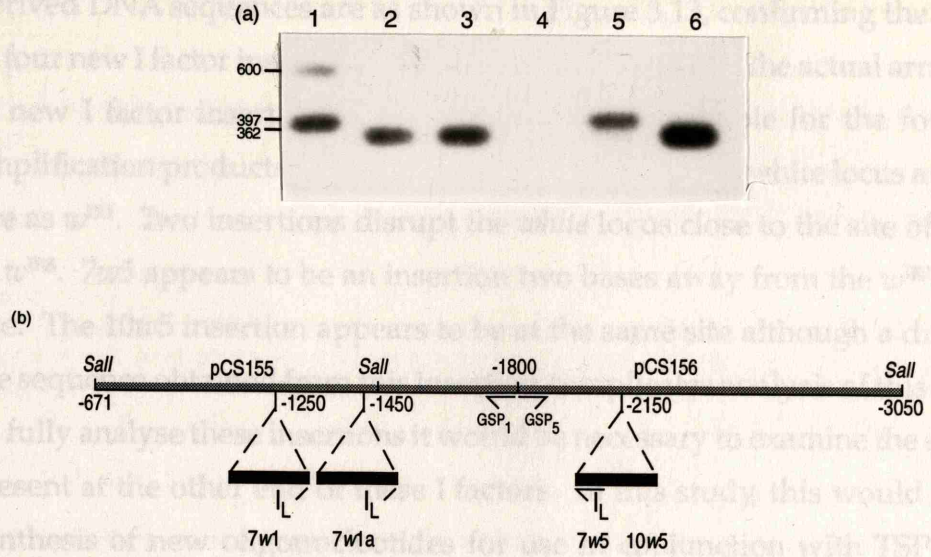


Figure 3.11 *Approximate Location of New Insertions at white*
The batches of DNA thought to contain positive bands were amplified with various combinations of primers. The three reactions shown above represent the repeatable results obtained. From the primer combinations necessary to generate these amplification products, and the size of the bands found on gels, the approximate location of these insertions can be predicted. (a) Autoradiograph showing positively hybridising fragments obtained from the reactions described: 1, DNA 7 with GSP-1 and TSP-IL; 2, DNA 7 with GSP-5 and TSP-IL; 3, DNA 10 with GSP-5 and TSP-IL;4, empty track; 5, w^{IR1} with GSP-1 and TSP-IL; 6, w^{IR5} with GSP5 and TSP-IL. (b) Schematic representation of the results obtained in (a).

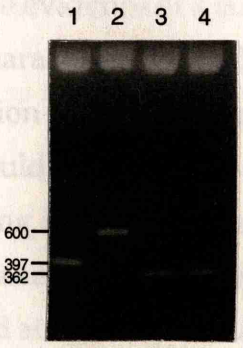


Figure 3.12 *Recombinant Clones Representing the New Insertions at white*
Recombinant clones were selected on the basis of size and hybridisation pattern. The Agarose gel here shows clones restricted with *Kpn*I and *Sst*I. 1; 7w1a, 2; 7w1, 3; 7w5, 4; 10w5.

cloned for each insertion. The ends of the cloned amplification products were therefore sequenced using primers flanking the cloning site in pBS KS$^-$. The derived DNA sequences are as shown in Figure 3.13, confirming the detection of four new I factor insertions at *white*. Figure 3.14 shows the actual arrangement of new I factor insertions at the *white* locus responsible for the four cloned amplification products. The insertion 7*w*1a disrupts the white locus at the same site as w^{IR1}. Two insertions disrupt the *white* locus close to the site of insertion of w^{IR5}. 7*w*5 appears to be an insertion two bases away from the w^{IR1} insertion site. The 10*w*5 insertion appears to be at the same site although a disparity in the sequence obtained from this insertion complicates analysis of this insertion. To fully analyse these insertions it would be necessary to examine the sequences present at the other end of these I factors. In this study, this would entail the synthesis of new oligonucleotides for use in conjunction with TSP-IR. This analysis was judged to be unnecessary for this project. A fourth insertion 7*w*1 disrupts the *white* locus in the third intron (according to the predicted structure of O'Hare *et al.*, 1984). As discussed, the w^{IR1} site of insertion is already recognised as a potential hot spot for P-element insertion; the isolation of another insertion at this site is further evidence for this theory. The isolation (in two different batches) of new insertions close to the w^{IR5} insertion site might indicate that this too is a potential hot spot for I factor insertion. No insertions had been previously characterised at the site of the insertion 7w1.

3.5 Conclusions

Although without the tangible evidence of a surviving fly line with a *white* eye phenotype, this study has characterised a molecular phenotype, proving the presence and efficient detection of new I factor insertions at *white* after an I factor mutagenesis. Steps could be built into the experimental design which would safeguard against losing flies bearing new insertions. For instance, an extra egg collection could be taken from all positive cages. These eggs could be reared and the flies subdivided at eclosion. DNA could be made from the adults in these sub-populations after a few days laying, and this DNA could be subsequently screened using PCR as described above. This inefficient sib-selection procedure could be followed (with some effort) to a single fly, a descendant of the original mutagenised female.

(a)

```
7w1   [–GGCACGACTTATCTCTTCGGAGG]TTGAAGTGGTACTGACGAGTAAnATTTTAATTTGCAGA
                                ||||||||||||||||||||||
white                           AATATTAATGAGATGCGAGTAACATTTTAATTTGCAGA
```

(b)

```
7w1a  [–GGCACGACTTATCTCTTCGGAGG]TTGAAGTGGTAATGATTTGCATA-TAATAATTTACTAA
wIR1                            TTGAAGTGGTAATGATTTGCATATTAATAATTTACTAA
                                ||||||||||||||||||||||||||
white                           CTTGTTTAGAATACATTTGCATATTAATAATTTACTAA
```

(c)

```
7w5   [-GGCACGACTTATCTCTTCGGAGG]TTGAAGTGGTACTG-AGTTATTTCGGATATCGGATATA
10w5  [----ACGACTTATCTCTTCGGAGG]TTGAAGTGGTACTGAGATTATTTCGGATAT
wIR5                            TTGAAGTGGTACTG...TTATTTCGGATATCGGATATA
                                ||||||||||||||||||||||||||
white                           AAAAAAAAAACAAGCAGTTATTTCGGATATCGGATATA
```

Figure 3.13 *Sequence Analysis of the Four Recombinant Clones.*
To accurately localise their insertion sites, the ends of the four recombinant clones were sequenced and compared to the published *white* locus and I Factor sequences. Sequences determined from the end of the amplification product associated with the GSP agreed with the published sequence and are not shown. Sequences determined from the other end show I Factor sequence followed by *white* locus DNA. Inconsistencies with the published sequence are shown in lower case. I Factor sequences are underlined. TSP-IL primer sequence is enclosed by square brackets.
(a) 7w1. Junction point of the insertion 7w1 aligned to white locus DNA (nt -1249 to -1287). At the junction point, one base is found which is not present in either the nascent *white* locus sequence or the I factor. (b) 7w1a. Junction point of the insertion 7w1a aligned with *white* locus DNA (nt -1422 to -1460) and w^{IR1} insertion. Although the junction is the same, these sequences differ by one base at the 3' end of the known target site duplication in w^{IR1}. To discover whether this difference is a consequence of insertion, the other junction point must be sequenced. (c)7w5 and 10w5. Junction points aligned to w^{IR5} insertion site and nascent white locus sequence (nt -2138 to -2107). The insertion 7w5 appears to be two nucleotides away from the w^{IR5} insertion. The insertion 10w5 possesses a further one nucleotide not found in the original *white* locus or I factor sequences. The origin of this nucleotide is unclear and would require sequencing of the other junction point, along with resequencing of this junction point, to facilitate clarification of its true origin.

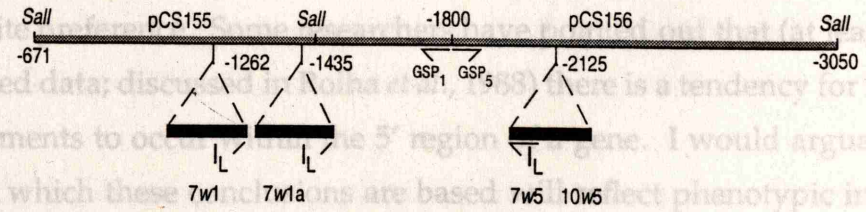


Figure 3.14 Exact Location of New I Factor Insertions at *white*
Sequence analysis (Figure 3.13) determined the exact location of new I Factor insertions at white. The insertion 7w1a is at the same site as the w^{IR1} insertion. The insertions 7w5 and 10w5 are at the same site as w^{IR5}. The insertion 7w1 represents a novel insertion.

This study has proven that I factor mutagenesis is efficient to carry out and should require no more work than a P-element mutagenesis of a similar scale. The methodology described here can efficiently detect new insertions within populations of at least 1,000 mutagenised flies. As a result, screens utilising wild populations (similar to the nature screen of Clark *et al.*, 1994) are possible. Although studied in detail for more than 20 years now, the range of phenotypes induced after hybrid dysgenesis are not yet fully understood and optimisation of a mutagenic cross is difficult. For instance, the age (at mating) of reactive females over a number of generations prior to the mutagenic cross is thought to contribute to the degree of transposition triggered (Bucheton, 1979). These and other considerations make the screening of wild populations (which if properly chosen can comprise large numbers of independent insertions) an attractive prospect. It should be remembered of course that the insertions present in a wild population are unlikely to represent severe or deleterious mutations as these types of mutation will be strongly selected against in wild populations if they appear as homozygotes. One of the main advantages of the original SSM method is the ability to detect new insertions before they have become homozygous.

The initial basis for this research is that I factors might possess a different insertion site preference to P-elements. The mechanism of transposition differs greatly between the two elements. P-elements move by a non-replicative excision from one site and invasion of another (Engels *et al.*, 1990). I factors do not excise; instead, an RNA species corresponding to the whole 5.4kb element is transcribed using an internal promoter (McClean *et al.*, 1993). This RNA species is thought to invade DNA at a double strand break and is then reverse transcribed. The enzymes which mediate the invasion events are structurally unrelated. Of the many P-element insertions analysed, there seems to be no target site preference. Some researchers have pointed out that (at least within published data; discussed in Roiha *et al.*, 1988) there is a tendency for insertion of P-elements to occur within the 5′ region of a gene. I would argue that the data on which these conclusions are based will reflect phenotypic insertions. P-element insertions into heterochromatic or intergenic regions are unlikely to elicit a phenotype. Likewise, insertion into introns might not disrupt translation as P-elements used for mutagenesis (e.g. *Birm*-2 elements; Robertson *et al.*, 1988)

are often relatively small (<2kb). A perturbation in transcription caused by disruption of a promoter region after P-element insertion would be most likely to cause a mutant phenotype. So few I factor mutageneses have been carried out and so few mutations have been analysed that it is difficult to form conclusions from the small data set. The six *white* locus insertions characterised by Fawcett *et al.*, (1983) all fell within coding regions, 3' untranslated regions or introns. That insertions in introns and 3' untranslated regions are phenotypic insertions may reflect the larger size (5.4kb) of the inserted element. Analysis of another set of I factor mutations at the *yellow* locus (Busseau *et al.*, 1989) revealed a tendency for rearrangements associated with I factor transposition. This might be a consequence of the I factors mode of transposition which entails a double strand break. As reported by Crozatier *et al* (1988), the *Drosophila melanogaster* genome is filled with incomplete I factor sequences (I elements). A number of these defective elements lack their 5' end. This type of defective element can arise if the process of reverse transcription is interrupted. This might explain why the I element sequences found at the breakpoints of the rearrangements analysed by Busseau *et al* correspond to incomplete I element sequences.

A third transposable element, *hobo* (Blackmann & Gelbart, 1989) can also be induced to transpose under appropriate genetic conditions. The H-E system (*hobo* containing and Empty strains) is similar to the other hybrid dysgenesis systems described above. The *hobo* element itself is a member of the same family of transposons as the P-element. It possesses short inverted terminal repeats and is known to transpose by a cut-and-paste mechanism similar to that utilised by the P-element. The exact nature of control of transposition and *hobo* activity is however still unknown. Consequently efficient mutagenic crosses cannot be set up. Once more is known of the biology of *hobo*, this transposon shall be another suitable mutagenic element through which mutations may be targeted using 'site-selected' mutagenesis.

The main disadvantage to using I factors and *hobo* however is the relative lack of engineered elements. Aside from the constructs made to investigate its mode of transposition, no engineered I factors have been constructed. In theory, mutator and ammunition elements (like those used by Robertson *et al.*, 1988)

may be easily engineered. In addition, elements similar to the enhancer trap P-element constructs which carry exogenous genes and plasmid replication origins (discussed in Chapter One) could be made. The range of intricately engineered and imaginatively designed P-elements already available might make re-engineering I factor equivalents somewhat pointless. The rate at which new P-element insertions are being generated and collected in stock centres and laboratories around the world ensures their continued dominance as tools for dissection of the *Drosophila* genome. That I factors transpose by a replicative mechanism also makes them less appealing than P-elements as genetic tools. A novel means of obtaining transposon induced mutations is to find a strain containing a P-element close to the gene under study. This P-element can be easily remobilised. Although in most cases, sister chromatid repair after P-element excision retains a copy of the P-element at the original site, P-elements are lost in approximately 10% of excision events (Engels *et al.*, 1990; Gloor *et al.*, 1991). In a portion of these, some genomic DNA will also be lost as the P-element excises. By denying sister chromosome repair, the frequency of imprecise excision can be increased considerably. Due to the I factors duplicative mode of transposition, re-excision never occurs and this secondary strategy is not possible for I factor insertions. It is these reasons, rather than its efficiency as a mutagenic element which argue most against the routine use of I factors as mutagenic elements. Certainly, the I factor should not be ignored as a potential tool for creating mutations. For the time being, many researchers continue to study I factors and it is their findings which will determine the extent to which it is utilised in future.

Chapter Four

Generation and Screening of the *eya* Head cDNA Library

4.1 Introduction

The primary objective of this project was to initiate the study of a number of cloned DNA sequences selected on the basis of their expression pattern. As discussed in Chapter One, this approach should provide an effective means of studying genes which function within the *Drosophila* brain. An approach which selects cDNA clones in this way is a powerful tool in the study of *Drosophila* where molecular technologies which can help us gain knowledge about these clones are already well established. As we have seen in Chapters One and Three, several tools which can augment this 'reverse genetic' screening approach are currently available and powerful new tools are constantly being developed.

The decision as to the specific approach used in this study was governed by a number of factors. A differential approach (St. John & Davis, 1979; Sargent & Dawid, 1983) was chosen. This method (screening an un-subtracted library with cDNA probes made from different RNA populations and selecting clones which react with only one probe) was judged to be the most direct means of obtaining new clones. In this respect, the screening strategy most closely resembles that of Levy *et al* (1982). It was decided however, that for this study, a cDNA library constructed using RNA derived from head tissue should be screened. As a result, the design of this screen differs in two important aspects from that study. These differences constitute significant improvements in the design of the screen. First, by screening a cDNA library rather than a genomic DNA library subsequent screening of the differentially expressed clones is facilitated. This is because no initial characterisation of the clones is needed to identify the coding region. A second advantage of using cDNA as the source is that careful choice of starting material can favour a specific type of expression pattern. In this study for example, the use of head cDNA functions as a pre-screen, enriching the library for sequences expressed in the head before the differential screen is carried out.

A final, important feature of the current strategy which distinguishes it from previous studies is that it utilises (as starting material) non-wild type cDNA derived from *eyes absent (eya)* flies. This strain of *Drosophila* lacks the optic lobes and much of the associated neural tissue (Sved, 1986). The use of *eya* head RNA in the construction of the library overcomes what may be considered

the major disadvantage of the original differential screen carried out by Levy *et al.*, (1982). As observed by Fryxell & Meyerowitz (1987), most clones selected by the Levy differential screen (using a genomic library and screening with wild type DNA) were specific to the eye or generally expressed throughout the brain. The use of *eya* head RNA (whilst not biasing this screen against genes expressed throughout the head) allows this screen to avoid those genes (such as the opsins cloned by Zuker *et al.*, (1985) and Shieh *et al.*, (1989)) which are expressed at very high levels only within the repeated structures of the eye.

Aside from the choice of starting material, several other features can influence the type of clones obtained by a differential screen. All were considered before the library was made. Factors influencing the choice of vector and the eventual strategy chosen for construction of the library are discussed below. The screening strategy used in the project, and the rationale behind its choice will be discussed later in this Chapter.

Vector Choice

The vectors used in the construction of the vast majority of cDNA libraries are derived from the bacteriophage λ. We chose to use the most sophisticated and versatile vector available at the time, (λZAPII, Stratagene). This vector (Short *et al.*, 1988) has been extensively engineered and includes appropriate restriction sites to allow directional cloning (a useful feature as will be discussed later). Also, this vector contains sequences from the bacteriophage f1 and the whole of the plasmid vector pBS SK$^-$. In the presence of proteins supplied by a helper phage (e.g. R408; Russel *et al.*, 1986), single stranded phagemids are produced. When these are made double stranded and circularised, they behave as plasmids and may be introduced into cells, the cells grown and plasmid DNA harvested for further study. This 'automatic subcloning' technique greatly enhances the efficiency of the screening process and should in theory reduce the risk of artefacts being studied as there is no actual subcloning (restriction and ligation into a new vector) involved at this stage.

Construction Protocol

The vector λZAPII is supplied as part of a kit. In addition, this kit supplies all the components necessary to construct a cDNA library in the vector using a

protocol derived from that of Gubler & Hoffman (1983). This method, which uses an oligo dT primer to initiate first strand cDNA synthesis and RNAse H to facilitate second strand synthesis has become the method of choice for cDNA library construction and there was no good reason to deviate from this protocol significantly. The reverse transcriptase used in this study was however unusual. The enzyme recommended in the λZAPII library construction protocol was the Moloney Murine Leukaemia Virus Reverse Transcriptase (Mu-MLVRT; Kotewicz *et al.*, 1985). At the outset of this study, an engineered version of this enzyme (SuperScript RT, Gibco BRL; Kotewicz *et al.*, 1988) became available. The engineered version lacked the RNAse H activity found in the original enzyme, and was stable at higher temperatures. RNAse H activity digests the RNA strand of DNA-RNA hybrids. If RNA-oligo dT hybrids are destroyed prior to initiation of second strand cDNA synthesis then the effect upon the RNA population is equivalent to removing a portion of the RNA population. This effectively reduces the complexity of that population. The lack of RNAse H activity in the SuperScript RT enzyme means that there should be no degradation of the template prior to the initiation of first strand cDNA synthesis. The increased thermal stability of the SuperScript RT enables first strand synthesis reaction to be carried out at 45°C instead of the normal 37°C recommended for the native enzyme in the λZAPII protocol. Single stranded RNA is extremely flexible and forms complex secondary structures in solution. These secondary structures interfere with first strand cDNA synthesis but their occurence can be reduced by increasing the reaction temperature. Together, these two features of the SuperScript RT make it significantly more effective as a Reverse Transcriptase for cDNA cloning. The yield of cDNA produced by it is likely to be greater and the cDNAs themselves are more likely to be of full length.

This study used the SuperScript RT, modifying the reaction conditions from those used in the λZAPII cDNA synthesis kit to suit it. The modifications made to the kit protocol relate mainly to the amount of RNA used, the reaction volume and the buffer conditions. A further deviation is during the size fractionation of the cDNA prior to ligation. These alterations will be described in more detail in Section 4.2.2. Parallel to this work, a second cDNA library was made together with my colleague Chris Mackenzie. This library used the same starting material and followed the λZAPII protocol, using the native M-MuLVRT supplied in

the kit. Controls carried out during the construction of that library will be shown for comparison to the SuperScript RT derived library. A schematic representation of the construction of the SuperScript RT cDNA library is shown in Figure 4.1.

4.2 Library Construction

4.2.1 Preparation of Starting Material

The quality of clones in a cDNA library is directly dependent upon the quality of the mRNA used to make that library. Obtaining large quantities of undegraded head RNA and efficiently enriching it for polyadenylated sequences is perhaps the most important process in the construction of a cDNA library. In recent years, many methods have been developed which are geared towards generating this high quality template. For the library, mRNA was prepared as described in Section 2.9, and purified with minimal manipulation by centrifugation through Caesium TriFluoroAcetate. This RNA should be of the highest quality as it has undergone minimal manipulation. The RNA was bound three times with oligo dT cellulose to remove as much non-messenger RNA as possible. Fractions were removed at each stage and the final mRNA obtained was precipitated. Samples of each fraction, as well as the original total RNA were run on a denaturing gel and probed with α-tubulin α1 (Kalfayan & Wensink, 1982) as a control to check the quality of the RNA obtained. This quality check is shown in Figure 4.2. As it was of sufficient quality, construction of the library could begin.

4.2.2 Construction of the Library

Construction of the library followed the λZAPII protocol wherever possible. However much of the first and second strand synthesis was adapted to fit with the conditions recommended by BRL for the use of SuperScript RT. Controls were removed at various points in the process and these will be shown where appropriate. For an overview of the whole process refer back to Figure 4.1.

First Strand Synthesis

3.5µg Poly A^+ RNA was mixed (in a final volume of 20µl buffer; 10mM DTT, 50mM Tris-Cl pH8.3, 75mM KCl, 3mM $MgCl_2$) with 2µg oligo dT primer (see

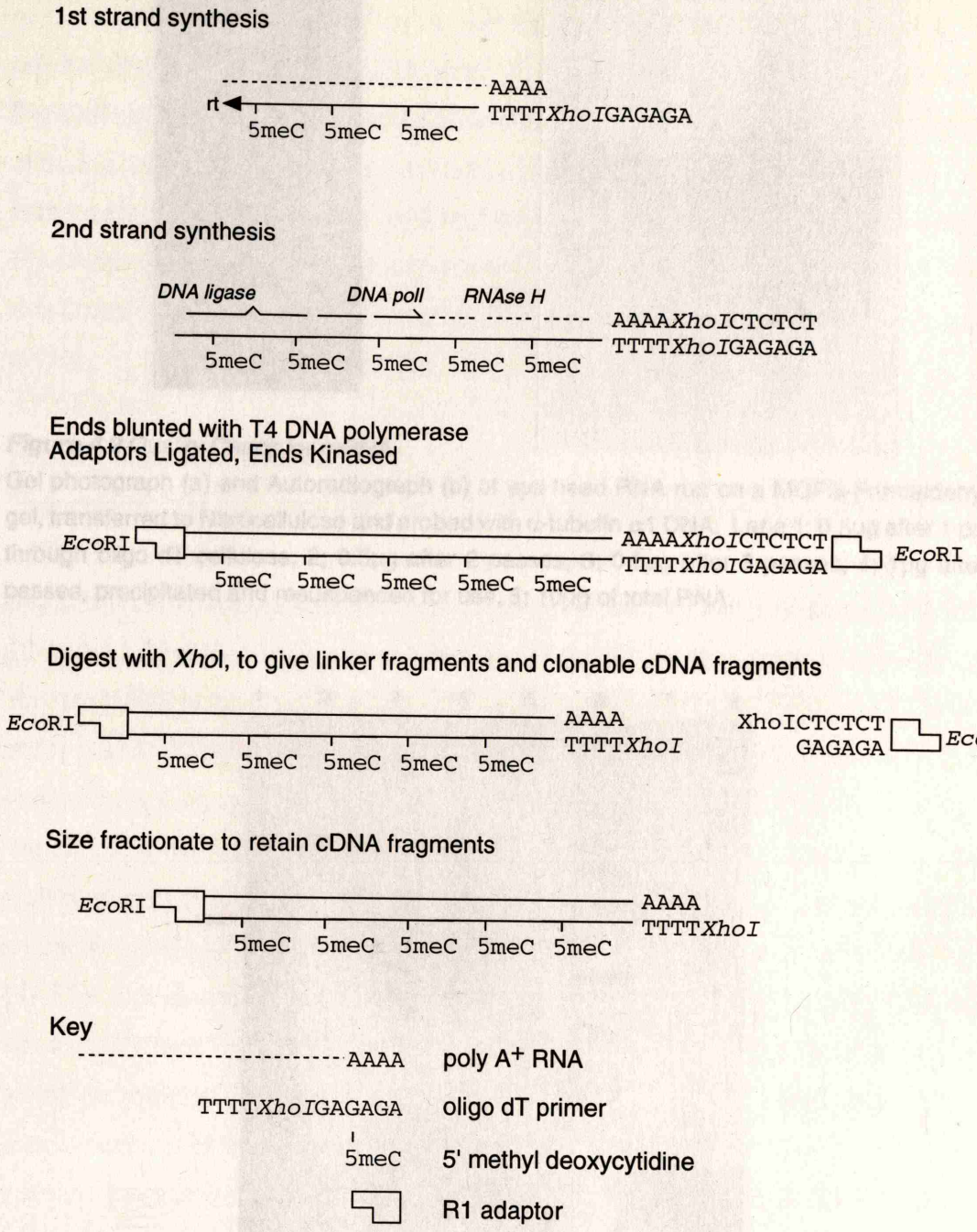


***Figure 4.1** Construction Protocol*

Schematic representation of the construction of the cDNA library. This will be discussed in detail in Section 4.2.2. rt; reverse transcriptase

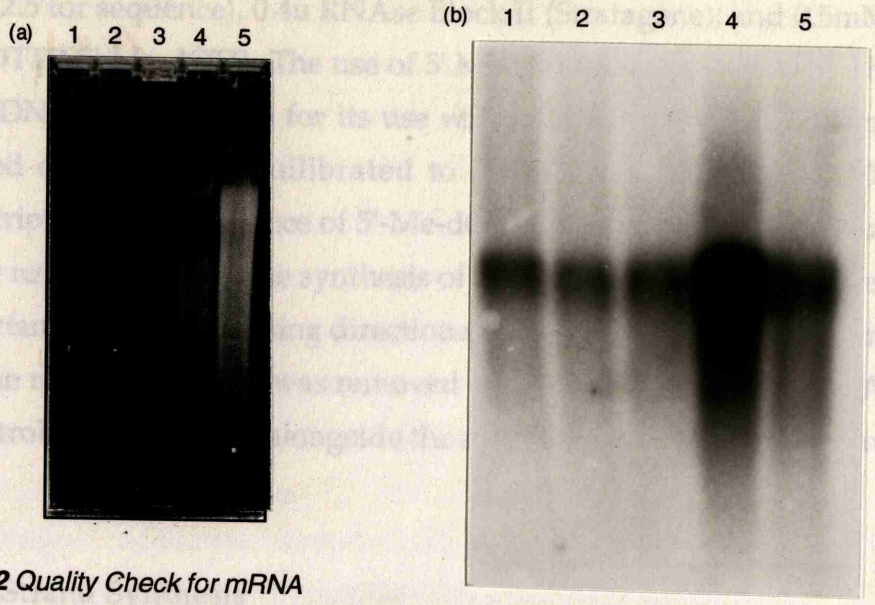


Figure 4.2 *Quality Check for mRNA*
Gel photograph (a) and Autoradiograph (b) of *eya* head RNA run on a MOPS-Formaldehyde gel, transferred to Nitrocellulose and probed with α-tubulin α1 DNA. Lane 1; 0.5µg after 1 pass through oligo dT cellulose, 2; 0.5µg after 2 passes, 3; 0.5µg after 3 passes, 4; 1µg after 3 passes, precipitated and resuspended for use, 5; 10µg of total RNA.

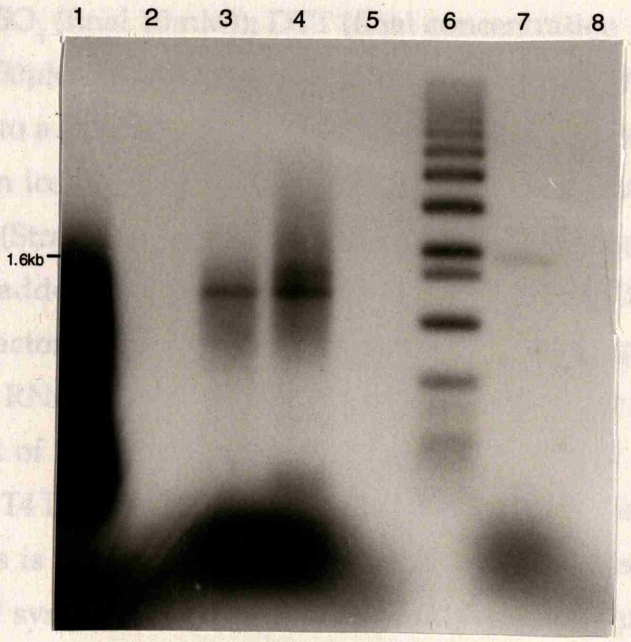


Figure 4.3 *First and Second Strand cDNA Synthesis*
Control samples taken during first and second strand synthesis were run on an alkaline agarose gel (1.5%) and the gel dried down and autoradiographed. Lane 1, Typical first strand cDNA size profile (using wild type polyA$^+$ RNA and Avian Myeloblastoma Virus RT(Pharmacia); 2, 2nd strand *eya* head cDNA made with Mu-MLVRT; 3, 1st strand *eya* head cDNA made with Mu-MLVRT; 4, 1st strand *eya* head cDNA made with SuperScript RT; 5, 2nd strand *eya* head cDNA made with SuperScript RT; 6, 1kb ladder, labelled with *E.coli* DNA polymerase I; 7, 1st strand control cDNA made with Mu-MLVRT; 8, 2nd strand control cDNA made with Mu-MLVRT.

Section 2.5 for sequence), 0.4u RNAse Block II (Stratagene), and 0.5mM dATP, dGTP, DTTP, 5'-Me-dCTP. The use of 5' Me-dCTP results in hemimethylation of the cDNA. The reasons for its use will be discussed later. This mix was annealed on ice, then equilibrated to 45°C before addition of 700 units SuperScript RT. The presence of 5'-Me-dCTP in the reaction mix ensures that cytidine residues used in the synthesis of this strand will be methylated. This is important later in facilitating directional cloning of the cDNA. A sample (1/10) of the reaction mixture was removed to a tube containing 5μCi dATP and this control was incubated alongside the main reaction for 1 hour, then placed on ice.

Second Strand Synthesis

The second strand synthesis reaction was carried out immediately. The following components were added (on ice) to the reaction mixture; 32μl 5x Second Strand Buffer (94mM Tris-Cl pH6.9 (final 25mM); 453mM KCl (final 100mM); 23mM $MgCl_2$ (final 5mM); 750μM β-NAD (final 150μM) and 50mM$(NH4)_2SO_4$ (final 10mM)); DTT (final concentration 5mM; dATP, dGTP, dTTP (final 200μM); dCTP (final 500μM); 20μCi dCTP. The reaction volume was adjusted to a final volume of 154μl by adding the appropriate volume of water. Still on ice, 2 units of RNAse H (Stratagene), 44 units of *E.coli* DNA polymerase I (Stratagene) and 15 units of *E.coli* DNA Ligase (New England Biolabs) were added. The reaction was then incubated for 2.5 hours at 16°C. β-NAD is a co-factor essential for the action of *E.coli* DNA ligase, a ligase which does not ligate RNA to DNA, only DNA to DNA. This ligase is found to facilitate a higher yield of full length clones (D'Alessio & Gerard, 1988) than when a ligase such as T4 Ligase (which may ligate RNA to DNA) is used. The precise reason for this is however unclear. An excess of dCTP is added during the second strand synthesis to compete out the 5'Me-dCTP used in first strand synthesis, thus ensuring that synthesis of the second strand uses unmethylated dCTP and that the double stranded cDNA molecules produced are hemi-methylated. The low temperature of incubation for this reaction is intended to reduce the occurence of hairpin structures in the cDNA synthesised as such structures are unclonable. After the reaction has incubated for the allotted time, the mixture was phenol-chloroform extracted twice, chloroform extracted twice and precipitated at -20°C overnight. The cDNA was collected by centrifugation

in a microfuge at 4°C for 1 hour. The pellet was washed in 80% ethanol and re-spun. After drying, the pellet was resuspended in 43.5µl H_2O and a portion (1/10) removed and kept as a control. The reaction volume was adjusted to 50µl with the addition of dNTPs (to a final concentration of 125µM), DNA Polymerase buffer (the components of this buffer are not detailed) and 10 units of T4 DNA Polymerase. The reaction was incubated at 37°C for 30 minutes and phenol-chloroform extraced, chloroform extracted and precipitated as before. This step digests back or fills in the uneven ends of the cDNA synthesised, readying them for addition of the *Eco*RI adaptors.

At this stage, the control samples collected during first and second strand synthesis were collected and electrophoresed on a denaturing alkaline agarose gel. An autoradiograph of this gel is shown in Figure 4.3. It can be seen from this gel that first strand synthesis has been successful and that the cDNA made by the SuperScript method seems to be of slightly higher average size than that synthesised by the native (unengineered) enzyme. No second strand products are seen for either of the two head cDNA samples, or the control DNA supplied with the kit. This suggests that the amount of cDNA loaded on the gel was not sufficient; perhaps the α^{32}P-dCTP incorporation at this stage was inefficient. Another gel (not shown) did indicate that second strand synthesis had been successful. The presence of a distinct band in both the head cDNA samples is puzzling. As may be seen in Figure 4.2, no corresponding band was seen in the starting RNA, although the size (~1.4kb) is similar to that of one of the ribosomal RNA species. This band may be peculiar to *eya* head cDNA as it had not previously been seen when first strand cDNA had been made from other *Drosophila* polyA$^+$ RNA (for example lane 1 of Figure 4.3).

Ligation and Kinasing of Adaptors

The next step is to ligate adaptors onto the ends of the cDNA. The precipitated cDNA was pelleted and washed as before and lyophilised for 10 minutes. The pellet was resuspended in a 1x ligase buffer containing 7µg *Eco*RI adaptors (the components of the Ligase Buffer supplied in the kit are not stated). After incubation at 37°C for 10 minutes, the reaction is made 1mM rATP and 1 Weiss unit of T4 DNA Ligase is added. This reaction was incubated overnight at 4°C, before heat inactivation of the Ligase (70°C for 30 minutes). The sequence of

the *Eco*RI adaptors are as shown below.

```
5' AATTCGGCACGAG 3'
   |||||||||
3'     GCCGTGCTC 5'
```

When ligated to the cDNA, these adaptors produce a molecule with an *Eco*RI compatible overhang at each end. Initially, the protruding ends are dephosphorylated to prevent adaptors from ligating to each other. After the adaptors have been ligated to the blunt ended cDNA, these ends are kinased to enable their eventual ligation onto the dephosphorylated ends of the vector. After inactivation of the kinase, the cDNA is kept in the same buffer, the solution is made 2mM rATP and 10 units of T4 Polynucleotide Kinase is added. The reaction volume is adjusted to 20µl. This mixture is incubated at 37°C for 30 minutes, then the Kinase is inactivated by a further 30 minutes incubation at 70°C.

Making the cDNA Directional

At this stage of the process we have double stranded cDNA with *Eco*RI sites at each end. In theory, this could be ligated into *Eco*RI digested λZAPII vector and packaged now. It is desirable however to know a little about a given cDNA clone before one starts to manipulate it. By ensuring that all cDNAs in a library are cloned in the same orientation, the 5' end of each cDNA should always be adjacent to a particular restriction site. This might be useful in (for example) the production of single stranded probes or in preliminary sequencing studies. This cDNA cloning strategy incorporates a number of features which together facilitate directional cloning of the cDNA. The original oligo dT primer used to prime first strand synthesis included an *Xho*I site. The first strand of the cDNA was made with methylated dCTP. For second strand synthesis, this was replaced by unmethylated dCTP. This results in double stranded cDNA which is hemimethylated along almost its whole length, aside from a small region at its 3' end downstream of the polyA tail. It is known (McLelland & Nelson, 1988) that the restriction enzyme *Xho*I will not cut this hemimethylated DNA. The *Xho*I site within the oligo dT primer is unmethylated and will act as a substrate for the restriction enzyme. Consequently, digestion of the cDNA produced here with *Xho*I should yield molecules with an *Eco*RI restriction site at their 5' end and an *Xho*I restriction site at their 3' end. This reaction is carried out in a

final volume of 50µl after the addition of a buffer supplement (composition undisclosed) and 90 units of the *Xho*I enzyme. After digestion at 37°C for 1 hour, the reaction was stopped and the solution adjusted to 1x STE (100mM NaCL; 10mMTris-Cl pH8.0; 1mM EDTA). The cDNA is then ready for size fractionation.

Size Fractionation of cDNA

After the *Xho*I digestion we are left with a population of cDNA molecules with the correct ends for directional cloning. In the same solution however, are an equal number of very short molecules which also possess the correct ends for cloning. These are the other products of the digestion of the *Eco*RI ended cDNA with *Xho*I. To separate these two populations of molecules it is necessary to pass the whole population over a size fractionation column. For this step, CL4B cellulose (Pharmacia) was used, poured to a bed volume of 3ml and equilibrated with 1xSTE. The whole of the cDNA population was added to the column and immediately single drop fractions were collected. The fractions were counted by Cerenkov scintillation and the values obtained plotted to give the results shown Figure 4.4. The larger molecules pass quickly through the column whilst a much larger volume of the column is available to smaller molecules and these take longer to pass through the bed of cellulose beads. This can be seen on the graph as an initial peak (corresponding to the cDNA) followed by a pause and a larger peak. As it was considered vital to avoid the small linker fragments, it was decided to pool fractions 1-32, and discard the remaining fractions. The saved fractions were extracted wiith an equal volume of phenol-chloroform, followed by chloroform extraction and then ethanol precipitated (the STE provides enough salt to precipitate the cDNA). This ethanol precipitate was pelleted and washed as before then resuspended in a total of 10µl of dH_2O ready for ligation into the λZAPII vector.

Ligation of cDNA to Vector Arms

One quarter of the cDNA (~0.15µg) was ligated to 1µg λZAPII arms (a molar ratio of approximately 3:1, insert:arms) in 1x ligase buffer supplemented with 1mM rATP and 2 Weiss units of Ligase. The ligation reaction was carried out at 4°C for a period of 48 hours after which the ligation products were packaged into phage using the Gigapack Gold II packaging extract (Stratagene). This packaging extract is optimised for packaging of cDNA libraries. After completion of the packaging reaction, phage buffer and chloroform were added and the packaged phage stored at 4°C.

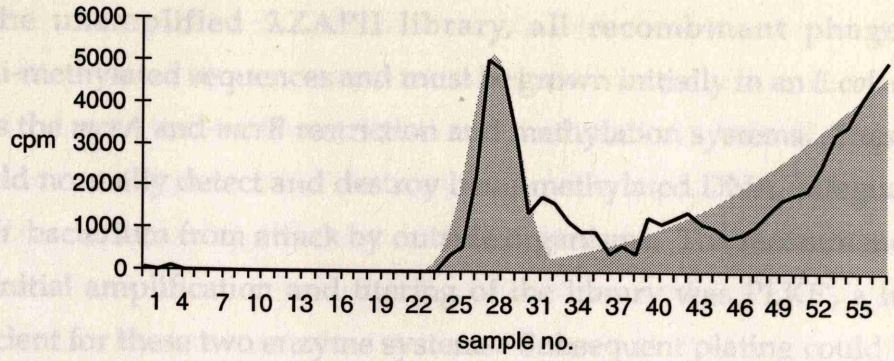


Figure 4.4 *Cerenkov Profile of Fractionated cDNA*
Graph showing activity of fractions eluted from the column. Fractions 1-32 were pooled and retained. Fractions 33 onwards were discarded. All activity had passed through the column by fraction 78 with a peak at fraction 70. In the graph above, only fractions 1-57 are shown. The solid line represents the recorded activity. The graph can be thought of as being composed of two different peaks, the first representing the cDNA fraction, and the other representing adaptors and unincorporated nucleotides. These two peaks are indicated by the shaded area.

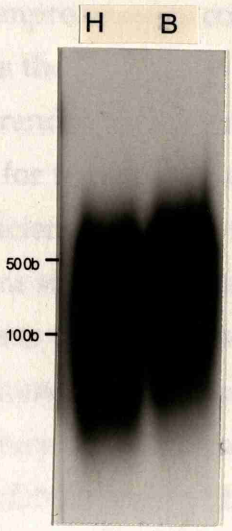


Figure 4.5 *First Strand cDNA Probes*
Head and body cDNA probes were prepared as described in section 2.18.2. After separation from unicoroporated nucleotides, but prior to HAP chromatography, samples were removed for analysis by electrophoresis. Probes were run on a 1% alkaline agarose gel. After electrophoresis, the gel was dried down (by capillary action) and an autoradiograph taken directly from the gel. H, Head; B, Body.

Plating Out Library

In the unamplified λZAPII library, all recombinant phage contain hemi-methylated sequences and must be grown initially in an *E.coli* strain that lacks the *mcrA* and *mcrB* restriction and methylation systems. These systems would normally detect and destroy hemi-methylated DNA, safeguarding the *E.coli* bacterium from attack by outside organisms. The recommended strain for initial amplification and titering of the library was PLKF', a host strain deficient for these two enzyme systems. Subsequent plating could be carried out on any standard plating strain.

When the SuperScript RT head cDNA library was plated out, the total titre was calculated to be 1.75×10^5 plaque forming units (pfu). This titre is very low for a cDNA library. The cDNA library made with the native Mu-MLVRT enzyme had a total titre of 6×10^5pfu. The disparity in the titres obtained may reflect the conditions used in construction of the two libraries. The cDNA library made with the native enzyme was constructed under optimum conditions whereas the library made with the SuperScript RT enzyme was constructed under conditions which had been compromised to conform to the special needs of the engineered enzyme as well as the conditions required for the construction of the library. During second strand synthesis in particular, the conditions used may well have been optimal for the production of long cDNA molecules but might not have favoured efficient synthesis in general, resulting in a lower overall yield of cDNA after this stage. Alternatively, the β-NAD, or one of the other non-standard components used in the second strand synthesis reaction may have carried been through and interfered with one of the subsequent stages in the construction of the library. Furthermore, it should be noted also that after size fractionation of the cDNA through the cellulose beads, only the very first fractions through the column were saved and pooled for the ligation reaction. Thus, a smaller portion of the cDNA synthesised was used in the construction of the library and a smaller titre would have been expected anyway. Together all these factors might account for the small titre obtained. A library of 175,000 clones was however more than sufficient for our purposes as this library was not to be screened *en masse*.

4.3 Screening Strategies

For this project, it was initially decided that a relatively small number of clones from the primary library were to be screened in arrays. Using an array-making device developed and constructed by Mackenzie *et al.* (1989), square plates containing arrays of ~900 bacteriophage clones may be quickly made and duplicated, facilitating efficient differential screening. As this screen was not intended to find only very rare clones (if it were, a subtraction based strategy would have been chosen) there was no real need to screen very large numbers of clones. The primary aim of this screen was to find differentially expressed genes; screening a relatively small number of these clones at low density should identify more than enough clones to work on. Within this population of clones, there may well be some which represent rare transcripts.

Of the 175,000 clones in the primary library, ~10,000 were kept for plating out at low density whilst the remaining 165,000 were amplified and plated at high density. Amplified phage were collected from plates (by over-laying with phage buffer). These phage were purified by ultracentrifugation through Caesium Chloride and stored. A portion of this library (10,000pfu) was screened at high density with Actin 5C (Fyrberg et al., 1981). Approximately 60 positives (0.6% of clones) were identified. The abundance of this gene is thought to be around 1%.

After preliminary checks had been carried out on this library, the remaining 10,000 unamplified clones were plated out at low density (200 per 9cm diameter plate) and the plaques picked with glass capillaries and placed into arrays in the device. A small number of plaques (representing non-recombinant phage and the clone λST41, which is expressed only in heads) were also picked and placed into the array for future reference as controls. The array device was then used to inoculate a freshly poured lawn of XLI-Blue plating cells with each set of 900 clones. The plates were incubated at 37°C for 10 hours until the plaques were large but not touching. Three filters were lifted from each plate (as described in Section 2.12) and stored until required.

Differential Plaque Hybridisation

Filter sets were screened one at a time using first strand cDNA probes synthesised from head and body RNA as described in Section 2.18.2. From this

set, 13 potential positives were chosen for further analysis. These clones were re-plated and screened again with the aid of Simon Tomlinson. Unfortunately, only two clones (λC12 and λC13) could still be classified as differential after this screen. The differential screen seemed not to have been very efficient. Over 900 clones had been screened and only 2 clones isolated. A significantly larger number of clones would have been expected at this stage. A second screen, gave even less heartening results. Due to high levels of background, no positives could be identified. Alternative differential plaque hybridisation strategies were considered and attempted. It was thought that the use of oligo dT to prime the synthesis of the cDNA probes might result in non-specific hybridisation of these T-tracts to the Poly-A tails present in all recombinant clones. As a result, random primed cDNA probes were tried. In addition, the hybridisation solution was supplemented with a 50 base poly A tract (at 20μg/ml), a strategy employed originally by Scherer *et al.*, (1981). Neither of these strategies made a significant difference to the quality of results obtained. Finally, a normalisation strategy as described by Fargnoli *et al.*, (1990) was used. This low-ratio subtraction hybridisation approach works on the same principal as normal subtraction hybridisation, but seeks only to remove abundant sequences common to both tissues, rather than all common sequences. The head cDNA probes were pre-hybridised for 24 hours in the presence of 5μg body RNA (~5x excess of driver to probe). Unfortunately, the normalisation procedure seemed to be too efficient, and no strong hybridisation was seen on the head filter probed with the normalised probes. In a previous differential screen carried out in this laboratory, Steven Russell (Russell, 1989) was unable to select male specific cDNA clones using a similar approach. As an alternative, he utilised a genomic library for his screen. In all the differential plaque hybridisation experiments carried out only the initial 2 clones were obtained. Consequently, other strategies were considered. The strategy chosen was Reverse Northern screening.

Reverse Northern Screening

At this stage in the experiment, it was decided that a more efficient way of screening these clones would be to use a Reverse Northern screening strategy like that described in Fryxell & Meyerowitz, (1987). For this, plasmids are cut, loaded on duplicate gels and electrophoresed. The DNA is then transferred to nitro-cellulose filters and screened with radioactive cDNA probes synthesised

from mRNA derived from the two tissues under study. Clones which represent differentially expressed genes can be easily selected on comparison of the pattern of hybridisation seen. To choose the clones to be screened in this way, it was decided to utilise the data acquired in the preliminary screens. Using the existing arrays of bacteriophage clones, and the results of the earlier differential screening experiments, clones which appeared to show no expression in the body were selected. Selection on this feature acts as a pre-screen, eliminating clones expressed at high levels in the body. Furthermore, these clones may be digested with appropriate restriction enzymes and examined to enable disposal of those clones which do not contain inserts. The selected bacteriophage clone inserts were excised as described in Section 2.10, the plasmids recovered and DNA made. After 100 clones with inserts had been collected, the plasmid DNA was digested (with *Eco*RI and *Xho*I to excise the insert) in large quantities. Samples were run on gels under carefully controlled conditions to ensure that the same DNA samples run on different gels would have migrated the same distance. This makes it possible to overlay autoradiographs of duplicate gels after hybridisation has been carried out. In addition to the clones from the cDNA library, each gel contained the following controls: pBSSK$^-$ DNA linearised with *Eco*RI, α-tubulin α1 excised from its vector, and λST41, a bacteriophage lambda clone provided by Simon Tomlinson excised from its vector. The inclusion on the gels of pBSSK$^-$ should indicate the level hybridisation in each track that is due to the presence of the vector (electrophoresis should in most cases separate the insert from the vector anyway). α-tubulin α1 is expressed equally in heads and bodies and should provide a gauge of expression levels seen on the head and body filters. λST41 represents the major opsin in *Drosophila*, an eye specific gene. This clone should hybridise only to head cDNA.

Once the filters were prepared, cDNA probes were made from head and body *Oregon R* mRNA as described earlier. A sample of the cDNA was run on a denaturing agarose gel to check its length. Using the primer extension synthesis method as described, first strand cDNA probes of average length 1.5kb were routinely synthesised. The autoradiograph in Figure 4.5 shows the probes prepared for this experiment. After the probes were made, an aliquot was assayed by Cerenkov scintillation. For the two probes used in this experiment, the specific activities obtained were 1.9×10^9cpm/μg (head) and 1.6×10^9cpm/μg

(body). To remove any double stranded probe (formed when hairpins occur during DNA synthesis and the DNA strand acts as its own template, the cDNA probes were purified by HAP chromatography as described in Section 2.18.2. For these probes, ~95% of both the head and body cDNA was recovered in the single stranded fraction, as measured by Cerenkov scintillation.

The probes (eluted from the HAP column in 120mM $NaPO_4$ buffer) were added to filters which had been prehybridised for four hours in a prehybridisation solution containing polyadenylic acid, as described in Section 2.18.2. The filters were hybridised and washed as described , then autoradiographed for 36 hours. The resulting autoradiographs with their corresponding agarose gels are shown in Figure 4.6. The largest group of clones (with 44 representatives) are those which are not significantly differentially expressed. 22 clones seem elevated in or specific to the head whilst a similar number (27) cannot be classified at this stage. Only a small number (7) seem to be elevated in the body. For those 27 clones which were not yet classified, a second screen was carried out, similar to the first screen. As can be seen from Figure 4.6, an inability to classify these clones was invariably due to their proximity (on the gel) to strongly expressed clones. By running gels with only the unclassified clones on them, and using cDNA probes of similar size and activity the hybridisation results shown in Figure 4.7 (a,b) were obtained. On the basis of this re-screen, a further eight clones could be classified as head elevated or head specific. Three clones remained unclassified, whilst the rest were classified as either body elevated (7) or non differential (9). These cDNA probes were also used to screen filters carrying DNA made from the 13 clones selected by differential plaque hybridisation. The results of this hybridisation are shown in Figure 4.7 (c, d) and agree with the secondary screen data quoted earlier. The initial classification of the pDnn series of clones is presented in Figure 4.8. At this stage of the experiment, 32 clones seem to be elevated in or specific to the *Drosophila* head. Preliminary classification of these clones is described in Chapter 5.

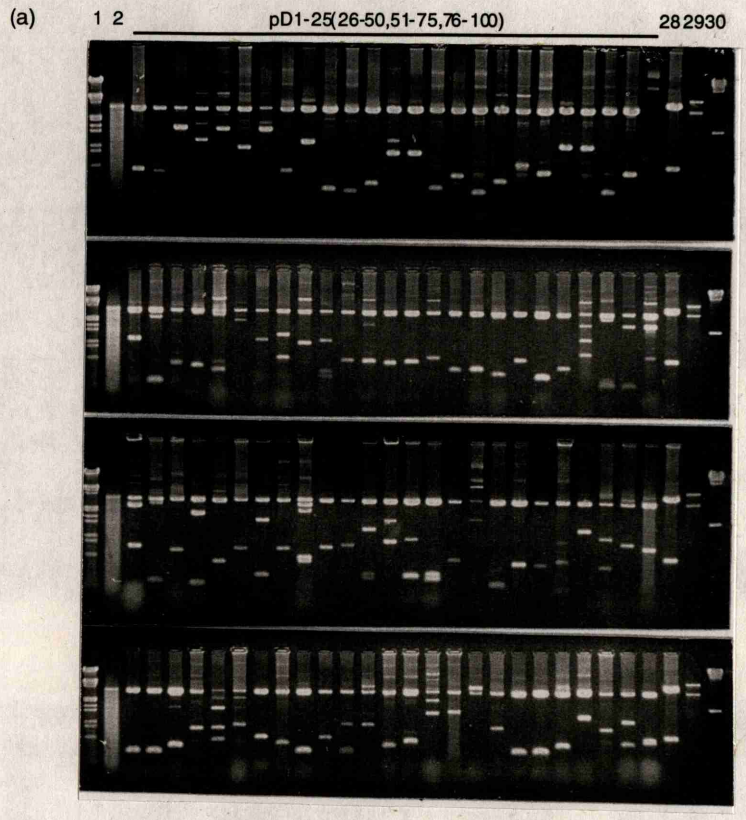
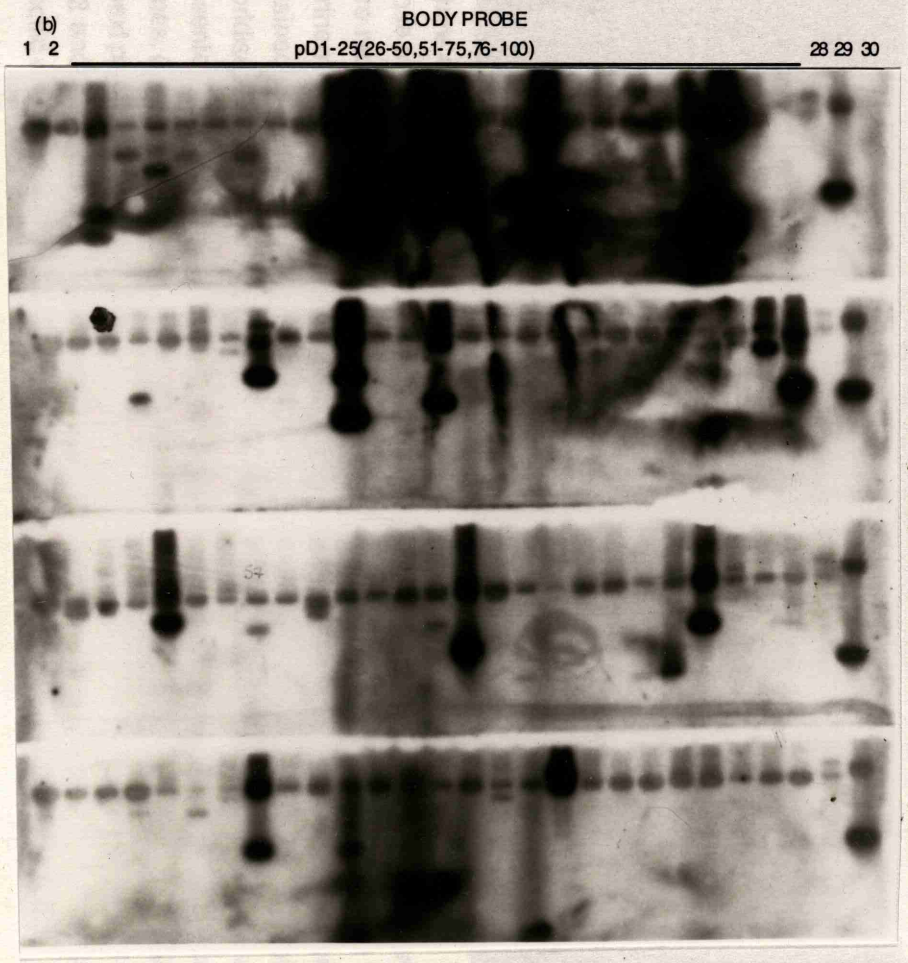


Figure 4.6 *Reverse Northern Experiments*

100 clones were screened by hybridising duplicate filters with cDNA probes made with head and body mRNA. Each filter contained 25 clones and control DNAs: Lane 1, λ digested with *Hind*III/*Eco*RI; 2, pBS digested with *Eco*RI; 3-27, DNAs (pD1-25, pD26-50, pD51-75, pD76-100) digested with *Xho*I/*Eco*RI; 28, pmsfK10 (a male specific clone kindly provided by Diane Harbison, University of Glasgow, as a general negative control for the head probe); 29, RP49 digested with *Eco*RI/*Hind*III; 30, λST41 digested with *Eco*RI/*Hind*III. (a-this page) photograph of EtBr stained gel. (b-next page) filter probed with body probe, (c-next page) filter probed with head probe. Filters were hybridised and washed to high stringency as described in Section 2.18.4. The filters were autoradiographed and equalised exposures obtained (for some reason, the head probe always yielded a stronger signal, both with the non-differential control and with the vector sequences present in each track). A number of the clones could not be characterised on the basis of these experiments and were rescreened as described in the text and shown in Figure 4.7. The initial classification of these clones is shown in Figure 4.8.



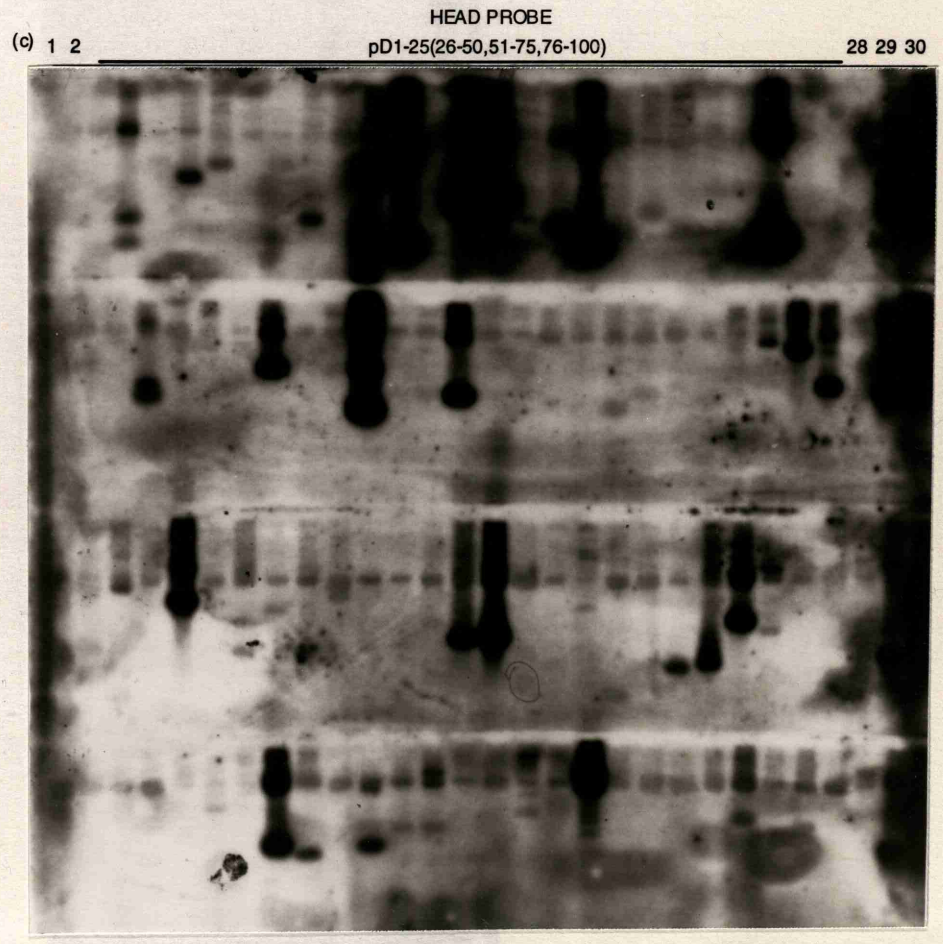
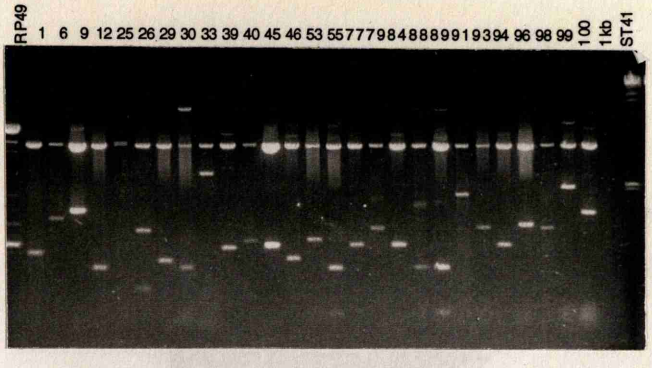
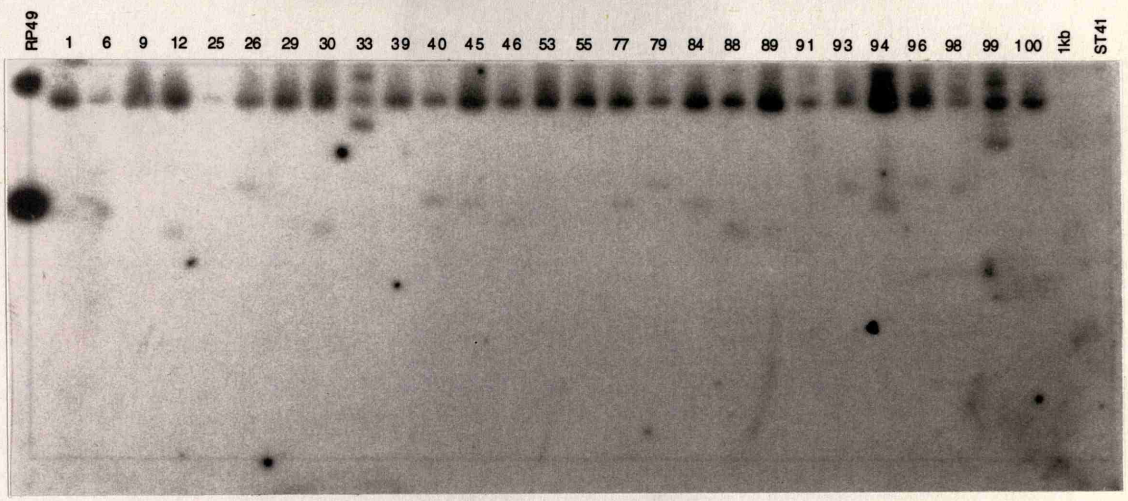


***Figure 4.6** continued* (b) body probe, (c) head probe.

(a)



(b)



(c)

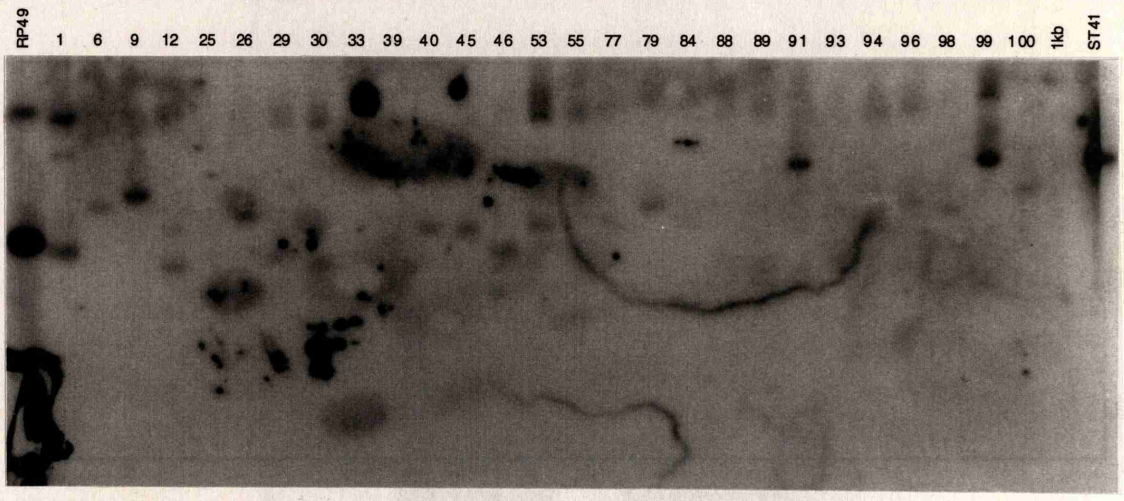
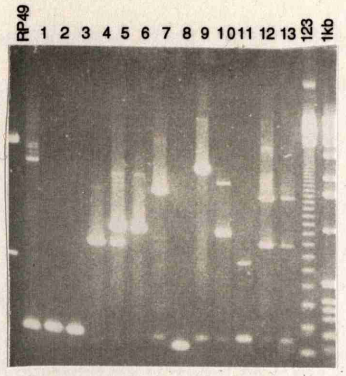


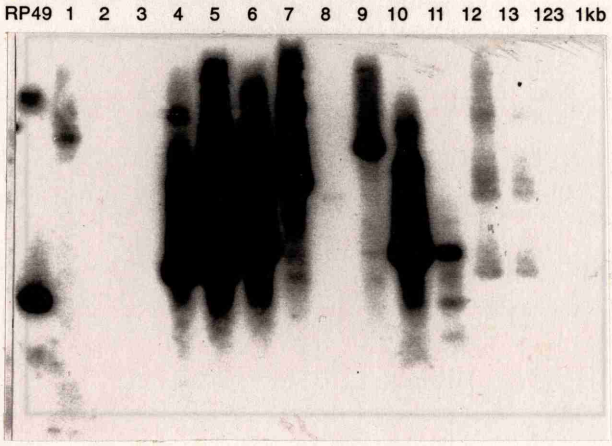
Figure 4.7 *Re-Screening of Clones*

Those clones which could not be classified on the basis of the initial Reverse Northern screen were redigested and run on gels with appropriate controls as described before. (a) Reverse Northern gel with all unclassified clones. (b) autoradiograph showing patterns of hybridisation obtained with a body first strand cDNA probe. (c) autoradiograph showing patterns of hybridisation obtained with a head first strand cDNA probe. Classification of these clones is presented in Figure 4.8. (d-next page) reverse northern gel with original differentially expressed clones. (e-next page) autoradiograph showing patterns of hybridisation with body cDNA probe. (f-next page) autoradiograph showing patterns of hybridisation with head cDNA probe. Clones C12 and C13 were chosen for further analysis. RP49, non differential control, ST41, head specific control.

(d)



(e)



(f)

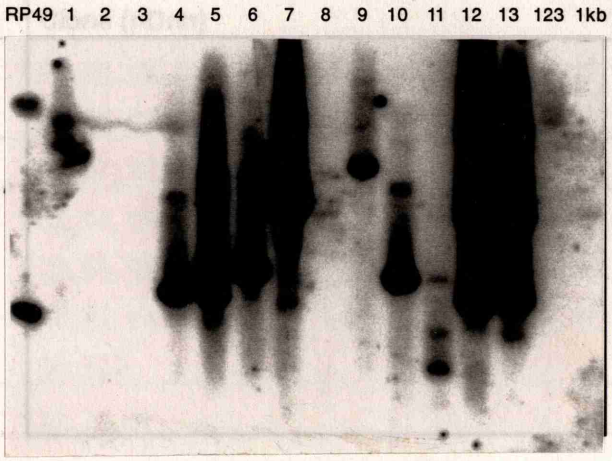


Figure 4.7 *Re-Screening of Clones (continued)*

(d) gel photograph with original differentially expressed clones. (e) autoradiograph showing patterns of hybridisation with body cDNA probe. (f) autoradiograph showing patterns of hybridisation with head cDNA probe. Clones C12 and C13 were chosen for further analysis.

Designation	Clone (pDnn)	Total
Head Elevated	**01, 06, 09, 15, 16, 20, 26, 33, 36, 41, 43, 44, 52, 53, 56, 58, 61, 62, 65, 67, 68, 69, 73, 83, 86, 87, 90, 91, 97, 99**	(30)
Body Elevated	03, 07, 22, 30, 45, 57, 66, 78, 81, 84, 93, 94, 96, 98	(14)
Not Differential	02, 04, 05, 08, 10, 11, 12, 13, 14, 17, 18, 19, 21, 23, 24, 27, 28, 29, 31, 32, 34, 35, 37, 38, 40, 42, 46, 47, 48, 49, 50, 51, 54, 59, 60, 63, 64, 70, 71, 72, 74, 75, 76, 77, 79, 80, 82, 85, 88, 89, 92, 95, 100	(53)
Not Classified	25, 39, 55,	(3)

Figure 4.8 *Final Analysis of Clones*

Simple classification of the clones screened by reverse northern analysis. In (a), a designation is given for each of the original 100 clones. Underlined clones are those which were initially unclassified The clones shown in bold are those chosen for further analysis, along with clones pC12 and pC13.

4.4 Discussion

Using a variety of low density differential screening procedures, this study has selected a number of clones representing genes which seem to be elevated or specific to the head. Although the initial differential screening of plaque arrays met with little success, the preliminary results were of some use in focusing the Reverse Northern screen towards those genes which seemed to be expressed at low levels within the *Drosophila* body. For this type of differential screen where the number of clones to be screened need not be large, a Reverse Northern approach seems to be most useful because the data obtained are more informative than that acquired from plaque hybridisation. In addition, the ability to screen out those clones which contained no inserts made the Reverse Northern screen more productive than the plaque hybridisation experiments. In theory, the increased amount of DNA bound on the filter should increase the sensitivity of the screen. The failure of the majority of the clones selected by the initial plaque hybridisation to prove positive in the secondary screen highlighted a significant potential drawback of that strategy.

Chapter Five

Initial Characterisation of Positive Clones

5.1 Introduction

The 32 cDNA clones isolated after the initial differential screening experiments represent a resource which may be utilised in our subsequent studies of the *Drosophila* brain. In the short term, the full characterisation of these clones was not feasible, therefore, preliminary characterisation of these clones was initiated in an attempt to provide information which would allow selection of a small number of these clones for further study.

A number of factors might influence the final choice of which cDNA clones to work with. The degree of tissue specificity is an important factor (although specificity is by no means an essential pre-requisite). Head (and not eye) specific cDNA clones are likely to represent an interesting class of genes which might perform specific roles in well defined groups of cells. Those cDNA clones which are head elevated are more likely to be representatives of a class of genes with more general neural functions. The abundance of the transcripts represented by these cDNA clones is likewise a plausible criteria on which to select clones. Although rarer transcripts are more likely to perform specific roles, an abundant neural gene which had not previously been characterised would be of great interest to a study like this one. A developmental expression profile would indicate which genes might have roles specific to the adult nervous system and those whose products function throughout development. Any transcripts which show a sexually dimorphic pattern of expression would be particularly interesting as this might lead our investigation towards the sites of control of sexual behaviour in the adult fruit fly. The possession of an interesting sequence feature might make a clone more attractive for study because it may be possible to predict a function for the corresponding transcript and direct research towards gaining evidence to support or disprove that prediction.

On a more mundane level, having reason to believe that a cDNA clone is an artefact is a valid reason for abandoning its study, even if its apparent expression profile is novel. Finally it would be wasteful of resources to study two related clones independently; studying related clones in tandem (or choosing to work with one representative) leads to significantly quicker progress.

To provide this preliminary data, a number of simple experiments were undertaken. A second Reverse Northern was carried out; this time to examine expression of all the transcripts at once. Reverse Northern analysis was also carried out with cDNA probes made from other developmental stages and male and female tissue. Restriction mapping and cross hybridisation studies were carried out in an attempt to identify related clones and to provide information for future experiments (such as subcloning and *in situ* hybridisation). Finally, preliminary sequence analysis of the 5′ and 3′ ends of each clone was carried out to address a number of questions as follows:

- Did any of these clones result from cloning artefacts?
- Might any of these clones represent other artefacts arising from contamination of the source material with DNA or RNA?
- Have any of these genes been investigated before (and their sequence logged in the sequence database) or do they represent clones related to those isolated and characterised in a similar project carried out by Simon Tomlinson in this laboratory? The terminal sequences of the cDNA clones examined in that study were available for sequence comparison.
- Are there any sequence features present which might further our investigation into the role of these genes?

Some of these analyses were carried out in parallel. For clarity however, they have been presented sequentially.

5.2 Reverse Northern

Until this point, all the clones chosen had been situated on several different nitrocellulose filters and all the Reverse Northern screening carried out *en masse*. To gain a more accurate impression of the relative abundance of the transcripts represented by these clones, duplicate filters were prepared containing all selected cDNA clones and appropriate controls. These filters were screened as before (gels not shown). From this Reverse Northern, it was possible to quantify the relative abundance of the transcripts represented by these clones in head and body tissue. In Figure 5.1, a more accurate designation is presented on the basis of the relative hybridisation signal obtained for each of these clones.

clone	design	ratio	emb	mp	size	5'X	Internal sites and Map Info
C12	he	10:1	nd	nd	700	no	none
C13	he	10:1	nd	nd	1400	no	*RI* ~300bp from 5' end
D1	whs	>10:1	-	+	640	no	none
D6	whs	>10:1	-	+	1050	no	*HIII* 200bp from 3' end
D9	whs	>10:1	-	+	1150	no	none
D15	he	12:1	-	+	380	yes	none, no *R1* site in Polylinker
D16	he	6:1	-	++	550	yes	none
D20	he	6:1	-	++	560	no	none
D26	whe	6:1	+	+	1100	yes	*HIII* near 5' end
D33	whs	>10:1	+	+	1950	yes	*XhoI* ~650bp from 3' or 5' end
D36	he	6:1	-	++	620	yes	none
D41	he	3:1	-	++	520	no	*HIII* near 3' end
D43	whs	>10:1	-	+	500	yes	*KpnI* 150bp from5' end
D44	he	6:1	-	++	540	yes	*HIII* ~250bp from 3' end
D52	he	10:1	-	++	250	no	none
D53	whs	>8:1	-	+	800	no	none
D56	he	16:1	-	++	740	yes	none
D58	he	16:1	-	++	740	yes	none
D61	he	6:1	-	++	860	yes	none
D62	he	10:1	-	++	1230	no	none
D65	he	6:1	-	++	640	no	*EcoRI*~300bp from 3' end
D67	he	4:1	-	++	1645	yes	*2xPstI* unplaced (*EcoRI*(300/750)*PstI*(750/300)*PstI*(550)*XhoI*
D68	he	10:1	-	++	250	no	none
D69	whe	4:1	-	+	550	no	none
D73	he	12:1	+	++	1600	yes	*RI&HIII* -same place *EcoRI*(500)*EcoRI&HIII*(1120)*XhoI*
D83	he	9:1	-	+	550	yes	none
D86	he	6:1	-	++	640	no	*EcoRI~300 b from 3' end*
D87	he	16:1	-	++	1000	yes	none
D90	he	12:1	-	++	1300	yes	*BamHI*~350bp from3'end
D91	whs	>10:1	+	+	1600	no	none
D97	he	20:1	-	+	1230	no	*KpnI*~350bp from3'end
D99	he	12:1	+	++	1600	yes	*RI&HIII* -same place *EcoRI*(500)*EcoRI&HIII*(1120)*XhoI*

Figure 5.1 *Simple Restriction Maps and Features of the 32 Clones.*
Features of clones chosen for further characterisation. codes:(w)h(s)(e); (weak) head (specific) (elevated). nd, not determined; -, not detected in embryos/mid pupal tissue; +, weak expression in embryos (emb)/mid pupal (mp) tissue; ++, expression in embryos/mid pupal tissue. Sizes are given in base pairs. ratio; relative expression levels in head and body cDNA. design; designation. 5'X, presence of *Xho*I site at 5' end of clone. The restriction sites in the pBS SK- polylinker are arranged as follows;
T3 primer; *Sst*I-*Xba*I-*Bam*HI-*Pst*I-*Eco*RI-cDNA INSERT-Atail-*Xho*I-*Apa*I-*Kpn*I; T7 primer.

During the initial characterisation of the sets of 100 cDNA clones presented in Chapter Four, cDNA probes made from embryonic and mid-pupal mRNA, along with probes derived from male and female tissue were synthesised by Simon Tomlinson in our laboratory. These probes were made available to me and were used to screen the original filters. The table in Figure 5.1 also tabulates the abundance of these transcripts in embryonic and mid-pupal tissue. No evidence for sexually dimorphic expression was seen for any of these genes studied here.

5.3 Restriction Mapping

Restriction analysis of cloned DNA provides a quick and efficient means of obtaining information about a clone. Although of little biological significance, a simple restriction map may provide the key to identifying related clones. The multiple cloning site (mcs) of pBS SK- contains unique sites for 8 Type II restriction enzymes (all with 6 base recognition sites). Digestion of the cDNA clones with each of these enzymes singly, will quickly identify those enzymes which cut within any of the cDNA sequences. The presence or absence of sites for all of these enzymes was determined for each cDNA clone and the results tabulated in Figure 5.1. A by product of the restriction mapping exercise is that approximate insert sizes are obtained. These are also presented in Figure 5.1.

Upon examination of the restriction site data, a number of observations can be made. Most significantly, a number of the cDNAs seem to possess *Xho*I sites at or near to their 5′ ends (a similar restriction pattern is seen on digestion with *Eco*RI/*Xho*I double digests as with *Xho*I alone). The presence of a 5′ *Xho*I site might indicate an aberrant cloning event. Such clones might not be suitable for further characterisation as they may not reflect true *Drosophila* transcripts. Sequence analysis of the 5′ ends of these clones might provide clues as to how these aberrant clones were generated. Aside from the 5′ *Xho*I sites, clone pD15 lacks a 5′ *Eco*RI site, another potential cloning artefact. A few of the clones appear related: clones pD99 & pD73 have the same insert size and share the same complex restriction pattern; similarly, clones pD65 & pD86 share restriction sites and have the same size insert, suggesting that they are related. A number of the other clones seem to be the same size, but without DNA sequence or

cross-hybridisation data it is impossible to say whether they are actually related. These types of analysis will also confirm the relationship of clone pD99 to clone pD73 and of clone pD65 to clone pD86.

5.4 Cross Hybridisation Studies

A more robust method of determining whether any of the cDNA clones are related is to look for cross-hybridisation. DNA from each clone (and appropriate controls) was spotted onto nitrocellulose filters (for experimental details see the legend to Figure 5.2). Dot blots were prepared and hybridised (with random primed probes made using only insert DNA and 5μCi of dCTP) at high stringency. The results of two of these dot blot hybridisations are shown in Figure 5.2.

From this analysis, there seemed to be 5 groups of overlapping cDNA clones: as predicted already, clones pD65 & pD86 cross-hybridise, as do clones pD73 & pD99. Clones pD52 & pD68 and clones pD56 & pD58 also cross hybridise. Earlier, southern analysis had indicated that pC12 and pC13 were related. Referring to the data from the restriction mapping, in all cases except pC12/ pC13, both members of the clone pair are the same size and are likely to represent the same cloning event. Although the original library was a primary one (which would argue against two identical clones being isolated unless they were full length), it must be remembered that these clones have undergone a significant amount of manipulation (picking, re-picking and excision) and the duplications observed almost certainly represent an inadvertent error in one of these processes. pC12 and pC13 are not the same size and their relationship is discussed in more detail later. Apart from the groups discussed, it should be noted that the rest of the filters showed no evidence for any homology between the remaining clones. From these studies we can conclude that the original screening experiments yielded a total of 27 independent cDNA clones.

5.5 Initial Sequence Analysis

By far the most informative analysis carried out at this stage was sequencing of the ends of these cDNAs and their subsequent analysis. The information gained allows us to address a number of questions as highlighted in Section 5.1.

(a)

15	16	20	36	41
43	44	52	56	58
61	62	65	67	68
69	73	83	86	87
90	97		rp49	SK-

(b) (c)

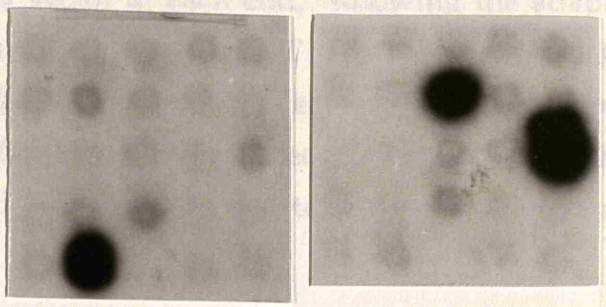


Figure 5.2 *Cross Hybridisation Studies Of The 30 Clones.*
After heating to 95ºC and chilling on ice, 100ng of each DNA (in a total volume of 2µl, 10x SSC, with Bromophenol Blue dye) was spotted onto a nitrocellulose filter in a grid as shown in (a). After application of all DNA samples, filters were air dried briefly, then denatured and neutralised as for plaque lifts. Transferred DNA was UV-crosslinked in a Stratalinker. (b) clone pD97 appears unique. (c) clone pD52 appears to cross hybridise with clone pD68.

```
                                                        XhoI
XHOI PRIMER          GAGAGAGAGAGAGAGAGAGA-ACTAGT-CTCGAG.TX18.....

I   D56/D58T3 CACGAG ........gagagagagaga-ACTAGT-CTCGAG-Tx20-GTTTCGTTCT
      D62T3 CACGAG ........gagagagagaga-ACTAGT-CTCGAG-Tx43-CGTTTAAATA
      D15T3 CACGAG ......gagagagagagaga-ACTAGT-CTCGAG-Tx18-TGTTTGTTGA
      D16T3 CACGAG ......gagagagagagaga-ACTAGT-CTCGAG-TX18-GTTTTTTTTT
      D26T3 CACGAG ......gagagagagagaga-ACTAGT-CTCGAG-Tx25-AACAATATAT
      D61T3 CACGAG ......gagagagagagaga-ACTAGT-CTCGAG-Tx23-GTATTATTTA
      D67T3 CACGAG ....gagagagagagagaga-ACTAGT-CTCGAG-Tx21-AATTTAAAGT
      D83T3 CACGAG ....gagagagagagagaga-ACTAGT-CTCGAG-Tx36-AACTGTTTTT
      D33T3 CACGAG ..gagagagagagagagaga-ACTAGT-CTCGAG-Tx21-AGATGGCGGC
      D44T3 CACGAG ..gagagagagagagagaga-ACTAGT-CTCGAG-Tx29-ATTTTTACAC
      D36T3 CACGAG ........gag.gagagaga-ACTAGT-CTCGAG-Tx25-GTGGAAAACT
    D65/D86T3 CACGAG ..gagaga.agagagagaga-ACTAGT-CTCGAG-Tx05-ATATTGCATT
      D87T3 CACGAG ......gagagagagagag.-ACTAGT-CTCGAG-Tx17-GAGTTGTATT
      D43T3 CACGAG ..gagagaga.agagagaga-actaat-CTCGAG-Tx25-GGCTTTTTTT
      D06T3 CACGAG ....gagagagagagagaga-actag.-.tcgag-Tx29-AAAGCGAAAG
II    D41T3 CACGAG gagagagagagagagagaga-ACTAGT-ctcga.-....-ACCACCAGAA
    D73/D99T3 CACGAG ........gagagagagaga-actag.-CTCGAG-....-AAGCAAAAAC
      D90T3 CACGAG ........gagagagagaga-a.tagt-CTCGAG-....-TATATATGAT
      D69T3 CACGAG ..gagagagagagagagaga-act...-......-....-CGCCAAACGT
      D53T3 CACGAG ..gagagagagagagaaaga-......-......-....-GACAGGAGAC
      D01T3 CACGAG ..gagagagagagagagag.-......-......-....-AAGAAAAGAA
      D20t3 CACGAG ....gagagagaga.aga.a-......-......-....-GAAAAGAAGA
      D09T3 CACGAG ..........gagagaga a-......-......-....-TGAGAGCGAT

III C12/C13T3 CACGAG ....................-......-......-.... ACGATTTTCA
     D52/68T3 CACGAG ....................-......-......-....-GGGGCAGCGG
      D91T3 CACGAG ....................-......-......-....-GAGAGAGGGA
      D97T3 CACGAG ....................-......-......-....-GTAGAGATTG
```

Figure 5.3 *Alignment Of Sequences At The 5' End Of Sequenced Clones.*
T3 sequences from all clones studied were aligned to highlight homology to the *Xho*I primer used in construction of the library (top line). The CACGAG sequence at the left of each line is part of the *Eco*RI adapter used in the cloning process. The last 10 bases of each sequence represent the first 10 bases of each insert. Group I clones possess a 5' polyA tail. Group II clones possess homology to the rest of the *Xho*I primer. Group III clones contain no significant homology to the *Xho*I primer.

Sequencing of the ends of the cDNA clones allows detailed analysis of the junction point of the cDNA and vector at each end. Knowing the adapter sequences, along with the sequence of the oligo-dT primer and the vector sequences, it is possible to predict the sequences at each of the vector-insert junction points. Any deviation from the predicted sequences would indicate that the cDNA clone in question might be an artefact, thereby making subsequent analysis difficult.

After eliminating obvious cloning artefacts, the sequences may be examined for those features which typically signify a *bona fide* cDNA clone. Although there is evidence for some *Drosophila* genes whose transcripts do not possess a polyA tail (e.g. the Histone H3.1 genes; Matsuo & Yamakazi, 1989), the vast majority do, and the cloning strategy used here relies on the presence of a polyA tail in the original mRNA substrate to facilitate cloning. Any cDNA clones characterised which do not possess a 3' polyA tail will not represent real cDNA clones and should be discarded. In addition, around 10-35 bases upstream of virtually all polyA tails in *Drosophila* lies a sequence known as the polyadenylation site which is thought necessary for the correct addition of a polyA tail. Although not invariant, a consensus sequence has been determined for eukaryotic cDNAs (AATAAA; Birnstiel *et al.*, 1985). If a cDNA clone contains a polyadenylation site, it almost certainly represents a real transcript. On the other hand, if no polyadenylation site is seen, this may indicate priming from an A-rich region of the genome rather than the real polyA tail from an mRNA molecule. cDNA clones lacking a consensus polyadenylation site should not however be dismissed.

The main resource generated by the sequence determination is that the sequence itself can be used to search local and remote databases of DNA sequences. This searching allows previously characterised genes to be identified. Sequence analysis was especially important as a means of examining potential overlap between my project and that carried out in this laboratory by Simon Tomlinson, allowing us each to direct our research more effectively. Finally, even if exact matches are not found, database searching can highlight the presence of sequence similarities and sequence motifs in the cDNA clones. Knowing this sort of information about a clone is very useful in providing a basis on which to formulate further biological questions. Such findings allows us to cross from theory to practise.

The 5′ and 3′ end of each cDNA clone were sequenced, generating up to 250bp of sequence (in one strand) for each end of each cDNA. Two cDNA clones (the shortest; pD15 and pD68) were sequenced in their entirety. All sequences were inspected as described above, to look for artefacts, polyadenylation sites, polyA tails and similarity to previously cloned sequences. The results of these investigations are described below.

Investigation of the Sequences at Junction Points.

The 5′ and 3′ end of each cDNA clone was sequenced and the sequences aligned according to the vector sequences present. All sequences were then compared to the expected structure of the ends of the insert as predicted from the sequence of the vector, adapters and primers used in the construction of the library. At the 3′ end of each cDNA, the *Xho*I site should flank a polyA tail of ~18 bases as specified by the *Xho*I primer used in the construction of the library. All clones sequenced do indeed have the 3′ *Xho*I site and a polyA tail (data not shown). All PolyA tails sequenced were of 17 bases or more. In addition, the 3′ ends of all clones sequenced were examined for polyadenylation sites (Birnstiel *et al.*, 1985). Of the 27 clones sequenced, (ignoring duplicates), 15 perfect polyadenylation sites were found, with another 6 clones possessing non consensus polyadenylation sites corresponding to sites recognised from previously cloned genes (Wickens & Stephenson, 1984). Five clones had no accepted polyadenylation site. For one clone (clone pD26) a very long polyA tail prevented reliable reading of sequences upstream of the polyA tail, and so no polyadenylation sequence was searched for.

At the 5′ end of the insert, a more complex picture emerged. Figure 5.3 shows an alignment of the junction sequences for all these cDNA clones. As predicted from the restriction mapping data, many of the cDNAs possess *Xho*I sites at their 5′ end. These sites seem to be derived from undigested *Xho*I primers used in first strand synthesis, and the inserts consist of the normal 5′*Eco*RI adapter followed by the *Xho*I primer, a polyA tail (in most cases) and insert sequences. Of the 27 cDNAs sequenced (ignoring duplicates), 15 possess 5′ polyA tails, whilst a further 8 possess *Xho*I primer sequences at their 5′ ends. Only four cDNAs possessed no homology to the *Xho*I primer at their 5′ end. Possible causes of these artefacts will be discussed at the end of this chapter.

Database Searching

Database searching was carried out with all sequences, whether or not they originated from cDNA clones subsequently classified as artefacts. PolyA tracts and vector sequences were discarded and the remaining DNA sequences (150-225bp) were used to search the GenEMBL sequence database using the FASTA program (Pearson & Lipman 1988) from the GCG software suite. In addition, all sequences were compared (with FASTA) to other sequences generated by myself and Simon Tomlinson in this laboratory.

Perhaps surprisingly, none of the sequences generated provided exact matches to sequences already present within the GenEMBL database. Indeed on the whole, no significant homology was found between any of the cDNAs and sequences already present. The exception was the cDNA clone pD83, which possesses an OPA repeat (Wharton *et al.*, 1987). This sequence feature (Figure 5.4) is a transcribed repeated sequence found within many genes in *Drosophila* (Wharton *et al.*, 1987) and other eukaryotes (Gerber *et al.*, 1994). The OPA repeat is a short tandem repeat of the sequence $(CAX)_n$, where X is G or A. The number of copies (n) of this sequence is usually less than 30. The repeat is found in a variety of genes and codes for a poly-Glutamine repeat. In almost all genes in which it is found, the OPA repeat is known to be translated (Gerber *et al.*, 1994). The clone pD83 possesses a perfect OPA repeat encoding an 11 residue poly-Glutamine tract as shown in Figure 5.4. The OPA repeat is part of an ORF which extends from the start of the known sequence of this clone and ends in an in frame termination codon (TAA). An accepted polyadenylation site (AATTAA; Wickens & Stephenson, 1984) is seen 16 bases upstream of the poly A tail in the 37 base 3′ UTR. Once the OPA repeat is excluded, no homology is seen when this sequence and its translation products are compared to DNA and protein databases. Often, poly-Glutamine stretches are found in transcription factors where they are thought to act as activation domains (Gerber *et al.*, 1994). Examining the SwissProt protein sequence database, Gerber *et al.*, found that 33 of the 40 'best' OPA repeats in that database were from known transcription factors, and these transcription factors seem to have specific rather than ubiquitous roles. By creating GAL4 constructs containing poly-Glutamine tracts, they were able to activate transcription from GAL4 responsive promoters considerably (compared with constructs lacking the OPA repeat sequences).

```
          50         60         70         80         90      + 100
pD83   GGCTGGCGAAGAATGTACGATTTTTTG--AAGGATCAGCTAGGATCGAAAAGAACAACAGCAA
       |||| | ||||  | | || | || | | | || |||| ||||||
YGal11 TCACTACAGCAAATGCAGCATTTACAGCAATTGAAAATGCAGCAGCAACAACAACAGCAGCAA
         2470       2480       2490       2500       2510       2520

         110        120        130        140        150        160
pD83   CAACAACAGCAGCAGCAACAACAACTGGCACA--GGCAGTAGCAG-AAATGTATAACAAAAAC
       |||||||||||||| |||||||| | | |||   ||| | ||| | || |||
YGal11 CAACAACAGCAGCAACAACAACAGCAGCAACAACAACAGCAACAGCACATATATCCCTCCTCG
         2530       2540       2550       2560       2570       2580
```

Figure 5.4 *The OPA Repeat Found Within Clone pD83*

FASTA alignment of clone pD83 with the DNA sequence from a gene encoding a transcription activator for yeast galactose inducible genes (Genbank Accession, M22481), the best match found in the GenBank database. The two sequences show 67% homology in a 106bp overlap. Starting at base 98 of the pD83 seuence (+), the sequence CA(A/G) is repeated 11 times. Within this region, all deviations from the Yeast sequence are in the third base of each triplet. A further 8 repeats of this triplet are found in the Yeast sequence.

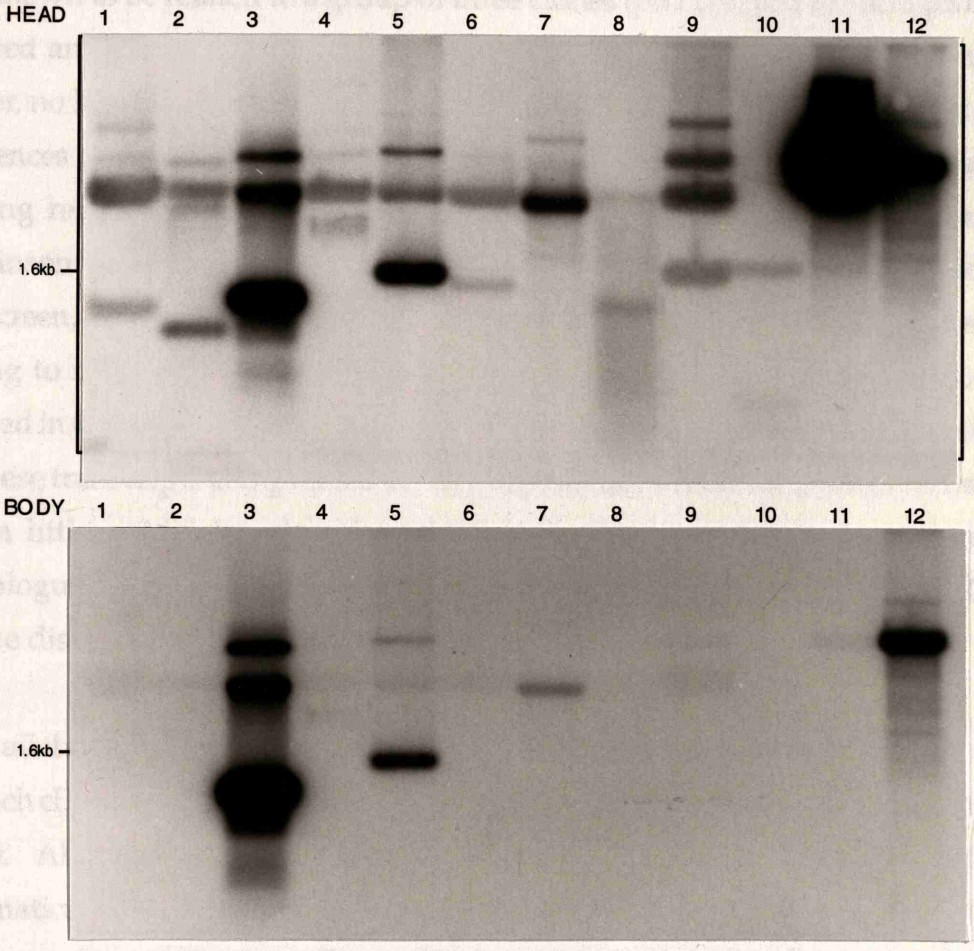


Figure 5.5 *Reverse Northern Analysis of the cDNAs Chosen for Further Study*

Plasmid DNA (0.5µg) was digested with appropriate restriction enzymes to release inserts. Digested DNA samples were split and run on agarose gels before transfer to Nitrocellulose membranes. Filters were hybridised with high specific activity cDNA probes made from head and body mRNA as described previously. After hybridisation, filters were washed and autoradiographed. The exposure shown is 2 hours. Lane 1, pC13; 2, pC133, 3;non-differential control; 4, pD68; 5, pD682; 6, pD91; 7, pD913; 8, pD97;9, pD974; 10, 1kb ladder; 11, λST41 (head specific control); 12, α-tubulin-α1.

The pD83 clone possessed vector sequences at its 5′ end. Using 3′ sequences, it might be possible to screen for other representatives of the transcript represented by this clone. The pD83 cDNA clone may possibly represent a tissue specific transcription factor, and if so, would be of considerable interest.

More success was obtained in sequence analysis of those cDNA clones that had been previously generated and characterised in this laboratory. Aside from confirming the relationship of cDNA clones pD52 & pD68, pD56 & pD58, pD65 & pD86, and pD73 & pD99, this local analysis also indicated that clones pC12 and pC13 were related at their 5′ and 3′ ends, though other findings suggest that their internal structure is different. In addition, these two cDNAs were also shown to be related to a group of three clones (pST170, pST133 and pST59) isolated and characterised by Simon Tomlinson in this laboratory. As stated earlier, no homology was detected between cDNA clones pC12/pC13 and the sequences of previously cloned *Drosophila* genes in the GenEMBL database, a finding reflected the results of database searching carried out by Simon Tomlinson with the pST170 group of clones. 2 of 32 cDNA clones isolated in this screen, along with 3 of 18 cDNA clones isolated in Simon Tomlinson's screen belong to this class of related transcripts, accounting for 1/10 of the clones isolated in these two screens. This observation suggests that the gene responsible for these transcripts is expressed at very high levels within the *Drosophila* head. It is a little surprising therefore that such a highly expressed gene (or a homologue from a different organism) has not so far been characterised. pC13 will be discussed further in a later chapter.

Once all this information had been obtained, a final classification could be made for each cDNA, and decisions made on which cDNA clones to choose for further study. All the clones designated 'Group I' in Figure 5.3, were discarded automatically because their structure suggests that they may represent two unrelated clones. The four 'Group III' clones were all chosen for further study as these were the most likely to represent real transcripts. All four clones had been classified as at least 10 fold enriched in head tissue as judged from Reverse Northern analysis. Clones pD68, pD91 and pD97 possessed consensus polyadenylation sites. In addition to these four clones, clone pD9 was also chosen. This cDNA was a member of the 'Group II' clones described in Figure

5.3. This clone possessed only limited homology to the *Xho*I primer sequence, but had other features which made its further characterisation attractive. cDNA pD9 had been classified as 'weak head specific' in Reverse Northern analysis.

5.6 Screening for cDNAs

As the quality of the *eya* head cDNA library had been called into question, it was decided that the five cDNA clones chosen for further analysis should be used as probes to obtain cDNA clones from a second cDNA library previously in use within the laboratory. This cDNA library, constructed by Steven Russell in the vector λNM1149 used head mRNA from Oregon R flies as its starting material. As the selection stage of this project had been concluded, it was no longer important to use a library made from *eya* RNA. The new cDNAs isolated might be considerably larger than the original cDNAs, allowing a more complete and meaningful characterisation. Initially, 50,000 bacteriophage clones were screened with each of the five cDNAs chosen for further study. pD68 identified 120 positively hybridising plaques of which six were chosen for further analysis. 4 clones were selected (from 75 positives found) for pC13, but no positives were found for cDNAs pD9, pD91 and pD97. A further 150,000 clones were screened with these 3 clones. 4 positives were found for both pD91 and pD97, but no positives were obtained for clone pD9. No further work was carried out on this clone.

All the positively hybridising clones were purified to single plaques and DNA made from the isolated bacteriophage. Crude restriction mapping of the bacteriophage allowed the selection of the longest subclone in each set and subcloning fragments into pBS SK$^-$ plasmid DNA, cut with the appropriate enzyme (*Eco*RI, *Hind*III or a double digest of both enzymes). Cross hybridisation by Southern analysis was used to show that in each case, the subcloned fragment corresponded to the original bacteriophage insert (or part of it) and was related to the original clone. For clone pC13, a subclone pC133 was chosen which contained a 0.9kb *Eco*RI fragment. Similarly, for clone pD68, a 1.5kb *Eco*RI/*Hind*III subclone was chosen. The bacteriophage clone λD913 had a complex pattern of restriction, so a positively hybridising *Eco*RI fragment of 2.8kb was chosen initially. This was contained within the subclone pD913. Finally, for clone pD97, a 1.5kb *Eco*RI fragment

was subcloned (clone pD974). The λNM1149 library was made with *Eco*RI/*Hind*III cut vector and therefore, all these bacteriophage clones should have *Eco*RI/*Hind*III fragments as inserts. For clones pD91 and pD97, at least one further *Eco*RI/*Hind*III fragment representing the extreme 3′ end of each clone should be present. Although no further bands were seen in restriction digests of the λD974 bacteriophage, it is possible that this 3′ fragment may be very small and not visible on an Ethidium Bromide stained gel. It is important to note that there was no internal *Eco*RI site in the original clone pD97 (see Figure 5.1).

It is possible that the newly isolated cDNA clones do not share the same restricted pattern of expression that the originally selected cDNAs showed; perhaps due to differential splicing to produce families of related transcripts with different expression patterns. It was therefore decided that the longer cDNAs should also be subjected to Reverse Northern analysis. Reverse Northerns were carried out as before and the results presented in Figure 5.5. As can be seen from the Reverse Northern hybridisation, the expression pattern obtained does not always match that of the original clone. This is especially true for pD682 whose expression pattern no longer seems significantly head elevated. It is important to recall that the cDNA clone pD68 was very short. Perhaps this fragment represents a differentially expressed transcript whilst the longer clone represents a non-differentially expressed transcript from the same locus. There are many cases of alternative processing of 3′ ends in several *Drosophila* genes (e.g. *exuperantia*, Hazelrigg & Tu, 1994; *doublesex*; Burtis & Baker, 1989). The full implications of the expression data shown here will be discussed in the next three chapters which deal with attempts at further characterisation of the remaining four clones, pC13, pD68, pD91 and pD97.

5.7 Conclusions

The initial characterisation of this selection of 32 cDNA clones has provided valuable information enabling the selection of a subset of cDNA clones to characterise further. These findings do raise a number of questions regarding the fidelity of the original library and its construction. It is encouraging to note that on the whole, repeated isolation of the same transcripts did not occur.

Furthermore the lack of matches to clones representing previously cloned *Drosophila* genes was also encouraging, although this final observation may be misleading and will be discussed.

It is disappointing to find artefacts in any library of cloned sequences, especially at the high levels found here. Examining the construction strategy for this library (as seen in Figure 4.1) it is difficult to pinpoint one step which might be responsible for the artefacts observed. Indeed, it is evident that the generation of artefacts of the type seen are reliant on each stage of the library construction process working to some degree. The presence of *Xho*I primer sequences at the 5′ end of all artefact clones indicates an inappropriate ligation event. From the sequences obtained, it is evident that in all the artefacts, the *Xho*I primer has been ligated to the *Eco*RI adapter. This would of course occur as a normal stage in the cloning process, but all such molecules should be cleaved by subsequent digestion with *Xho*I. The presence of correct junction sites at the 3′ end of all clones indicates that *Xho*I digestion was successful in at least some cases. Those fragments not cleaved by *Xho*I do not form a suitable template for ligation into *Xho*I/*Eco*RI cut vector. Molecules which had not been cleaved at their primer *Xho*I site (due to problems with *Xho*I digestion) would still be unclonable as they would possess *Eco*RI sites at both ends. Head to head cloning of two cDNA clones, each correct at their 5′ end, but with one possessing an *Eco*RI site downstream of its poly A tail is possible. The relative lack of internal *Eco*RI sites (present in 2 of 23 artefact clones) argues against this possibility. In any case, such events are likely to occur with far less frequency than correct cloning events. Other combinations of events can also explain the type of artefact recovered, but they all have one feature in common. They would be expected to occur at much lower frequency than the ligation events which are necessary to bring about a *bona fide* clone. It is as yet unclear whether the artefacts seen here represent two clones joined head to head, or one clone, along with extra primer sequences at its 5′ end. Although the second possibility would allow interpretation of expression data for these clones, the uncertainty led me to conclude that it would be prudent to abandon characterisation of these suspect clones and concentrate instead on those that seemed real. It is unfortunate that more positive criteria (such as expression profile or abundance) could not be

used to select clones to work on. The four clones chosen do represent a variety of expression profiles and abundances. As a result, their characterisation may tell us something of many aspects of neural function in *Drosophila*.

A perceived disadvantage of differential screens is their relative insensitivity. A great deal of thought went into the design of this screen in an effort to increase the final sensitivity of the screen without increasing its technical complexity. Some of the findings of this work would vindicate this approach. None of the sequences generated matched sequences already present in the database. If a wild type library had been used, a significant proportion of the clones selected as head specific would represent genes such as those encoding opsins or other abundant eye-specific proteins. Indeed a single clone selected from a wild type library screen carried out by Simon Tomlinson was sequenced and found to encode a *Drosophila* opsin, previously cloned by O'Tousa *et al.*, (1985). Similarly, with the exception of the pC13/pST170 series of cDNA clones, there was no evidence of very highly abundant cDNAs hindering the isolation of clones representing rarer transcripts. Indeed, pC13 was selected precisely because it was an abundant, unknown transcript and therefore of considerable interest.

The lack of homology or identity between the sequences generated so far in this study may have been due to the methods used to obtain preliminary sequence data. For all 3' sequences, the first ~200 bases are likely to fall outside coding regions. The 3' untranslated regions of most genes have a significantly higher proportion of A/T residues than the rest of the genome. When searches are performed with A/T rich sequences, random matches to other 3' untranslated regions and A/T rich genomes such as mitochondrial DNA are found. Very few of the sequences obtained were not A/T rich so it is perhaps unsurprising that no matches were found. Database searching with full length sequences from the cDNAs chosen for further study will alleviate this problem. Furthermore, the longer sequences generated will be examined for Open Reading Frames and the potential translation products used to search protein databases. This approach is a more sensitive means of identifying homology between related sequences.

Chapter Six

Characterisation of Clone pD682, a Mitochondrial Phosphate Carrier Protein

6.1 Introduction

The first *Drosophila* cDNA clone to be studied in more detail was pD682. This cDNA was represented by a 1.5kb subcloned *Eco*RI/*Hind*III fragment. Although Reverse Northern analysis with the original clone (pD68) had indicated a differential pattern of expression, the same analysis of the subclone pD682 had indicated a ubiquitous expression pattern. The initial clone pD68 was very short (250bp); this might make the original designation less reliable.

Despite its apparent ubiquitous expression, it was decided that the characterisation of clone pD682 should continue as there was reason to believe that it represented a previously uncloned *Drosophila* gene.

The further analysis of the subclone pD682 involves a number of experiments. Full length sequencing in both strands should facilitate detailed database searching. It was hoped that this might enable a function to be assigned to the gene encoding this cDNA. If homologous genes have been found in other organisms, then knowing the size of these genes, it may be ensured that full length clones are isolated and characterised. Southern analysis carried out on genomic DNA with this cDNA subclone as the probe should provide some information on the genomic organisation of the clone. For instance, whether it is present as a single copy or repeated throughout the genome. By studying hybridisation to polytene chromosomes, the gene encoding this cDNA clone can be localised to a particular chromosome band. Consultation of Online databases (such as FlyBase, maintained at the University of Indiana), and directories (Lindsley & Zimm, 1992) might allow identification of candidate genes. Although expression studies may be carried out to further investigate the distribution of this transcript, its ubiquitous expression profile suggests that these results will be largely uninformative.

6.2 Results

6.2.1 Sequencing of the Clone pD682

Initially it was decided to sequence the clone pD682 in both strands in an attempt to identify its function by examining sequence characteristics. The first strand was sequenced by creating nested deletions (Henikoff, 1984). Once this first

strand was sequenced fully, the sequence generated was used to design seven oligonucleotide primers to allow determination of the sequence of the second strand. This sequencing strategy is summarised in Figure 6.1. The sequences of the oligonucleotides used in this sequencing project are tabulated in Section 2.5. The 1474 base sequence generated by these methods was compiled and assembled using the IBI MacVector and AssemblyLign sequence analysis software. This DNA sequence is shown in Figure 6.2. A long Open Reading Frame (ORF), stretching from base 5 to base 907 was found using TESTCODE and CODONPREFERENCE analysis. Although lacking a translation start codon (ATG), this long ORF does end with a recognised stop codon (TAA) whilst the 3' untranslated region contains a consensus polyadenylation site (AATAAA) 26 bases upstream of the 16 base polyA tail. The likely translation product from this subclone is shown in Figure 6.3.

The GCG program FASTA (Pearson & Lipman, 1988) was used to compare the likely translation product to previously known protein sequences. It became immediately apparent that, although not cloned previously in *Drosophila*, the putative polypeptide encoded by the clone pD682 is very closely related to a number of previously characterised mammalian proteins. The three proteins which show greatest homology are from Bovine, Rat, and Human sources (Runswick *et al.*, 1989; Ferreira *et al.*, 1989; Dolce *et al.*, 1991). Each encodes a protein called the mitochondrial Phosphate Carrier Protein (PCP), an ubiquitous protein found on the inner mitochondrial membrane in all cells. The function of the PCP is to mediate the transfer of phosphate over the mitochondrial membrane. Figure 6.4 shows a BESTFIT alignment obtained when the *Drosophila* and rat polypeptide sequences are compared. The two sequences show ~74% identity and ~80% similarity.

The transcript represented by the pD682 subclone isolated and characterised here lacks the first 50 or so amino acids present in the rat mitochondrial PCP and other cloned PCPs. It was known that the pD682 subclone ended in an *Eco*RI restriction site, and that this site did not appear to be part of an *Eco*RI linker added during construction of the cDNA library. If we examine the other DNA sequences around this region of the PCP gene, it is evident that this *Eco*RI site may actually form part of the sequence of this gene. Perhaps the original

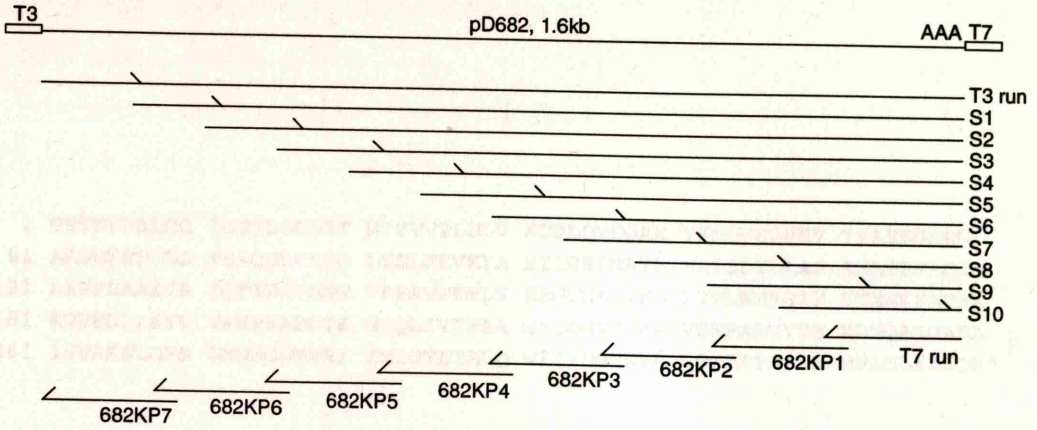


Figure 6.1 *Strategy for Sequencing of Clone pD682*

Overlapping exonuclease III deletions (using *Sst*I to produce 3' overhangs) were generated (S1-S10). Once a contig had been generated from the sequences, oligonucleotides 682KP1-7 were synthesised and these used to generate sequence from the second strand. The 682KP series of oligonucleotides point towards the 5' end of the gene. AAA; poly A tail.

```
   1 attcGGCAGC ACCAAGTACT TCGCTGTGTG CGGAATTGGT GGCATTCTCA GCTGCGGCAC
  61 CACCCACACC TTCGTGGTCC CACTGGATCT GGTGAAGTGC CGTCTGCAGG TCGACCAGGC
 121 CAAGTACAAG AATCTGGTGC ACGGATTCAA GGTCACCGTG GCGGAGGAGG GCGCCCGCGG
 181 ACTGGCTAAG GGCTGGTTCC CCACTCTGCT CGGCTACTCG GCACAGGGTC TGTGCAAGTT
 241 CGGTCTGTAC GAGTTGTTCA AGGTGAAGTA CGCCGAAATC ATTGGCGAGG AGAACGCCTA
 301 CCTGTACCGC ACCTCCTTGT ATCTGGCTGC TTCCGCTTCG GCCGAGTTCT TCGCCGATAT
 361 CGCTCTGGCG CCGTTCGAGG CCGCCAAGGT GAGGATCCAG ACTATTCCTG GATACGCCAA
 421 CAACTTCCGC GAGGCAGTGC CCAAGATGCT AAAGGAGGAG GGCGTCAATG CCTTCTACAA
 481 GGGTCTGGTT CCCCTGTGGA TGCGACAGAT CCCATACACC ATGATGAAGT TCGCTTGCTT
 541 CGAGCGCACC GTGGAGCTGC TCTACAAGTA TGTGGTGCCC AAGCCTCGTG CCGACTGCAC
 601 CAAGGGCGAG CAGCTGATCG TGACCTTCGC CGCTGGCTAC ATCGCCGGTG TGTTCTGCGC
 661 CGTGGTGTCG CATCCCGCTG ATGTGGTGGT GTCCAAGCTG AACCAGGCCA AGGGAGCTAG
 721 CGCCATCAGC GTGGCCAAGT CGCTGGGCTT CAGTGGCATG TGGAACGGAT TGACTCCTCG
 781 TATTATCATG ATCGGTACCC TGACCGCTCT GCAGTGGTTC ATCTACGATG GTGTGAAGGT
 841 CGCCCTGGGC ATTCCCCGCC CACCACCACC AGAGATGCCA GCTAGCCTGA AGGCCAAGCA
 901 GCATTAAGTG GGCTAGTCCA ATTGGTTACC TTAAGGACAG AGTTCACTAG TTGGAGCGAT
 961 TGTGGGCAGA CACAGGAGCA ACATACTTTT ACTTTCTAAC GAATTTCGGT TTCAGTATAG
1021 CGAAAATGCT TTATACTTGT TTCTCAGCAA GAATACATTA ATGTTGCGTG TGTGCCATTG
1081 CACATGTACA TTCCAAAACA ACACTTTACG TATTTCTCTA ACACAGCACT TATACACACT
1141 TTTTTATAAA CACATCTTTC TAACTTTAAA AGCCAAACAT TCTATCGTAC GAAACTAAGC
1201 TACACTCAAG TTCAAAGTAT TCAATTCCAT ATATATTCTG TTATACTATG TGCTATATTT
1261 AATCCTATGT GTGTCTAGCT TAGCAGCTGG AAAGTATATC CGTATGTTGT ACCTTTGATG
1321 GCCAGCCCCC TTGTTATCTC CTCTCCGCTG CCCCCACAGG CCACACATGC GATATGATAT
1381 TAATGCTGTT TCAATTGTCT GACCCGATAC CCAAGAAAGC ACTCTGAATA AATTAAATTA
1441 TAAAATATAA AAAACTCTAA AAAAAAAAAA AAAA 1474
```

Figure 6.2 *DNA Sequence of the Subclone pD682*

This subclone was 1474 bases in length. A 16bp poly A tail is preceded 26bp upstream by a consensus polyadenylation site (underlined). The remains of the 5' *Eco*RI site are shown in lower case. A long ORF runs from base 5 to base 907. The in-frame stop codon is also underlined.

```
  1 GSTKYFALCG IGGILSCGTT HTFVVPLDLV KCRLQVDQAK YKNLVHGFKV TVAEEGARGL
 61 AKGWFPTLLG YSAQGLCKFG LYELFKVKYA EIIGEENAYL YRTSLYLAAS ASAEFFADIA
121 LAPFEAAKVR IQTIPGYANN FREAVPKMLK EEGVNAFYKG LVPLWMRQIP YTMMKFACFE
181 RTVELLYKYV VPKPRADCTK GEQLIVTFAA GYIAGVFCAV VSHPADVVVS KLNQAKGASA
241 ISVAKSLGFS GMWNGLTPRI IMIGTLTALQ WFIYDGVKVA LGIPRPPPPE MPASLKAKQH*
```

Figure 6.3 *Predicted Polypeptide Product from the Subclone pD682*
The ORF identified within the pD682 sequence shown in Figure 6.2 is 300aa long. This ORF was predicted using TESTCODE and CODONPREFERENCE analysis.

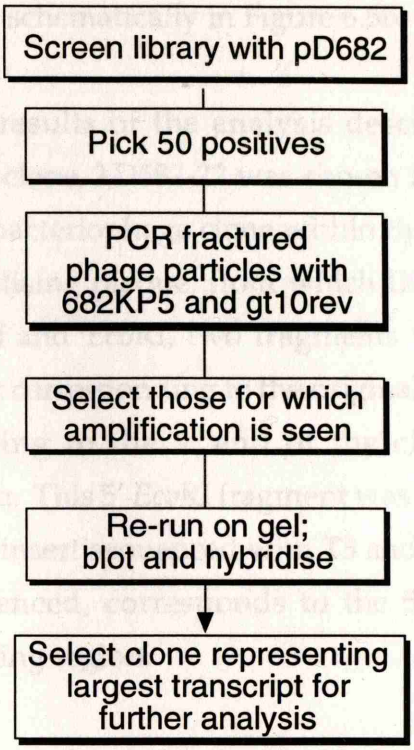
```
Dro682   1 GSTKYFALCGIGGILSCGTTHTFVVPLDLVKCRLQVDQAKYKNLVHGFKVTVAEEGARGL
           || ||:||||:||:|||| ||| ||||||||||:|||..|||.:.:||.:|: :|:.|||
Ratpcp  54 GSMKYYALCGFGGVLSCGLTHTAVVPLDLVKCRMQVDPQKYKGIFNGFSITLKEDGVRGL

Dro682  61 AKGWFPTLLGYSAQGLCKFGLYELFKVKYAEIIGEENAYLYRTSLYLAASASAEFFADIA
           |||| |||:||| |||||||:||:||. |.:|:||||.||:||||||||||||||||||
Ratpcp 114 AKGWAPTLIGYSMQGLCKFGFYEVFKALYSNILGEENTYLWRTSLYLAASASAEFFADIA

Dro682 121 LAPFEAAKVRIQTIPGYANNFREAVPKMLKEEGVNAFYKGLVPLWMRQIPYTMMKFACFE
           |||:|||||||| |||||.:|||||||.||||:||||||:.|:|||||||||||||||
Ratpcp 174 LAPMEAAKVRIQTQPGYANTLREAVPKMYKEEGLNAFYKGVAPVWMRQIPYTMMKFACFE

Dro682 181 RTVELLYKYVVPKPRADCTKGEQLIVTFAAGYIAGVFCAVVSHPADVVVSKLNQAKGASA
           |||| |||:|||||||.:|||:|||:|||.||||||||||:|||||| ||| ||..||..|
RatPCP 234 RTVEALYKFVVPKPRSECTKAEQLVVTFVAGYIAGVFCAIVSHPADSVVSVLNKEKGSTA

Dro682 241 ISVAKSLGFSGMWNGLTPRIIMIGTLTALQWFIYDGVKVALGIPRPPPPEMPASLKAKQH*
             | ..|||.|:|.|||:||||||||||||||:||| : :|||||||||.||| |
RatPCP 294 SQVLQRLGFRGVWKGLFARIIMIGTLTALQWFIYDSVKVYFRLPRPPPPEMPESLKKKLG
```

Figure 6.4 *Alignment of Drosophila and Rat Amino Acid Sequences for the PCP Polypeptide.*
Top line, *Drosophila* Sequence, Second line; Rat Sequence

bacteriophage clone λD682 did extend beyond this *Eco*RI site. If the 5′ *Eco*RI fragment was less than 100bp long, it is unlikely that it would be visible on a gel. Furthermore, a fragment from this region would not be recognised by the pD68 probe and so would not be visible on an autoradiograph. Rather than attempt to clone the 5′ *Eco*RI fragment from the bacteriophage clone λD682 (it was thought unlikely to contain all the information desired), it was decided that a significantly longer cDNA clone (containing the whole coding region) should be sought.

As the mitochondrial Phosphate Carrier Protein transcript seems relatively abundant (0.2% in the head libraries which were screened), an approach similar to that used by Hamilton *et al.*, (1991b) was thought appropriate. The strategy for isolation of longer clones was as detailed in Figure 6.5. A further 10,000 plaques were screened and 50 positive plugs picked into sterile water. Each plug contained bacteriophage representing approximately 20 different bacteriophage clones. It was assumed that in each, at least one clone related to clone pD682 would be present, whilst the remainder would be unrelated. PCR was then performed on the 50 bacteriophage plug samples. Figure 6.5a summarises the steps involved in this process. Two primers were used. One primer (gt10rev, see Section 2.5) hybridises to the bacteriophage vector at a site flanking the cDNA cloning site and adjacent to the 5′ end of the cDNA insert. The second primer used in these PCR experiments was the oligonucleotide 682KP5, one of the original sequencing primers which points upstream from around 500 bases inside the original clone pD682. Although linear amplification may occur from all templates because of the presence of the gt10rev sequence in the vector, only clones related to pD682 (and possessing a site for 682KP5) may act as templates for geometric amplification. In addition, this screening strategy allows an initial qualitative analysis of the clones amplified, because the length of the amplification products (as visualised on a gel) reflects the length of new sequence present in the bacteriophage clone. Any fragment shorter than ~500bp would represent a clone that was shorter than the original. Fragments of 500-650bp would represent longer, but still not full length clones (based on the assumption that the *Drosophila* cDNA will have roughly the same length of mitochondrial import sequence as the three mammalian PCPs). Only amplification products in excess of 650 base pairs should represent clones

(a)



(b)

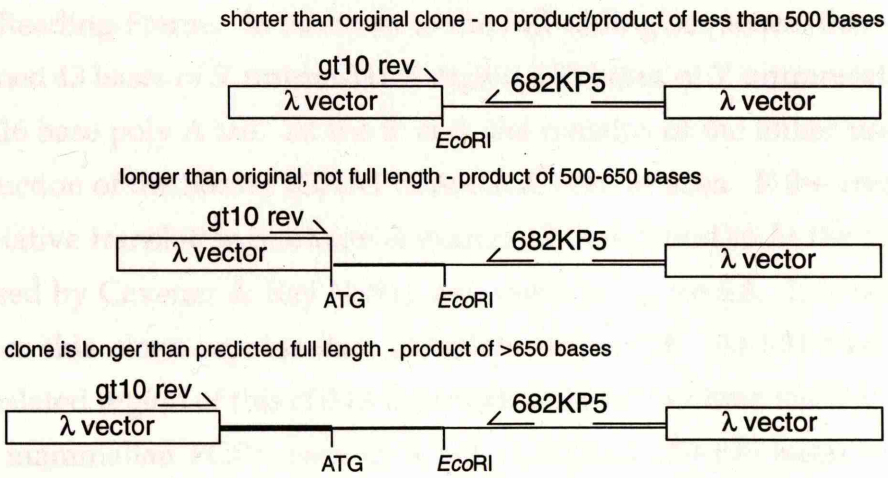


Figure 6.5 *Strategy Used to Isolate Longer Clones Representing the Drosophila Mitochondrial PCP*
(a) summary of the isolation strategy. (b) schematic representation of the different classes of amplification products which might be obtained.

containing the full coding sequence; it is these which should be chosen for further analysis. This is shown schematically in Figure 6.5b

Figure 6.6 shows the results of the analysis described above. The plug harbouring the longest clone, λD682-22 was chosen for further analysis. The positively hybridising bacteriophage clone within this plug was purified to a single positively hybridising plaque, from which DNA was isolated. Upon restriction with *Hind*III and *Eco*RI, two fragments were released; a 1500bp *Hind*III/*Eco*RI fragment corresponding to the original clone and a 200bp *Eco*RI fragment corresponding to the 5′ end of the clone and therefore the mitochondrial PCP gene. This 5′ *Eco*RI fragment was subcloned into *Eco*RI cut pBS SK$^-$ vector and the insert sequenced with T3 and T7 vector primers. This fragment, when sequenced, corresponds to the 5′ end of the *Drosophila* mitochondrial PCP coding region.

When assembled, the whole cDNA is 1684 bases long. Figure 6.7 shows TESTCODE analysis of the new sequence for the likely Open reading Frame. TESTCODE and CODONPREFERENCE analysis mirror closely the predicted Open Reading Frame. In addition to the full coding sequence, this subclone contained 43 bases of 5′ untranslated region, 551 bases of 3′ untranslated DNA and a 16 base poly A tail. At the 5′ end, the remains of the linker used in the construction of the library (GAATTCAAGGC) can be seen. If the area around the putative translation start site is examined, it is found to fit the consensus proposed by Cavener & Ray (1991), as shown in Figure 6.8. It is not known whether this clone represents a complete transcript. At 551 bases, the 3′ untranslated region of this cDNA is considerably longer than the 3′ UTRs seen in the mammalian PCPs (each of which is around 170-190 bases in length). Perhaps this region has a role to play in *Drosophila* which is not seen in the mammalian PCPs.

The sequence of the whole cDNA isolated, along with the predicted translation product, is shown in Figure 6.9. The predicted Open Reading Frame runs from base 55 to base 1116 and encodes a 353 amino acid polypeptide of molecular weight 38,5KDa. As may be expected from the high degree of homology which they exhibit, the mammalian PCPs encode proteins of almost identical molecular

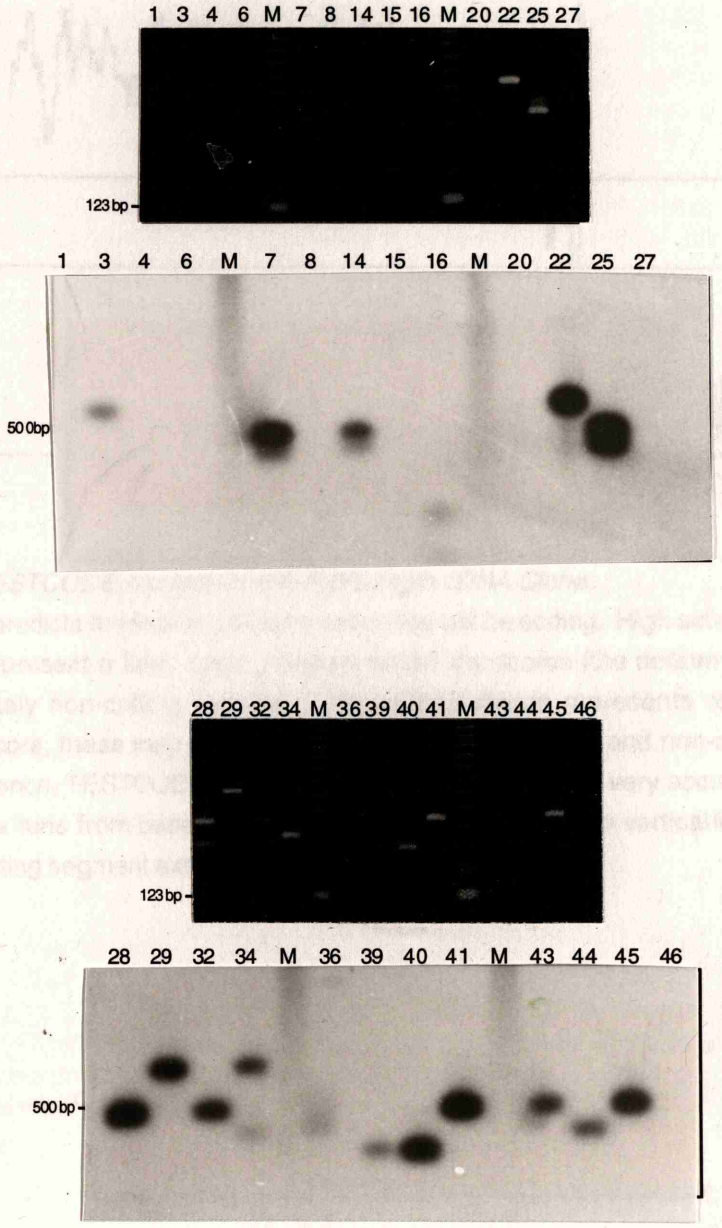


Figure 6.6 *PCR Amplification of Longer Clones Representing the Drosophila Mitochondrial PCP*
After PCR reactions were performed as described in Figure 6.5, those for which an amplification product was present were re-run on a 1.5% TBE agarose gel and the gel and the longest PCR product identified. EtBr stained gels and autoradiographs after hybridisation with pD682 DNA. Numbers above the lanes refer to the sample number. M; 123bp ladder. In this case, the longest amplification product was obtained from the clone designated λD682-22. This clone was selected for further analysis.

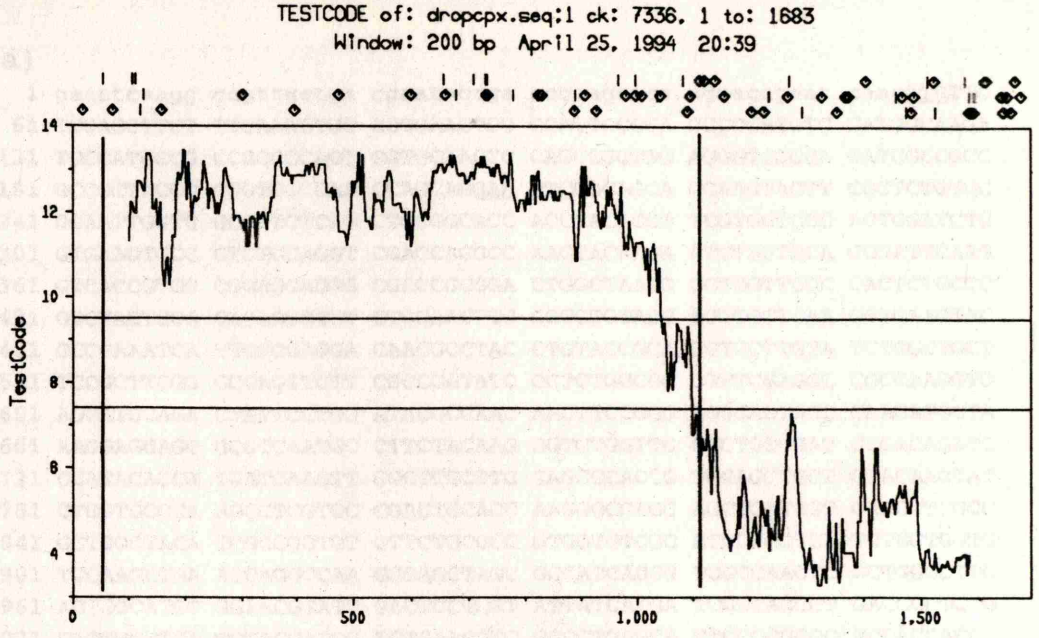


Figure 6.7 *TESTCODE Analysis of the Full Length cDNA Clone.*
TESTCODE predicts the likelihood that a sequence will be coding. High scores (the top portion of the plot) represent a likely coding region) whilst low scores (the bottom portion of the plot (represent likely non-coding regions). The middle region represents regions which give ambiguous score, these may reflect a transition between coding and non-coding sequences. For this sequence, TESTCODE predictions for this region proved very accurate. The ORF for this sequence runs from base 55 to base 1116 (between the two vertical lines. TESTCODE predicts a coding segment exactly mirroring this region.

```
Percentage 34  38  34  34  32  36  53  65  47  34  100 100 100 35  30
consensus  c a c a a c C A a a A T G g c
                 | |   | |   | | |   | | |
pD682-22   c a g a a t C A a g A T G t t
Percentage 34  38  16  34  32  22  53  65  47  19  100 100 100 20  20
```

Figure 6.8 *Comparison to Cavener's Consensus Drosophila Start Site:*
Top sequence; Caveners consensus sequence. Bottom sequence; sequence of pD682-22 around the translation start site (ATG). Upper case; consensus (found in >50% of *Drosophila* sequences), Lower case; most common base at this position. All consensus residues are conserved, along with 5/10 of the common residues (underlined). It should be noted, that bias in favour of one nucleotide need only be slight for it to be included in the 'most frequent' category. The numbers with each nucleotide represent its frequency in previously published *Drosophila* gene sequences (Cavener & Ray, 1991). A score of 25% would represent random distribution.

(a)

```
   1 gaattcaagg ccgttgctga cgaataccga ccccagaaca agaacagaat caagATGTTC
  61 TCCAGCTTCT TTGAAACTGC CCGGAACTCG CCATTCCGCA CCCCCATGTC GATGGCAAGA
 121 TGCGATGCCG CCGCCCCAGT GGTGGAACCC CAGCCGGTTG AGGGTCGCCA GATCGCCGCC
 181 GCCGCTACCC CGGTGGCCAA CCAGCAGGAA TTCGGCAGCA CCAAGTACTT CGCTCTGTGC
 241 GGAATTGGTG GCATTCTCAG CTGCGGCACC ACCCACACCT TCGTGGTCCC ACTGGATCTG
 301 GTGAAGTGCC GTCTGCAGGT CGACCAGGCC AAGTACAAGA ATCTGGTGCA CGGATTCAAG
 361 GTCACCGTGG CGGAGGAGGG CGCCCGCGGA CTGGCTAAGG GCTGGTTCCC CACTCTGCTC
 421 GGCTACTCGG CACAGGGTCT GTGCAAGTTC GGTCTGTACG AGTTGTTCAA GGTGAAGTAC
 481 GCCGAAATCA TTGGCGAGGA GAACGCCTAC CTGTACCGCA CCTCCTTGTA TCTGGCTGCT
 541 TCCGCTTCGG CCGAGTTCTT CGCCGATATC GCTCTGGCGC CGTTCGAGGC CGCCAAGGTG
 601 AGGATCCAGA CTATTCCTGG ATACGCCAAC AACTTCCGCG AGGCAGTGCC CAAGATGCTA
 661 AAGGAGGAGG GCGTCAATGC CTTCTACAAG GGTCTGGTTC CCCTGTGGAT GCGACAGATC
 721 CCATACACCA TGATGAAGTT CGCTTGCTTC GAGCGCACCG TGGAGCTGCT CTACAAGTAT
 781 GTGGTGCCCA AGCCTCGTGC CGACTGCACC AAGGGCGAGC AGCTGATCGT GACCTTCGCC
 841 GCTGGCTACA TCGCCGGTGT GTTCTGCGCC GTGGTGTCGC ATCCCGCTGA TGTGGTGGTG
 901 TCCAAGCTGA ACCAGGCCAA GGGAGCTAGC GCCATCAGCG TGGCCAAGTC GCTGGGCTTC
 961 AGTGGCATGT GGAACGGATT GACTCCTCGT ATTATCATGA TCGGTACCCT GACCGCTCTG
1021 CAGTGGTTCA TCTACGATGG TGTGAAGGTC GCCCTGGGCA TTCCCCGCCC ACCACCACCA
1081 GAGATGCCAG CTAGCCTGAA GGCCAAGCAG CATTAAgtgg gctagtccaa ttggttacct
1141 taaggacaga gttcactagt tggagcgatt gtgggcagac acaggagcaa catactttta
1201 ctttctaacg aatttcggtt tcagtatagc gaaaatgctt tatacttgtt tctcagcaag
1261 aatacattaa tgttgcgtgt gtgccattgc acatgtacat tccaaaacaa cactttacgt
1321 atttctctaa cacagcactt atacacactt ttttataaac acatctttct aactttaaaa
1381 gccaaacatt ctatcgtacg aaactaagct acactcaagt tcaaagtatt caattccata
1441 tatattctgt tatactatgt gctatattta atcctatgtg tgtctagctt agcagctgga
1501 aagtatatcc gtatgttgta cctttgatgg ccagcccccc tgttatctcc tctccgctgc
1561 ccccacaggc cacacatgcg atatgatatt aatgctgttt caattgtctg acccgatacc
1621 caagaaagca ctctgaataa attaaattat aaaatataaa aaactctaaa aaaaaaaaaa
1681 aaa 1683
```

(b)

```
   1 MFSSFFETAR NSPFRTPMSM ARCDAAAPVV EPQPVEGRQI AAAATPVANQ QEFGSTKYFA
  61 LCGIGGILSC GTTHTFVVPL DLVKCRLQVD QAKYKNLVHG FKVTVAEEGA RGLAKGWFPT
 121 LLGYSAQGLC KFGLYELFKV KYAEIIGEEN AYLYRTSLYL AASASAEFFA DIALAPFEAA
 181 KVRIQTIPGY ANNFREAVPK MLKEEGVNAF YKGLVPLWMR QIPYTMMKFA CFERTVELLY
 241 KYVVPKPRAD CTKGEQLIVT FAAGYIAGVF CAVVSHPADV VVSKLNQAKG ASAISVAKSL
 301 GFSGMWNGLT PRIIMIGTLT ALQWFIYDGV KVALGIPRPP PPEMPASLKA KQH*
```

Figure 6.9 *Sequence and Predicted Polypeptide Product from the pD682 cDNA Subclone.*
(a) DNA sequence of the full length clone. Non-coding regions are shown in lower case. The ATG start site and TAA stop codon are underlined, as is the *Eco*RI site which formed the 5' end of the original clone. The first 11 bases of this sequence constitute part of the linker used in construction of the λnm1149 library. (b) Full length predicted polypeptide sequence, representing the *Drosophila* PCP protein. The underlined Alanine residue (position 43) is thought to represent the first amino acid of the mature PCP protein.

weights. Figure 6.10 shows an alignment of this translation product to the other cloned mitochondrial PCPs. At the amino acid level, this protein shows 70% identity and 80% similarity to the previously cloned mammalian PCPs, as well as 42% identity and 69% similarity to the *S. cerevisiae* PCP (Phelps *et al.*, 1991), over the region in which they overlap (the *S. cerevisiae* PCP does not possess a mitochondrial import sequence; Phelps *et al.*, 1991). Amongst themselves, the mammalian PCPs show 90% identity and 96% similarity. In addition, they each show approximately 42% identity and 67% similarity to the *S. cerevisiae* PCP; scores very similar to those seen when the *Drosophila* PCP and the *S. cerevisiae* PCP are aligned. This suggests that the *S. cerevisiae* PCP has diverged to a similar degree from the mammalian and *Drosophila* PCPs. This relationship is summarised in the table and PILEUP derived dendrogram shown in Figure 6.11. Homology varies over the whole length of the protein, with the import sequence having diverged most.

The three mammalian PCPs each possess a mitochondrial import sequence which is cleaved once the protein has reached its final destination within the mitochondrial membrane (Schatz & Butow, 1983). In contrast, the *S. cerevisiae* mitochondrial PCP does not possess a mitochondrial import sequence, though why this is so is unclear (Phelps *et al.*, 1991). The *Drosophila* PCP cloned here bears most resemblance to the mammalian PCPs and that homology extends through the whole protein, including the import sequence. From direct analysis of PCPs and other mitochondrial proteins in mammals, it has been shown that the site of cleavage of the import sequence is after an Alanine residue, and 2-6 residues after an Arginine (Runswick *et al.*, 1987). In each of the PCP proteins, these two residues are separated by 4 amino acids. In the predicted *Drosophila* PCP polypeptide sequence, appropriately placed Arginine and Alanine residues are seen. It is likely (although no direct biochemical evidence is available), that these residues also form the cleavage site for this protein. The size of the import sequence is 43 amino acids and the mature protein has a predicted molecular weight of 34 KDa. This compares well to the predicted molecular weights of the proteins encoded by the cloned mammalian PCP genes.

Import sequences are usually acidic, although the presence of basic residues is not unknown. The *Drosophila* PCP polypeptide predicted here would contain 4

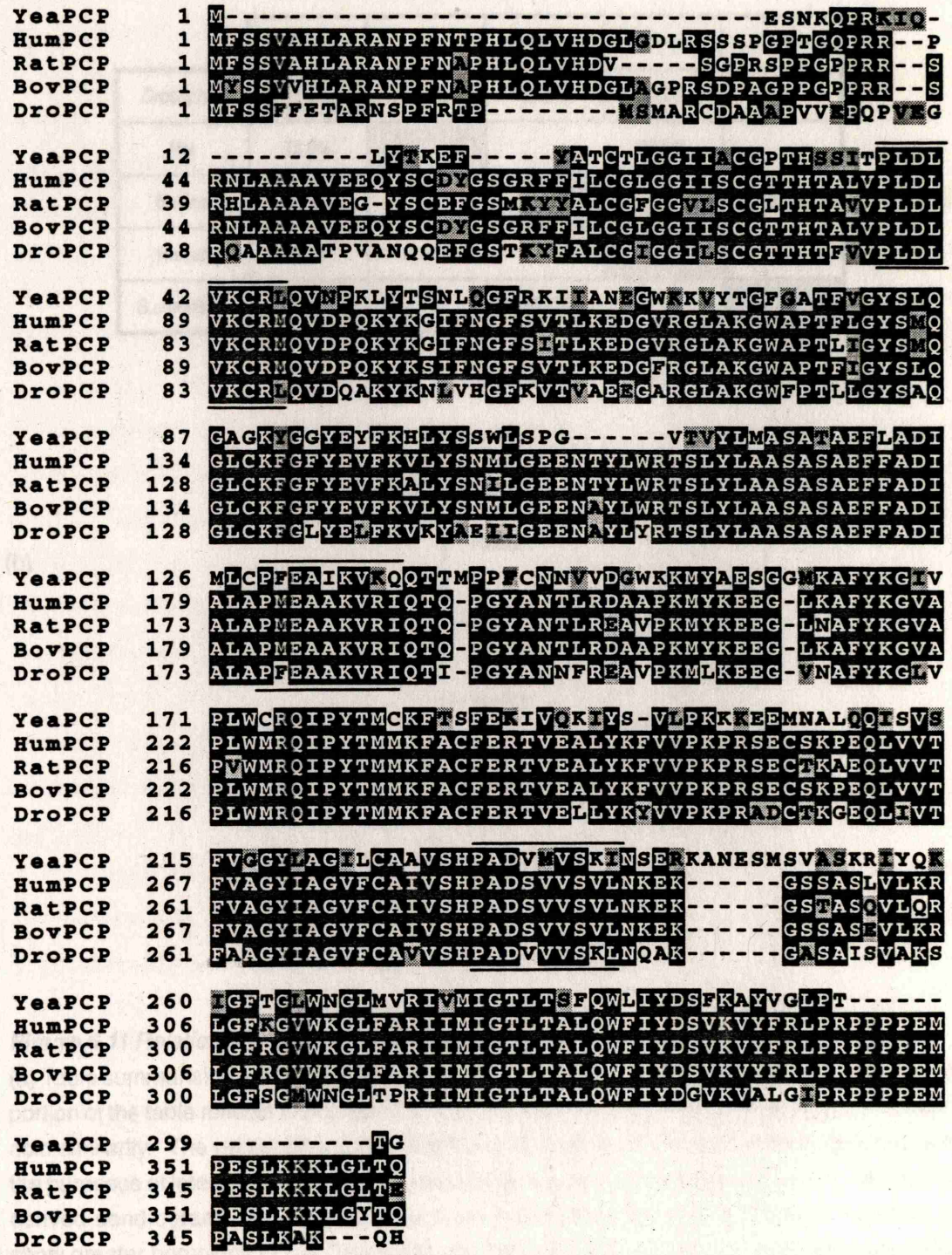


Figure 6.10 *Alignment of the pD682 Translation Product to Previously Cloned Mitochondrial PCPs.* Alignment was visualised using the BOXSHADE program. Stronger tones indicate greater consensus between the five sequences. YeaPCP, S.cerevisiae PCP peptide sequence; HumPCP, Human; RatPCP, Rat; BovPCP, Bovine; DroPCP, *Drosophila*. The bracketed residues represent the 3 copies of the mitochondrial energy transport protein subgroup signature.

(a)

	Drosophila	Rat	Bovine	Human	S.cerevisiae
Drosophila		69.5%	69.3%	70.2%	42.4%
Rat	78.9%			89.6%	
Bovine	81.7%				42.1%
Human	80%	96.7%			
S.cerevisiae	68.3	66.7%			

(b)

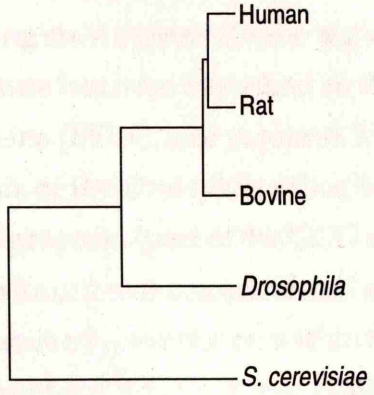


Figure 6.11 *Relationship of Cloned Mitochondrial PCPs*

(a) Table summarising the percentage similarity between the 5 cloned PCP genes. The top right portion of the table reflects amino acid identity, whereas the bottom left portion represents amino acid similarity. The rat sequence was used as an exemplar of the mammalian sequences for the purposes of internal comparison and comparison to the S.cerevisiae sequence. (b) PILEUP derived dendrogram. As expected, the three mammalian genes are clustered together yet show greater homology to the *Drosophila* polypeptide sequence than to the S.cerevisiae PCP polypeptide. In this case the relationship reflects closely the evolutionary distance of the organisms from which these polypeptide sequences were derived.

basic (Arginine) residues and 4 Acidic (3 Aspartic Acid, 1 Glutamic Acid) residues, however the sequence does have a slight net negative charge (isoelectric point 6.4). This is thought to be important for efficient transfer of the protein to its final destination within the mitochondrial membrane.

Within the mature peptide (after cleavage of the import sequence), there are many sequence features which identify this translation product as a typical mitochondrial PCP. As a group however, these proteins show more limited homology to other proteins involved in transport across the mitochondrial membrane. There are two other major groups of proteins in this family; the brown fat uncoupling proteins and the ADP/ATP translocases (reviewed in Klingenberg, 1990). As highlighted in Figure 6.10 for the PCP proteins, all of these proteins show evidence for two duplication events in their evolution. Each protein is comprised of 3 repeated units; each of approximately 100 amino acids. Each repeat encodes two membrane spanning α-helices, resulting in a final protein possessing six transmembrane regions. The most conserved region of this repeated structure has been identified as the 'signature' for 'mitochondrial energy transfer proteins (ETPs), and is found 2-3 times in all existing members of the family. Analysis of the *Drosophila* mitochondrial PCP cDNA cloned here utilising the MOTIFS program (part of the GCG suite), which searches a database of protein signatures found two complete and one incomplete instances of this ETP signature, appropriately positioned within the amino acid sequence. Figure 6.12 shows the sequences of the three copies of the signature from the mammalian, *S. cerevisiae* and *Drosophila* polypeptides, along with the consensus signature for the group. For the first copy of the signature, all of the PCP sequences fit the consensus. At the second copy, the *S. cerevisiae* sequence deviates in the last two residues. At the third copy, each sequence deviates at two positions (in each case 6 and 8) from the consensus. In addition, the *S. cerevisiae* sequence deviates in the fifth residue of the signature. Perhaps (within the PCPs) there is less functional constraint in this region of the protein than in the other copies of the signature.

The ADP/ATP translocase has been cloned from *Drosophila* (Louvi & Tsitilou, 1992) and humans (Cozens *et al.*, 1989). The protein is 296aa long and is related over its entire length to the mature polypeptide predicted from the sequence presented here. The *Drosophila* ADP/ATP translocase is 51% similar and 23% identical to the *Drosophila* PCP. This is comparable to the scores obtained when

Consensus		PXDXLRXLL		PXDXLRXLL		PXDXLRXLL	
		E IK RI		E IK RI		E IK RI	
		V V		V V		V V	
		A M		A M		A M	
		T F		T F		T F	
		Y		Y		Y	
Drosophila	I	PLDLVKCRL	II	PFEAAKVRI	III	PADVVvSkL	
Mammalian		PLDLVKCRM		PMEAAKVRI		PADSVvSvL	
Yeast		PLDLVKCRL		PFEAIKVkq		PADVmvSkI	

Figure 6.12 *Mitochondrial ETP Signature.*
The signature is composed of a sequence of 9 amino acids. Permitted residues for each position are shown in the top portion of this diagram. Deviations from the permitted residues (at positions 6 and 8 of the 3rd copy in all the sequences, and at additional residues in the S.cerevisiae sequence) are shown in lower case. The human, rat and bovine sequences are all invariant over each of the three copies of the signature, so a 'mammalian' consensus sequence is used. Sometimes only one residue is allowed elsewhere up to six different amino acids are permitted. the bottom portion of the figure shows the sequences found in the PCP's for the 3 repeats of the signature.

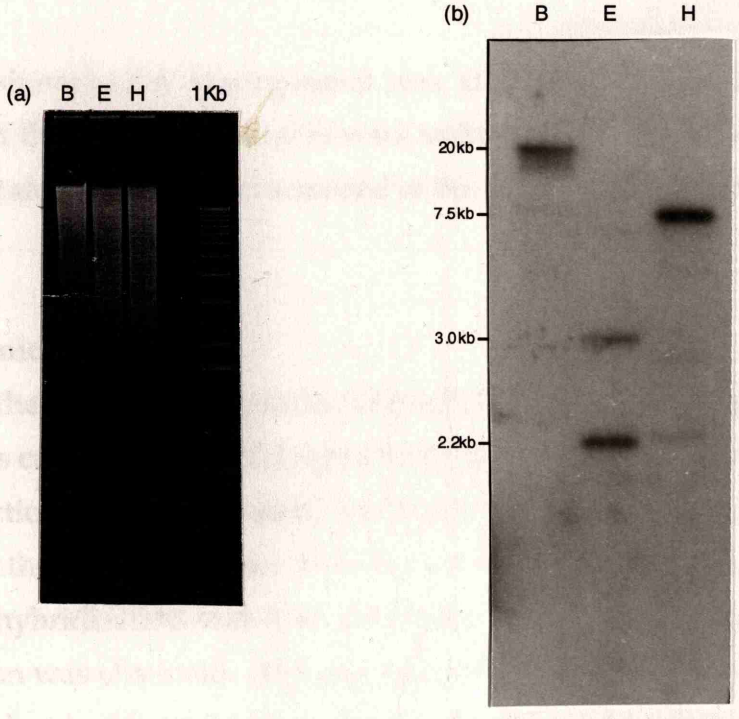


Figure 6.13 *Southern Analysis of the Region Represented by the pD682 Subclone.*
4µg *Drosophila* genomic DNA was digested with the three restriction enzymes *Bam*HI, *Eco*RI and *Hind*III. Digested DNA was separated on 0.8% TBE agarose gels, stained, photographed and transferred to nylon membrane for hybridisation with pD682 DNA. After hybridisation, filters were washed as described in Section 2.17 and autoradiographed for 72 hours. (a) Gel photograph. (b) autoradiograph. B; *Bam*HI; E, *Eco*RI; H, *Hind*III. A single band of >20kb was seen in the *Bam*HI digest, whilst the *Hind*III digest generated a single hybridising band of ~7.5kb. In the *Eco*RI digest, two smaller bands of 3.0 and 2.2 kb are identified.

the equivalent mammalian proteins are compared. Amongst themselves, the two ADP/ATP translocases share 79% homology, showing a similar degree of conservation as the PCP proteins do to each other.

The sequencing and analysis of the clone pD682 has told us what it is, and much about how it is related to previously cloned mammalian PCPs at the molecular level. This type of analysis tells us little about the functioning of this gene and the protein it encodes in the living fly. It is important that any characterisation attempts to integrate sequence analysis with other molecular analyses which might give an insight into how the mitochondrial PCP functions in the fly. One way to do this might be to create mutations at this locus, or to look for previously known mutations which affect this gene. In tandem, the expression of the mitochondrial PCP transcript might be investigated to examine whether any specific distribution control is required for the correct functioning of this gene product.

Although 'site-selected' mutagenesis was attempted in an effort to create mutations at this locus, no mutants were isolated (data not shown). Mutants were not obtained in other loci screened at the same time by other members of the lab.

6.2.2 Genomic Organisation

To examine the genomic organisation of the PCP gene, high stringency Southern analysis was carried out on wild type *Drosophila* genomic DNA cut singly with three restriction enzymes (*Bam*HI, *Eco*RI and *Hind*III). As can be seen from Figure 6.13, there is no obvious evidence for duplication at this locus. Lower stringency hybridisation was also attempted (data not shown) but no extra hybridisation was observed. Although the mammalian PCP genes all seem to be encoded by single copy genes, there is some evidence that the ADP/ATP translocase genes in *Drosophila* and other organisms are not present in single copies. Only one hybridising band is seen for *Bam*HI and *Hind*III, whilst for *Eco*RI, two bands are seen, indicating the presence of a site for *Eco*RI in this region of the genome. The probe used for this Southern hybridisation was the *Eco*RI/*Hind*III fragment from pD682; this cDNA subclone contained no internal *Eco*RI sites itself. No further genomic characterisation of this locus was carried out.

6.2.3 *in situ* Hybridisation to Polytene Chromosomes

Although no mutants were generated by SSM of this locus, mutant strains carrying defects at this locus may already have been isolated. The physical location of the PCP gene can be determined by *in situ* hybridisation of the pD682 clone DNA to spread polytene chromosomes. Once a cytological location has been assigned (by reference to cytological maps), databases may be examined to look for previously characterised genes which also map to this region of the *Drosophila* chromosome.

As seen in Figure 6.14, the *Drosophila* pD682 subclone hybridises to a region of the Left arm of the Third chromosome at position 70E. Consultation of the FlyBase database highlighted six previously recognised mutations which map to this location. These are *giant nuclei (gnu), Supressor of Variegation (Su(var)), shade (shd), stonewall (snw),* Tf_{GAGA}, and l(3)00564. Of these, Tf_{GAGA} (Soeller *et al.*, 1993) is a cloned Zinc Finger possessing transcription factor and can be discounted immediately as it bears no sequence resemblance to the PCP gene. The genes represented by the other mutant strains are as yet uncloned.

The locus represented by the mutation *gnu* (*giant nuclei*; Freeman & Glover, 1987) is thought to have a role in the regulation of nuclear division. This is a maternal effect lethal mutation; giant nuclei form in embryos from mutant mothers, whilst cytoplasmic division appears unperturbed. *Su(Var)* was the original Supressor of Variegation to be recognised (Spofford, 1967). These genes are recognised by their ability to suppress the effects of other mutations caused by the phenomenon of Position Effect Variegation (PEV). PEV is thought to occur when a gene lies close to a region of heterochromatin. Chromatin packing interferes with transcription of the gene, resulting in lowered expression. The phenotype seen is as if the gene is expressed in only a subset of genes. For example, a mottled white eye phenotype is seen with the mutation In(1)*w*[m4]. Although the *Su(Var)* gene itself has not been cloned, several other Supressor of Variegation genes have been cloned (*Su(Var)3-9*, Tschierch *et al.*, (1994); HP-1,Clark & Elgin (1992)). So far, all have been found to encode proteins which are associated with chromatin, and many (such as *Polycomb* and HP-1 (Heterochromatin Associated Protein-1)) posess a domain called the chromo-domain (Paro & Hogness, 1991) which is thought to be important for chromatin

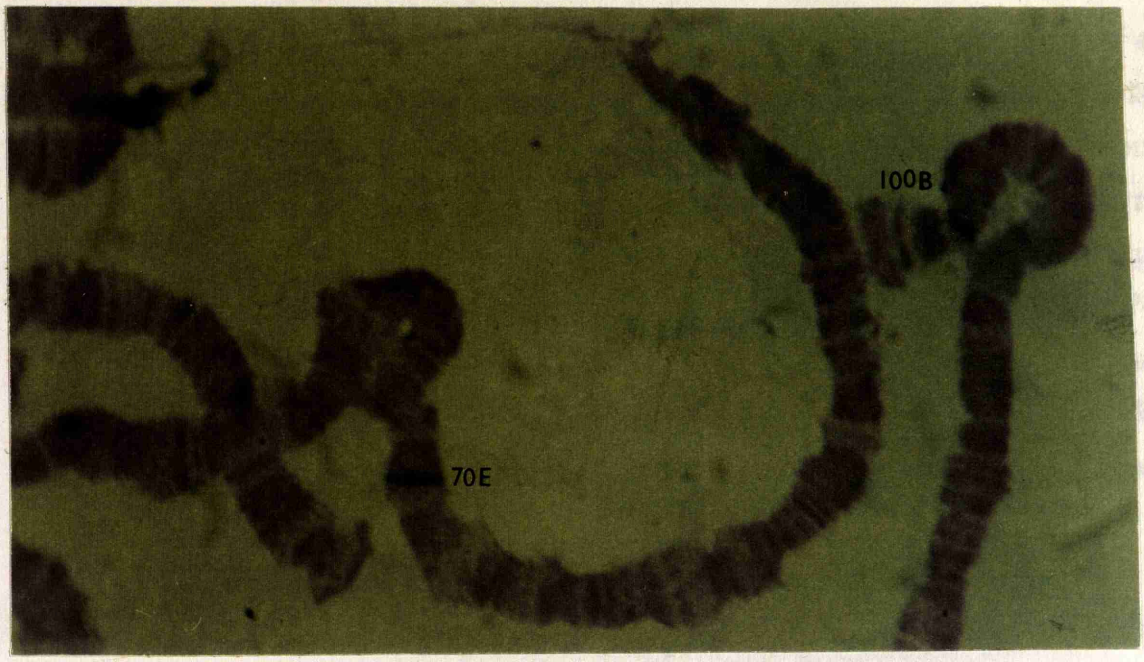


Figure 6.14 *in situ Hybridisation of pD682 to Spread Polytene Chromosomes.*
Spread polytene chromosomes from the enhancer trap line C507 were hybridised with plasmid DNA from clone pD682. Hybridisation is seen to 100B, where the P-element in this line is situated (it contains plasmid sequences). Hybridisation is also seen to band 70E, the predicted location of the *Drosophila* PCP locus. This hybridisation was kindly carried out by Zong Sheng Wang.

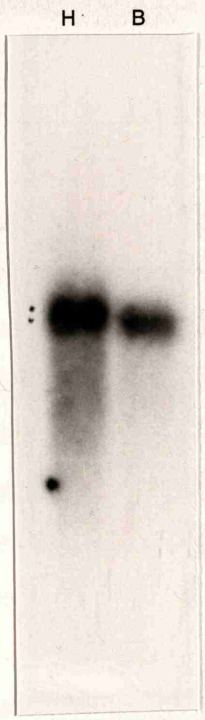


Figure 6.15 *Northern Analysis of the Transcript Represented by the pD682 Subclone.*
10µg total RNA from heads and bodies was run on 1.5% Agarose, MOPS formaldehyde denaturing gels as described in Section 2.17. RNA was transferred to nitrocellulose and hybridised with random primed pD682 DNA. After washing the filter was autoradiographed (exposure 36 hrs at -70ºC). Similar levels of expression are seen in the head (h) and body (b) tracks.

binding. The EMS induced mutation *shade* was initially recognised as part of a screen for mutants in embryonic cuticle formation (Jurgens *et al.*, 1984). *shade* is an embryonic lethal; mutant embryos show no differentiation of head skeleton or larval cuticle; although the molecular basis for this is as yet unknown (Lindsley & Zimm, 1992). The locus represented by the mutation *stonewall* (Berg & Spradling, 1989; Lindsley & Zimm, 1992) is a female sterile, caused by a P-element insertion. Although viable, females show malformed, tumorous ovaries. l(3)00564 is one of a series of lethal P-element insertions isolated by Karpen & Spradling (1992), although no information is known as to the phenotype of the mutation.

None of the genes discussed above look like obvious PCP loci candidates. It is possible that mutations at this locus would be lethal at some stage during early development and therefore the loci represented by the mutations *shade* and l(3)00564 would seem the most likely candidates. It might be possible to investigate these mutations further by examining their genomic organisation to see if they show disruption (compared to wild type) in the region covered by the pD682 subclone. This is particularly true of the line l(3)00564 as this mutation is caused by a P-element insertion and is likely to be accompanied by a significant disruption of the wild type restriction pattern at this locus.

Alternatively, a P- element construct, containing the PCP coding sequence under the control of a ubiquitous promoter could be introduced into the lethal lines by P-element transformation. In this case, the lethal phenotype (if caused by loss of functional PCP) might be rescued by the presence of a wild type copy of the PCP gene. For the chromosomal location 70E, the FlyBase database contains references to just these 6 genes. It is highly likely that other genes exist at this location for which no mutations have been recognised or for which mutations have not yet been accurately mapped.

6.2.4 Transcriptional Analysis of Clone pD682

Although the product of the gene encoded by the pD682 subclone is required in all cells, it is possible that some form of transcriptional control occurs at this locus. Indeed the possibility that the clones pD68 and pD682 might represent different transcripts makes it interesting to study the expression profile of this gene.

Although tissue *in situ* hybridisation to adult heads and whole mount embryos was carried out (data not shown), none of the results obtained suggest anything other than a ubiquitous pattern of expression. Northern blot analysis of this clone was also carried out. The Northern blot shown in Figure 6.15 shows a single band of 1.9kb present at similar levels in heads and bodies, consistent with the pattern expected for a transcript with an ubiquitous role.

6.3 Conclusions

This study has been successful in identifying and characterising a cDNA clone encoding the whole coding sequence for a previously uncloned *Drosophila* gene. Molecular approaches have allowed the investigation of protein sequence features of this gene's product and comparison of this gene's product with closely related genes in *Drosophila*. Other experiments have provided information on the expression and genomic organisation of this gene. However, as highlighted in Chapter Three it is difficult to assign function unequivocally to a piece of DNA unless molecular finding can be correlated with a physical (or at least biochemical) phenotype in a mutant fly. It is now evident that the most interesting investigations to perform with this gene would involve studying the structure-function relationship of the PCP gene product *in vivo* by mutational analysis and P-element transformation into PCP mutants. Once a PCP mutant is available and its nature known, it will be possible to consider whether *Drosophila* would provide a powerful model system in which to study the function of this gene.

Examining the relationship between of the PCP and ADP/ATP translocase genes in *Drosophila melanogaster* and other species may provide insight into the recent evolution of this gene family. These studies, coupled to the structure-function analysis described above might clarify the extent of functional constraints on variation in the third ETP signature and the rest of the protein.

Chapter Seven

Characterisation of Clones pD974 and pC133

7.1 Introduction

Subclones pC133 and pD974 represent the strongest and most weakly expressed clones chosen for further investigation. Subclone pD974 is a 1.5kb *Eco*RI/*Hind*III fragment. It was derived from the largest bacteriophage λ clone isolated in the library screen carried out with the subclone pD97. This clone was subjected to the same type of analysis as the clone pD682 introduced in Section 6.1. The transcript represented by subclone pC133 was expressed at levels in excess of those seen for the mitochondrial phosphate carrier protein described in the previous chapter. In addition, a further four clones related to this were isolated in the course of this study and the parallel study performed by Simon Tomlinson. Evidence from earlier sequencing studies of the original clone pC13 (not shown) suggested that this clone contained vector sequences. Subclone pC133 is a 900 base *Eco*RI fragment, derived from the largest bacteriophage λ clone isolated in the library screen carried out with the subclone pC13. Again this subclone was subjected to the analyses described in Section 6.1. No function has yet been assigned to these clones through the analyses described here. For this reason, they shall be treated together.

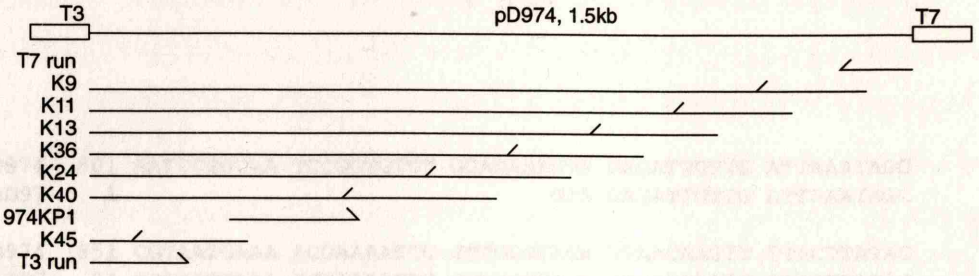
7.2 Characterisation of Clone pD974

7.2.1 Sequence Analysis

Clone pD974 contains an *Eco*RI fragment from the bacteriophage clone λD974, subcloned into pBS SK$^-$. This fragment is 1.5kb in length. Exonuclease III deletions were created as before and these were sequenced. One gap in the original sequence was 'closed' by primer sequencing using an oligonucleotide (974KP1) designed from the initial sequence. The whole clone has been sequenced in one strand, although due to overlap of the ExoIII deletions, most has been sequenced twice. The sequencing strategy and assembled DNA sequence for the pD974 subclone is shown in Figure 7.1.

Alignment of the pD974 sequence to the 5' (T3 sequence) from the original clone pD97 showed near perfect identity over the whole sequence, apart from the presence (in the pD974 subclone) of a 47 base insertion. The BESTFIT comparison of the two sequences is shown schematically in Figure 7.2. The relationship between the two sequences is shown schematically in Figure 7.3.

(a)



(b)

```
   1 AAGCAAAATA GTTAACGGCA AATGAATATT AACTTTCGGA TCGTAACAAG TAAAAGTCGA
  61 ATTTTCGAGT CTTCTTCACT TAAACACTTT CGCAACCCAA AACTAGGCCG TTGCATACGT
 121 GAGTACCTTA GAAAACAGTA ACAAAAGTTA AATAGAAATT ACGGCCTTCA CGATATCTGG
 181 CCAATATCAA AAAAACAGCA ACAGGTCCCG CTCCCCGTAA GTCATTTCGC TCAACTCGTA
 241 GTAATTCCCT CATTTCTCAC ATACAAGACT ACTAGAGATC CCCAAAATAC AATCGATTCG
 301 GTAGCATTAC TTTGCGATCC TCTGAAAGTT TCCCCCGAGA CCTAGGACAT CCATTGCGCT
 361 CATGGGGTAT ATATGTATCT CACGATCTAG AAACACTAGT GCACTCTTCA CTCGCTGAAA
 421 TTCCGCATGC ACACTATAGG CATTACGAAC AGCAGCAGCA GCAGCAGAAG CCACACTTTT
 481 CGATCTCTCA ATAGTTTCNA TAGAGANGAA TTAATTATAA TGATATTGAC ACGGATCGAT
 541 GATGGTCATT CGAAAGCGCG GAACATTAGC AAAAAACAAG CAAAACATCA TGGAAACTCA
 601 TTCAAAGTAA AAGTATAAAT CTNTTATGAT TTGACTGTGA TAATAATAAT GACAAAATGA
 661 GAAATGAAGT CGGAAGGTGT AAACAAATTA ACGTATATAG NGTATTTAAA TGTATGTAAA
 721 TACAATGCAA ATGTAATATT AACCTAGGTG CGAAACGAAA TGATGGCTCC AAGAAAACAG
 781 CGCGGAATAA GGCAAATAGC AATCCGGCAA TCCGGTGTGT GGAGAAAGTA GAGATTGTTG
 841 ATTAAATAGC CGTAATGAAA ACGAAAATTC GTTGCGCAGA GTAACAAGTT TTACTTATAC
 901 ATATTAGCAA AATATTTAAA TAAAATTTTA GCAAACCAAA TCCCTTTCTG AGTAGAAACG
 961 AACCCAAAAT AAAAGGCAAC TGAAGGTCCT AATTGAAGTA AACAAATTAA ATGCGAAATG
1021 AAATCAACGA TATTTAGTTA TGCTCTATTT ACAAAGTATA ATCAATTGAA GTTTTTTAAA
1081 AATCAAGTAG CAAAGCAAAC GATTTTATAG ATTTATTCAT TTAGCATTTG AAAATCGCTA
1141 TGGTAAAAGT TCACACATAA CCACGGTATA AAATAGGTTT TCTTCATTTC GCTCAATTTC
1201 ATTTAAAGAG GTATAGATTT GAACAAAAAA AAAAAAACAA AAACCAAAAC GTACGACAAG
1261 TAACAGAAAT AGTCTAGGTG AATCTTTAAA CAATTAAACA AAAGGTATCC CTACAAGAAG
1321 ATTCTTTAAG TAGCTCATTC AAGTTTTAAT AAATCGAGAA AAACATAATT TTTGCAATTA
1381 AACAAAACGA TAAGAAATTC TGTACAATTG CGAACGAAAA GTATGCAAAC AGTTACTCCT
1441 CTCTACATAC ATATATATGC ATAGCTATAT AACCCCAAGT GAATATACTC GAACGCTT
```

Figure 7.1 *Strategy for Sequencing and Assembled Sequence of Clone pD974*

(a) Overlapping exonuclease III deletions (using *Kpn*I to produce 3' overhangs) were generated (K9, K11, K13, K36, K24, K40, K45). Once two contigs had been generated from their T7 sequences, the oligonucleotide 974KP1 was synthesised and used to complete the sequence in one strand . The 974KP1 oligonucleotide points toward the 3' end of the gene (as orientated by alignment with the original clone pD97). (b) 1498bp in length. This sequence is 65.8% A/T rich. The location of the oligonucleotide 974KP1 is underlined. This sequence is aligned 5'-3'.

```
pD974   801 AATCCGGCAA TCCGGTGTGT GGAGAAAGTA GAGATTGTTG ATTAAATAGC
 pD97     1                                GTA GAGATTGTTG ATTAAATAGC

pD974   851 CGTAATGAAA ACGAAAATTC GTTGCGCAGA GTAACAAGTT TTACTTATAC
 pD97    24 CGTAATGAAA ACGAAAATTC GTTGCGCAGA GTAACAAGTT TTACTTATAC

pD974   901 ATATTAGCAA AATATTTAAA TAAAATTTTA GCAAACCAAA TCCCTTTCTG
 pD97    74 ATATTAGC

pD974   951 AGTAGAAACG AACCCAAAAT AAAAGGCAAC TGAAGGTCCT AATTGAAGTA
 pD97    82       AAACG AACCCAAAAT AAAAGGCAAC TGAAGGTCCT AATTGAAGTA

pD974  1001 AACAAATTAA ATGCGAAATG AAATCAACGA TATTTAGTTA TGCTCTATTT
 pD97   127 AACAAATTAA ATGCGAAATG AAATCAACGA TATTTAGTTA TGCTCTATTT


pD974  1051 ACAAAGTATA ATCAATTGAA GTTTTTTAAA AATCAAGTAG CAAAGCAAAC
 pD97   177 ACAAAGTATA ATCAATT 194
```

Figure 7.2 *Alignment of pD974 and pD97 Sequences Over Their Region of Overlap.*
Top line; pD974. Bottom line, pD97. A 47 base segment (bases 909-955) is seen in pD974 but absent from pD97.

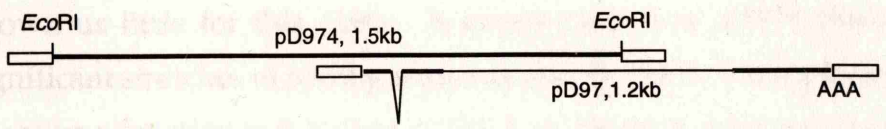


Figure 7.3 *Schematic Representation of the Relationship of pD97 and pD974*
The 2 cDNA inserts are shown, with the open boxes representing vector sequences. The heavy lines indicate regions for which sequence information is available. The two sequences overlap by approximately 670 bases. There is no *Eco*RI site in the pD97 clone corresponding to that at the 3' end of pD974.

If we examine the pD974 sequence in Figure 7.1, it is immediately apparent that this cDNA insert is very AT rich (65.6%). Moreover, it seems that there are many runs of Adenosine or Thymidine residues dispersed throughout the cDNA. Over 1498 bases, there are 62 runs of 5 or more A/T residues, compared to 8 C/G runs. Neither TESTCODE or CODONPREFERENCE were able to identify likely coding regions within this fragment. If the sequence is searched for the presence of translation stop sites (TAG, TAA or TGA codons), 93 are found within the 1498 bp sequence. 93 stops in 1498 bp is equivalent to 1 STOP codon in each frame every 48 bases, an average ORF length of 16 amino acids.

The lack of a candidate Open Reading Frame meant that efficient searching of ProSite and SwissProt could not be carried out. However database searching was carried out. The program FASTA was used to search for similar DNA sequences in the GenBank database. The program FASTA was also used to search for similarities between translation products of the three forward frames (with stops removed) and sequences in the ProSite and SwissProt databases. No significant homology was found in any of these searches. The only homology of any interest was that shown by a short OPA repeat. Bases 448-467 of the pD974 sequence presented in Figure 7.1 encode a short ($(CAG)_6$) perfect poly-Glutamine stretch. There are however 'in frame' stop codons 18 amino acids 5' and 9 amino acids 3' of this repeat. This OPA repeat may be untranslated.

In contrast to the clone pD682, examined in Chapter Six, sequence analysis has shown us little for this clone. It seems clear that pD974 does not contain significant stretches of coding sequence and therefore it has proved impossible to assign a function to this piece of DNA on the basis of homology to previously cloned genes.

As before, 'site-selected' mutagenesis was attempted for this locus, but without success. It is for clones such as this, where function is completely unknown, that 'site-selected' mutagenesis becomes a powerful and attractive tool. Further attempts at 'site-selected' mutagenesis would be a vital element of future investigation of this locus.

7.2.2 Genomic Organisation of pD974

High stringency (to 0.1 x SSC) Southern analysis was carried out on wild type *Drosophila* genomic DNA cut singly with three restriction enzymes (*Bam*HI, *Eco*RI and *Hind*III) in order to examine the genomic organisation of the locus represented by pD974. As can be seen in Figure 7.4, (although hybridisation is obscured a little) it is evident that only a single band hybridises in each lane, suggesting that this gene is present as a single copy in the genome. No further genomic characterisation of this clone was carried out.

7.2.3 *in situ* Hybridisation to Polytene Chromosomes

The physical location of the pD974 subclone was determined by *in situ* hybridisation to polytene chromosomes. Figure 7.5 shows spread chromosomes hybridised with clone pD974 DNA. Hybridisation can be seen at polytene band 85D. Unlike the mitochondrial PCP described in Chapter 6, we cannot use our knowledge of the function of the product of this locus to aid our evaluation of the candidate loci. Some of these loci (e.g. *ova*, *sic* and ms(3)85D) can be discounted on the basis of their phenotype. Other loci have been previously cloned and sequenced yet show no homology to the pD974 subclone. The loci which map to this region of the genome are shown in Figure 7.6. None of these loci are known to have specific role within the nervous system.

This region of the genome has been well characterised genetically, and there are a number of lethal mutations which map to 85D. As before, it is possible that the locus represented by pD974 gives rise to a lethal phenotype when mutated. Any future investigation of this locus might begin by examining the large numbers of lethal P-element insertions generated by Karpen & Spradling (1992) for evidence that they disrupt the wild type Southern hybridisation pattern shown in Figure 7.4. If any of these insertions do reside close to or within the locus represented by pD974 then remobilisation of the P-elements in these strains may generate weaker phenotypes which can be characterised.

7.2.4 Transcriptional Analysis of Clone pD974

As no Open Reading Frame has been identified, it would be useful to have some idea of the size of the transcript or transcripts from the locus represented by the pD974 clone. This might give an idea of what size of transcript is being

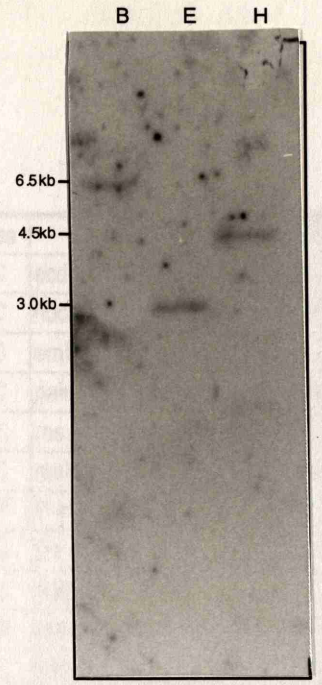


Figure 7.4 *Southern Analysis of the Region Represented by the pD974 Subclone.*
4µg *Drosophila* genomic DNA was digested with the three restriction enzymes *Bam*HI, *Eco*RI and *Hind*III. Digested DNA was separated on 0.8% TBE agarose gels, stained, photographed and transferred to nylon membrane for hybridisation with pD974 DNA. After hybridisation, filters were washed as described in Section 2.17 and autoradiographed for 72 hours. B; *Bam*HI; E, *Eco*RI; H, *Hind*III. A single band of 6.5kb was seen in the *Bam*HI digest, whilst the *Eco*RI digest generated a single hybridising band of 3kb. In the *Hind*III digest, a band of 4.5kb is identified.

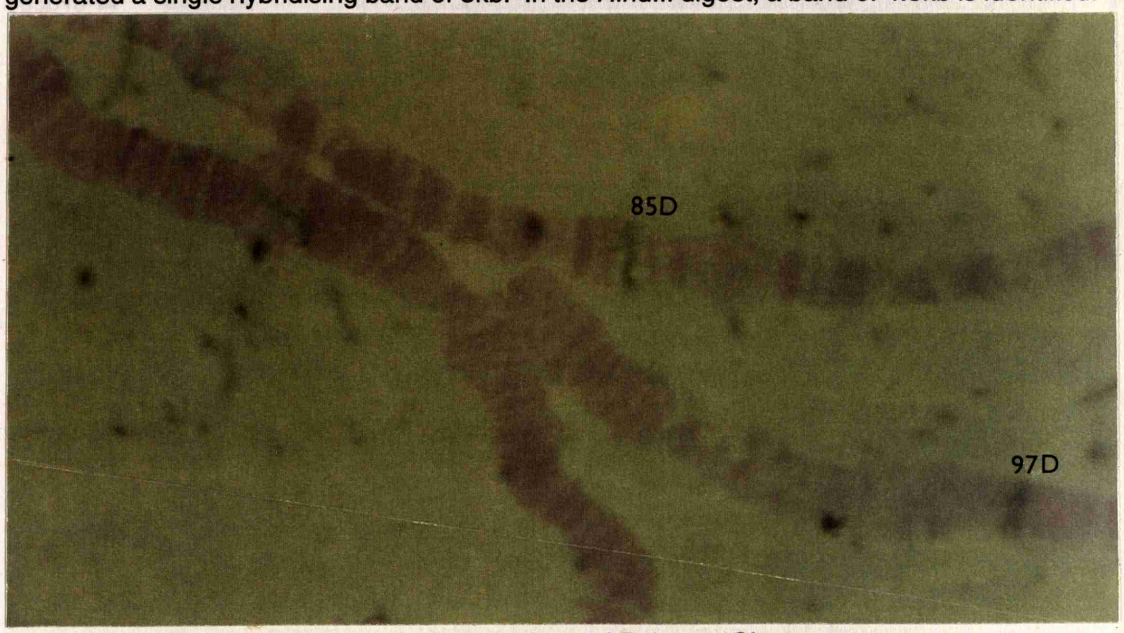


Figure 7.5 *in situ Hybridisation of pD974 to Spread Polytene Chromosomes.*
Spread polytene chromosomes from the enhancer trap line 187Y were hybridised with plasmid DNA from clone pD682. Hybridisation is seen to 97D, where the P-element in this line is situated (it contains plasmid sequences). Hybridisation is also seen to band 85D, the predicted location of the locus represented by pD974. This hybridisation was kindly carried out by Zong Sheng Wang.

Location	Gene symbol	des	Description, reference
84E7-87E11	*ImpE3*	C	ecdysone inducible gene, Natzle et al., 1988
85C-D	*Hrb85CD*	C	ribonucleoprotein destabiliser; Lindsley & Zimm, 1992
85C-D	*pum (pumilio)*	C	embryonic lethal; MacDonald, 1992
85C-D2	*prd7*	C	paired like 7; Frigerio et al., 1986
85D	*Ras85D*	C	ras oncogene homolog; Neumann-Silberberg et al., 1984
85D	*betaTub85D*	C	male specific beta tubulin; Kemphues et al., 1982
85D	*ova (ovarette)*	P	P-insertion assymetric cell divsion in females; Lin & Spradling, 1992
85D1-2	*D1*	C	D1 chromosomal protein; Ashley et al., 1987
85D1-27	*l(3)neo37*	L	lethal p-element insertion Cooley et al., 1992
85D1-27	*ms(3)85D*	S	male sterile P-element insertion, Castrillon et al., 1993
85D2-3	*l(3)01688*	L	lethal P-element insertion, Karpen & Spradling, 1992
85D5-6	*l(3)04837*	L	lethal P-element insertion, Karpen & Spradling, 1992
85D7-8	*l(3)01728*	L	lethal P-element insertion, Karpen & Spradling, 1992
85D7-8	*l(3)03559*	L	lethal P-element insertion, Karpen & Spradling, 1992
85D8-12	*sic*	L	sichel, maternal effect embryonic lethal, Lindsley & Zimm, 1992
85D8-13	*l(3)85Da*	L	X-ray induced lethal, Lindsley & Zimm, 1992
85D8-9	*l(3)05430*	L	lethal P-element insertion, Karpen & Spradling, 1992
85D10-13	*Fps85D*	C	fps oncogene homolog; Katzen et al., 1991
85D10-14	*l(3)85Dc*	L	X-ray induced lethal, Lindsley & Zimm, 1992
85D11-12	*l(3)85Db*	L	X-ray induced lethal, Lindsley & Zimm, 1992
85D11-17	*alpha-Man-II*	C	alpha mannosidase II, Foster et al., 1995
85D11-E3	*by*	R	blistery, Lindsley & Zimm, 1992
85D11-E3	*l(3)85Dd*	L	X-ray induced lethal, Lindsley & Zimm, 1992
85D11-E3	*l(3)85De*	L	X-ray induced lethal, Lindsley & Zimm, 1992
85D11-E3	*l(3)85Df*	L	X-ray induced lethal, Lindsley & Zimm, 1992
85D11-E3	*l(3)85Dg*	L	X-ray induced lethal, Lindsley & Zimm, 1992
85D11-E3	*l(3)85Dh*	L	X-ray induced lethal, Lindsley & Zimm, 1992
85D11-E3	*l(3)85Di*	L	X-ray induced lethal, Lindsley & Zimm, 1992
85D11-E3	*l(3)85Dj*	L	X-ray induced lethal, Lindsley & Zimm, 1992
85D17-19	*l(3)s01747*	L	lethal P-element insertion, Karpen & Spradling, 1992

Figure 7.6 *Loci Which Map to Band 85D of the Drosophila Polytene Chromosome.*
This region of the genome has been extensively mutagenised and has been well characterised.des; Designation: C, cloned; L; lethal; S; sterile, P; P-element insertion; R, rearrangement.

sought. Unfortunately, Northern analysis using the pD974 subclone as probe (though attempted a number of times) was never successful.

in situ hybridisation to whole mount embryos did however give a specific pattern of expression. As seen in Figure 7.7, transcription of the product represented by the pD97 clone (the original clone) is largely restricted to the Central Nervous System in late (stage 16) embryos.

in situ hybridisation to frozen head sections was also carried out. As can be seen from Figure 7.8, the pD974 antisense probe hybridises to cell bodies throughout the brain, consistent with a general nervous system expression pattern.

7.2.5 Considerations and Conclusions

Examining the data as a whole it is perhaps prudent to consider the possibility that the pD974 clone does not represent a cDNA but is in fact a fragment of genomic DNA. Although circumstantial, there are a number of pieces of evidence to support this. The A/T richness of the sequence (as already highlighted) indicates that this sequence is not coding. It may therefore constitute part of a very long 3' UTR or it may be a piece of genomic DNA, containing short exons interspersed with long intronic sequences. The OPA repeat found in pD974 might be part of a short exon, which would explain why it cannot be included in any long Open Reading Frame. Also, the 47bp insertion in pD974, which differentiates it from the pD97 sequence might be an intron. Assuming that pD97 represents a real cDNA, then the intron in pD974 would separate two exons found in pD97. Examining the sequences at the junctions of the 47bp insertion does identify a splice acceptor site (as reviewed in Padgett *et al.*, (1986)), although no splice donor site is seen at the 5' end of the inserted sequence. A genomic fragment might be cloned by false priming of genomic DNA at an A/T rich region. At the 3' end of the pD974 subclone is an *Eco*RI site, no *Eco*RI/*Hind*III fragment could be seen in the digests of the λD974 bacteriophage clone. Perhaps this is not a real *Eco*RI site and is an artefact generated during the 'forced' cloning of a genomic fragment; the *Eco*RI site is not present in the parent clone pD97.

Figure 7.7 *in situ Hybridisation to Whole Mount Embryos*
Embryos were collected as described in Section 2.20. A single stranded probe was made using the PCR protocol described in Section 2.20. Plasmid pD97, along with the T7 promoter was used, generating an antisense probe. Expression becomes localised after stage 14 and is largely restricted to the Central Nervous System (with diffuse staining outside this region) by stage 16 as shown.

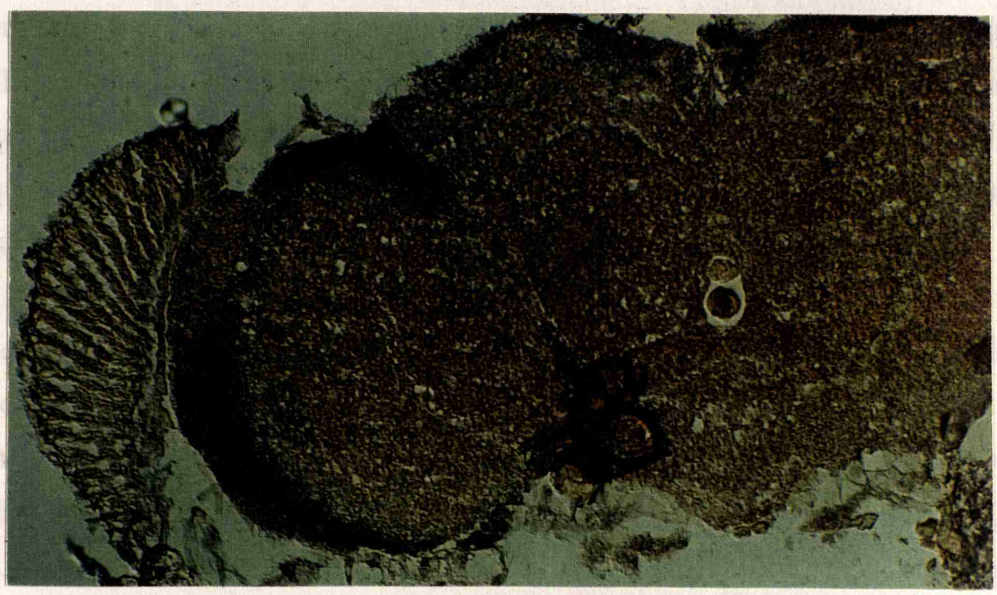


Figure 7.8 *in situ Hybridisation to Adult Head Cryostat Tissue Sections*
A single stranded probe was made using the PCR protocol described in Section 2.20. Plasmid pD974, along with the T7 promoter was used, generating an antisense probe. This probe was used to hybridise to head sections as described in Section 2.21. Strong expression is seen in cell bodies throughout the brain, possibly reflecting a general nervous system expression pattern.

Together these observations lend weight to the possibility that pD974 is a genomic fragment, however it should be stressed again, that the evidence is all circumstantial and inconclusive. Indeed it may be argued that the possible intron at position 908-955 is no more AT-rich than the pD974 subcloned fragment as a whole. Neither is this 47 base pair intron any more AT-rich than the 47 bases immediately upstream or downstream of this insertion. Also, it is known that not all OPA repeats are translated and in any case, a repeat of 6 Glutamines is rather short compared to most OPA repeats already known (in those sequences studied by Gerber *et al.* (1994), sequences with an OPA repeat length of less than ten were not considered). Further investigation (analysis of the other bacteriophage clones isolated in the original screen) would be required to resolve this uncertainty. This analysis should be carried out before any further analysis of the locus represented by the pD974 subclone.

7.3 Characterisation of Clone pC133

7.3.1 Sequence Analysis

Subclone pC133 represents a 0.9kb *Eco*RI fragment from the bacteriophage clone λC133. A 100 base pair *Eco*RI/*Hind*III fragment, believed to originate from the 3' end of this cDNA clone was not subcloned. Sequencing of the original clone pC13 was hampered when it was found that the insert fragment contained vector sequences. As a result, characterisation of this clone was delayed and the sequencing and analysis of this clone was not taken as far as some of the other clones in this study.

As described already, a parallel study had isolated three related clones typified by one clone pST170 (Tomlinson, 1994). pST170was sequenced entirely in one strand. It is 598 bases in length, including a 23 base poly A tail. At position 468 in this sequence is an internal *Eco*RI restriction site. An oligonucleotide (170ol1; CCAACAGCAGCCGCAAG) derived by Simon Tomlinson from position 237 to 253 of this fragment had been used to complete the sequencing of this clone. In an attempt to further investigate the relationship between pST170 and pC133, this oligonucleotide was used as a sequencing primer. In addition, terminal sequencing with the vector primers T3 and T7 was performed. As would be expected, the sequence generated by the oligonucleotide 170ol1 was identical

to the equivalent sequence generated from pST170. Sequence from the T7 primer corresponded to that of pST170 upstream of the *Eco*RI site at position 468 of the original sequence. This suggests that this *Eco*RI site is at the 3' of the pC133 clone, confirming the prediction that this clone would lack sequences at the 3' end of the transcript. At the T3 end of pC133, 246 bases of sequence were obtained which were not homologous to any sequence derived from this or the pST170 clone. The sequence obtained, along with the sequence of the pST170 clone and a schematic representation of their relationship is shown in Figure 7.9. Together, these sequences represent 750 bases at the 3' end of the gene represented by this family of cDNA clones. All the pC13 related sequences were searched exhaustively for long Open Reading Frames without success, Unfortunately, time prevented the further sequence characterisation of this clone.

7.3.2 Genomic Characterisation of Clone pC133

High stringency (to 0.1 x SSC) Southern analysis was carried out on wild type *Drosophila* genomic DNA cut singly with three restriction enzymes (*Bam*HI, *Eco*RI and *Hind*III) in order to examine the genomic organisation of the locus represented by pC133. As can be seen in Figure 7.10, it is evident that at most, two bands hybridise in each lane, suggesting that this gene is present at a single locus in the genome. No further genomic characterisation of this clone was carried out.

7.3.3 *in situ* Hybridisation to Polytene Chromosomes

The physical location of the pC133 subclone was determined by *in situ* hybridisation to polytene chromosomes. Figure 7.11 shows spread chromosomes hybridised with clone pC133 DNA. Hybridisation can be seen at polytene band 74A. In contrast to pD974 described earlier, this region of the *Drosophila melanogaster* genome has not been extensively studied and indeed only one gene *Arylsulfatase* (*Ars*; Lindsley & Zimm, 1992) is known to map here. Although not cloned, phenotypic studies suggest that this locus encodes the enzyme arylsulfatase. As it has not been cloned, it is not known whether the locus encoding this enzyme is also the origin of the pC133 cDNA. Other arylsulfatase genes have been cloned (e.g. Chang *et al.*, 1990) however although comparison of the pC133 sequence to them does not identify any significant similarity. As before, no mutants were obtained when 'site-selected' mutagenesis

(a)

pC133 T3 (5' end of pC133 subclone)

```
  1  AAGCTGCATT CCGCAGCAAC GACGGTTGGG CACATCCCGT CTCATCCCCA
 51  CTCTCTCGAA TTGTGTCCGT TGGTAGCCGA AGTTACGATT GCAAAACTTT
101  CGTATCTCTC TCTCACTCTC TCAAGCAAAA GCCAAGGCCG GAGCAGGTGC
151  ATATGGTGCC AACCCCCCTG GCCCCCCCAA AAGGGGATGG TNNTGGTGCA
201  ATTGCAAAGG CACAGGAGCA TAGGAGCAGC TAAATAAACA AGAAAG
246  BASES
```

pST170 sequence

```
  1  AAAATCATCA TCGTGTTATC AGGACGACCG CAAAAAATCT CTGAAAATCG
 51  TACAAAAAAA TATTAAAACC TCAAAATGCA TTGGTGTGTG TGTCCCGTTC
101  GTGTAATGCC TGCGCGCCAC GCCCCCCGCC CCTCCCACTG TCACGCGCTC
151  GCCACGCCCC CCAGTTGCCG CCCCATTAGC AGCGGGCAGA TGGACAACGT
201  GTGGGTGGCC CACAAGGACA TCTAGTAACA CGACGCCCAA CAGCAGCCGC
251  AAGGTCTGAT ACGCCGCCCG CCACGCCCAT CGTGTTTGGG CGGCAGAGGA
301  AGCGGTCCGG AAGCGGAAAC GGAAACGGAA ACGGGCGAGC AAAATGGTGG
351  CGCGGTATCG CGGGCAAGGC GACGGCGCGT CCACAAAAAA TAACCATAGA
401  GACGTTTAAG GCAAATTCTA AATGAACAAA AGTATAAACC AAAAACAATT
451  GTAACGAGAA AACAAACGAA TTCAAAAAAA TGAGCAATAT GAGCAACAAC
501  TAATAATGGC AAAAAAAGCA AAAACGACAA CTGCAAATTA CGACACAACA
551  CCTTCGAAAA AGACCTCAGT AATAATTTTT TTTTTTTTTT TTTTTTTT
598  BASES
```

(b)

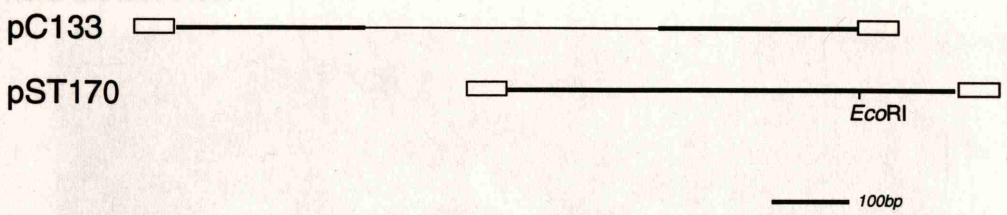


Figure 7.9 Sequence of Clone pC133 and its Relationship to pST170

(a) Sequence from the 5' end of this clone generated 246 bases which showed no homology to the original pST170 sequence quoted below. These sequences are thought to be separated by approximately 150 bases of DNA. pST170 sequence confirmed by sequencing the equivalent region of pC133 is shown italicised. The *Eco*RI site which defines the 3' end of pC133 is underlined (468-473). The oligonucleotide 170ol1 is also shown underlined (bases 237-253).

(b) Schematic representation of the relationship above. Thicker lines represent sequence regions of DNA.

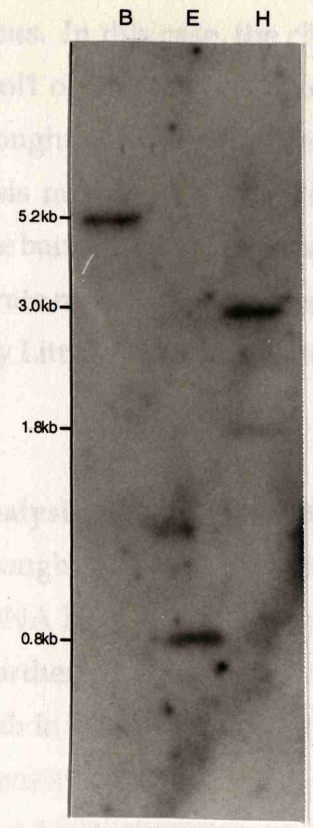


Figure 7.10 *High stringency Southern Analysis of the Region Represented by the pC133 Subclone.*
4µg *Drosophila* genomic DNA was digested with the three restriction enzymes *Bam*HI, *Eco*RI and *Hind*III. Digested DNA was separated on 0.8% TBE agarose gels, stained, photographed and transferred to nylon membrane for hybridisation with pC133 DNA. After hybridisation, filters were washed as described in Section 2.17 and autoradiographed for 72 hours. B; *Bam*HI; E, *Eco*RI; H, *Hind*III. A single band of 5.2kb was seen in the *Bam*HI digest, whilst the *Eco*RI digest generated two hybridising bands of 800bp and 3kb. In the *Hind*III digest, bands of 3.0kb and 1.8kb are identified.

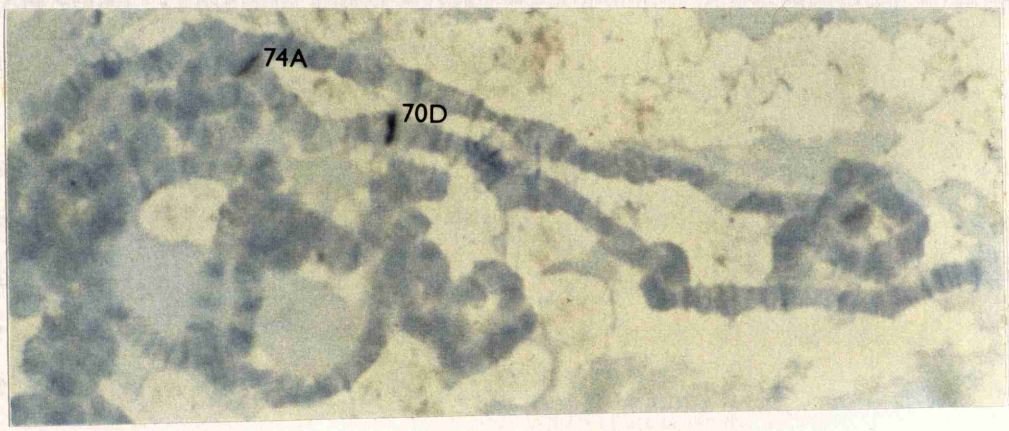


Figure 7.11 *in situ Hybridisation of pC133 to Spread Polytene Chromosomes.*
Spread polytene chromosomes from the enhancer trap line 208Y were hybridised with plasmid DNA from clone pD682. Hybridisation is seen to 70D, where the P-element in this line is situated (it contains plasmid sequences). Hybridisation is also seen to band 74A, the predicted location of the locus represented by pC133.

was attempted at this locus. In this case, the chances of success were reduced anyway because the 170ol1 oligonucleotide points towards the 3' of the gene whilst P-elements are thought to favour the 5' ends of genes. An alternative to 'site-selected' mutagenesis might be to look for existing mapped P-element insertions at this polytene band (or within one subdivision; 73F-74B) and try to remobilise them to generate new local P-element insertions. This strategy has been used successfully by Littleton *et al.* (1993) to mutagenise the *synaptotagmin* locus.

7.3.4 Transcriptional Analysis of Clone pC133

The pC133 subclone is thought to represent a very abundant transcript, present at high frequency in cDNA libraries and responsible for a strong signal on Reverse Northerns. Northern analysis shows a very strongly hybridising transcript of around 1.3kb in length in head RNA. A faint band of the same size (at around 1/10 intensity) is seen in body RNA as shown in Figure 7.12. This finding is reflected by a similar experiment carried out by Simon Tomlinson using the subclone pST170 as probe. From this analysis, it was estimated that this transcript might account for 0.5-1% of all mRNA in heads (Tomlinson, 1994).

in situ hybridisation to whole mount *Drosophila* embryos was performed as described before. In this case (as seen in Figure 7.14) an unlocalised expression before Stage 13 embryos seems to become restricted to the Nervous System of later embryos (Stage16).

in situ hybridisation to frozen adult head tissue sections was also performed. Sense and antisense probes were made as described in Section 2.20 and hybridisation carried out as described in Section 2.21. As with clone pD974, the antisense probe hybridises to cell bodies throughout the brain, including eye tissue. No significant localised expression is seen with the sense probe.

7.3.5 Conclusions

Although hampered by the lack of complete sequence for this clone, much is now known about its pattern of expression and genomic organisation. In addition, as the chromosomal location of the gene represented by this clone is

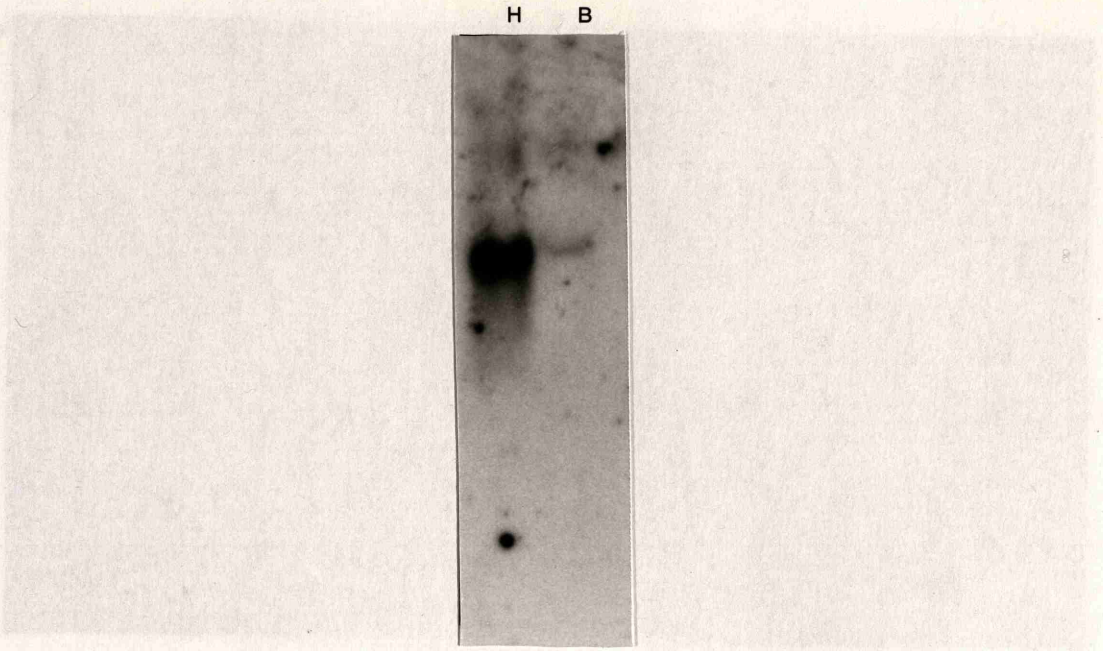


Figure 7.12 *Northern Analysis of the Transcripts Represented by the pC133 Subclone.*
10µg total RNA from heads and bodies was run on 1.5% Agarose, MOPS formaldehyde denaturing gels as described in Section 2.17. RNA was transferred to nitrocellulose and hybridised with random primed pC133 DNA. After washing the filter was autoradiographed (exposure 16 hrs at -70ºC). Expression levels in the head (h) are approximately 10-fold higher than in the body (b).

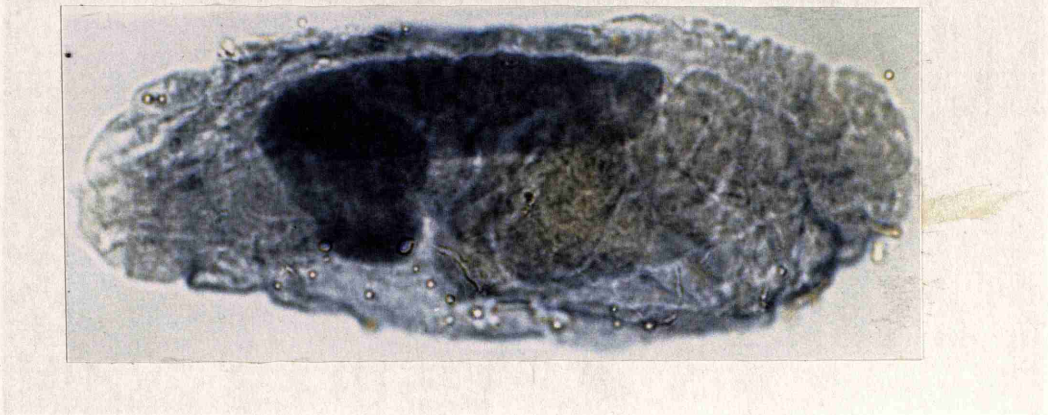


Figure 7.13 *in situ Hybridisation of the pC133 Subclone to Drosophila Whole Mount Embryos*
Embryos were collected as described in Section 2.20. A single stranded probe was made using the PCR protocol described in Section 2.20. Plasmid pC133, along with the T7 promoter was used, generating an antisense probe. Expression becomes localised after stage 14 and is largely restricted to the Central Nervous System by stage 16 as shown.

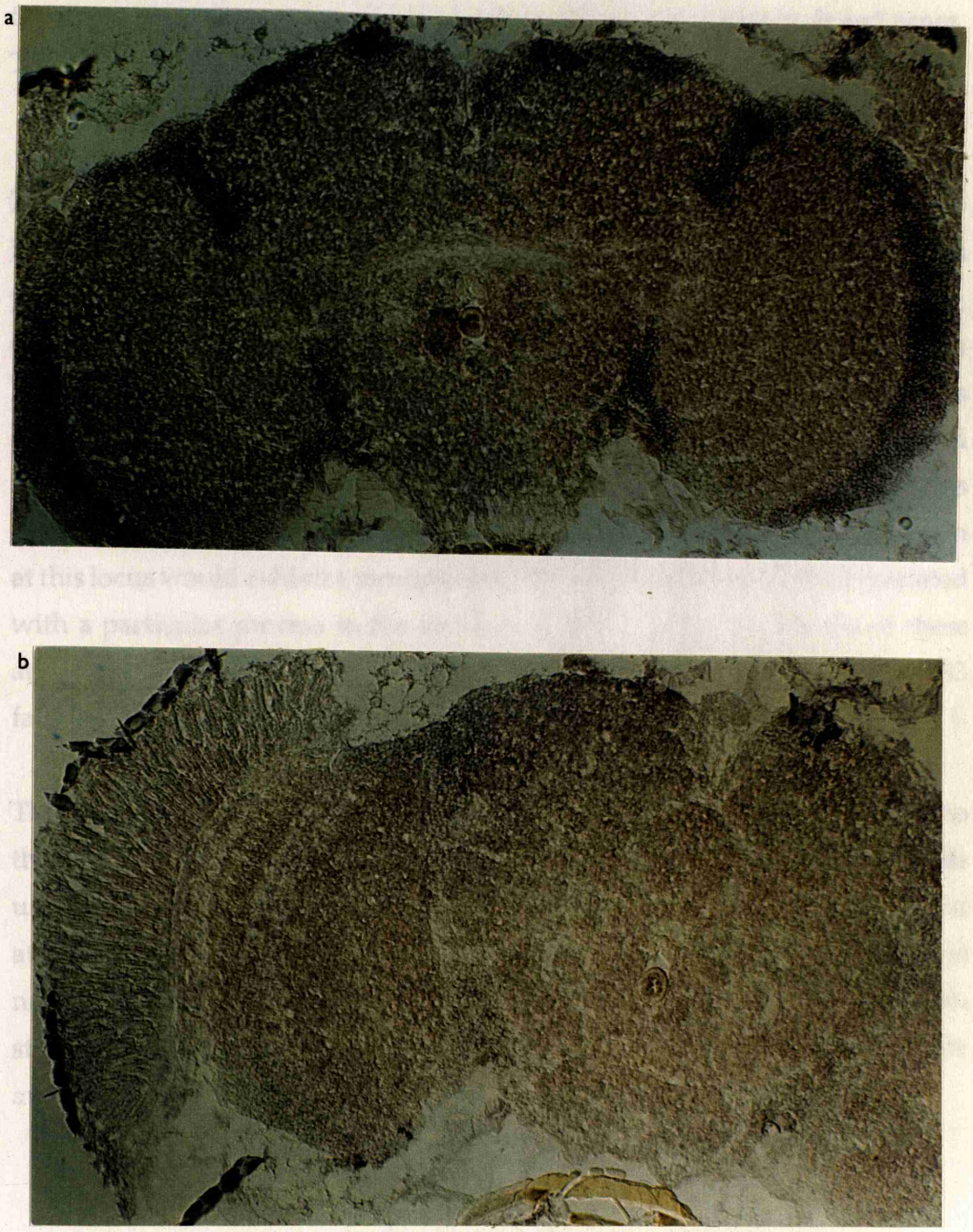


Figure 7.14 *in situ Hybridisation of the pC133 Subclone to Frozen Adult Head Sections*
A single stranded probe was made using the PCR protocol described in Section 2.20. Plasmid pC133, along with the T7 promoter was used, generating an antisense probe (a). This probe was used to hybridise to head sections as described in Section 2.21. Strong expression is seen in cell bodies throughout the brain, possibly reflecting a general nervous system expression pattern. In tandem, a sense probe was made using the T3 primer to generate DIG-labelled DNA as before. No significant specific hybridisation is seen with this probe, as expected (b).

known, it should be possible to obtain mutations at this locus in an attempt to characterise and study it further. Once full length sequence is known, a possible function may be deuced from the homology it shows to previously cloned genes. This should allow more detailed characterisation of this gene.

7.4 Discussion

For pC133, a long Open Reading Frame is now the main requirement for this project, so sequencing of the pC133 clone should be completed, and if necessary, other clones characterised. Once sequencing is complete, disruption of this locus is vital. A variety of expression studies all suggest that this clone represents a strongly expressed nervous system specific transcript, previously uncloned and not yet represented by a mutant *Drosophila* strain. Even if sequence analysis gives no clue as to the function of this gene, it would be hoped that a mutation at this locus would exhibit a recognisable phenotype which might be correlated with a particular process in the nervous system. Using the results of these analyses should help direct research on the locus represented by the pC133 family of clones.

The two clones described here highlight the need for mutants corresponding to the cDNA clones isolated in this study. Although attempts at creating mutants using 'site-selected' mutagenesis were unsuccessful in this study, a further attempt is certainly justified. In addition, the use of 'local jumping' to generate new mutations is also possible. Such a screen may be more efficient than standard SSM for the pD974 locus where many P-element insertion strains are available already.

Chapter Eight

Characterisation of Clone pD913, a Novel Esterase Gene

8.1 Introduction

The largest bacteriophage clone isolated in the secondary screen performed with pD91 was represented by the plasmid subclone pD913. From this, a 2.8kb *Eco*RI fragment cross hybridised to the original plasmid clone. Subclone pD913 does not have a 3' *Hind*III site and is therefore unlikely to possess a polyA tail at its 3' end. Other fragments from the λD913 bacteriophage were also subcloned. However once sequencing of the large *Eco*RI fragment had been initiated, it was clear that these fragments would provide no further information about the coding region of this clone. As a result, they were not characterised further. In addition to sequence analysis, Southern and Northern blot analysis and *in situ* hybridisation to both polytene chromosome spreads whole mount tissue were also carried out. The results of these experiments will be discussed in this Chapter.

8.2 Results

8.2.1 Sequence Analysis of Clone pD913

pD913 was the longest subclone to be characterised in this study. The clone contained internal sites for *Kp*NI and *Sst*I, the enzymes normally used to create 3' overhangs in the generation of Exonuclease III deletions. Production of nested deletions would have been possible using thio fill-in techniques (as described in the Promega '*Protocols and Applications Guide*') however as these techniques had not been tried in this laboratory and the reagents were not readily available, it was decided that a more conventional (albeit slower) approach should be used to obtain the sequence of this clone. In the first instance, terminal sequence was obtained using the vector primers T7 and T3. Once sequences from the fragment were available, these were used to design oligonucleotide primers to obtain more sequence. This strategy was used to design 3 primers at each end of the clone, generating long sequence contigs in each case. This sequencing strategy is shown in Figure 8.1a. In addition, a crude restriction map was obtained for the clone by digestion with enzymes which cut within the pBS SK$^-$ polylinker. The restriction map, shown in Figure 8.1b was used to decide which fragments should be subcloned. Several subclones were obtained and terminal sequencing of these was carried out in order to obtain more sequence from this locus. Although this strategy was not in itself enough to produce full length

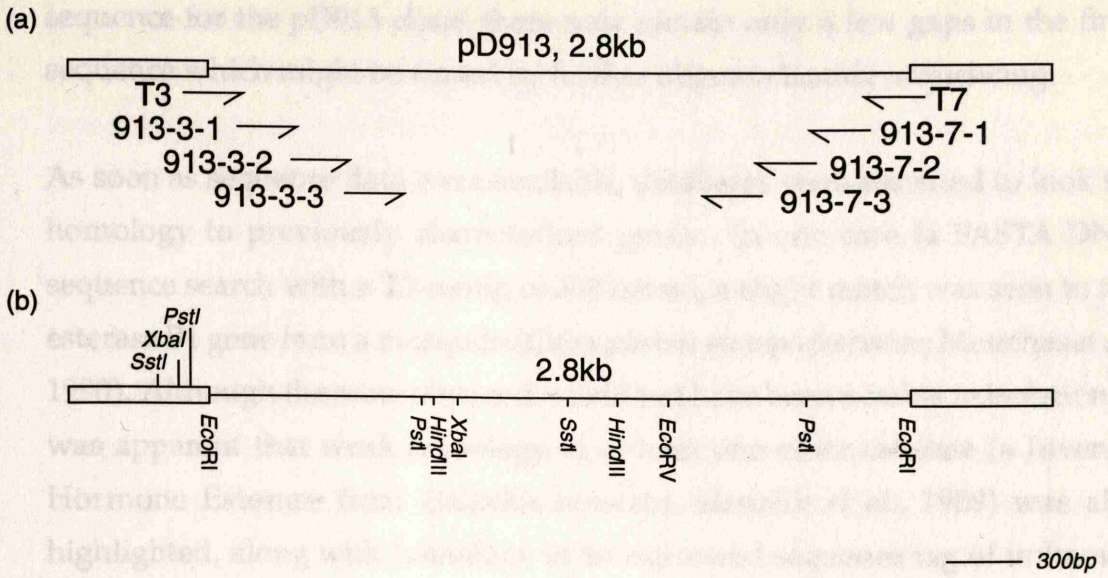


Figure 8.1 *Sequencing Strategy and Restriction Map of the Subclone pD913*
Ends of the clone were sequenced by designing primers from the end of the known sequence and in this way progressing in steps through the clone. Parts of the middle of the clone were sequenced by subcloning fragments. In this way, most of the coding region of the this gene was sequenced.

```
SCORES   62.4% identity in 133 bp overlap

         40         50         60         70         80         90
pD913    AAAGTCCAGCAGTATCGCCAGTCGACCAATGAAACAGTTGTCGCCGACACGGAGTACNGC
                                           ||||   |   ||   | ||   | |||:||
CulexB1  AAAAACTCCAATCTACGTAGGATGAGTTTGGAAAGCTTAACCGTTCAGACCAAATACGGC
            390        400        410        420        430        440

        100        110        120        130        140        150
pD913    CAAGTGAGGGG-TATCAGCGTCTATCTCTCTACGATGTGCCCTACTTCAGCTTCGAGGGT
         |  ||  ||||  | | |    ||||  |    |    |   ||| ||||||||  ||||
CulexB1  CCGGTCCGGGGCAAACGGAACGTATCGTTGCTGGGACAGGAGTACGTCAGCTTTCAGGGA
            450        460        470        480        490        500

         160        170        180        190        200        210
pD913    ATCCCGTACGCCCAGCCTCCGGTGGGGGAGTTGCGGTTTAAGGCCCCTCAGAGGCCCATT
         ||  ||||||||||| | | ||||  |||||| ||||||||||||
CulexB1  ATTCCGTACGCCCGGGCACCGGAAGGGGAGCTGCGGTTTAAGGTGAGAGTGGTAAATTGT
            510        520        530        540        550        560

         220        230        240        250        260        270
pD913    CCCTGGGAGGGAGTTCGCGACTGCAGCCAGCCGAAGATAAGGCCGTCCAGGTGCAGTTCG

CulexB1  TTCAAGTGCTGTTCAAATTTTATGGATGTGCAAGTGCATTTTTGTTCAAATAAAGAGCAA
            570        580        590        600        610        620
```

Figure 8.2 *Initial Alignment of pD913 to the Culex Esterase B1 Sequence.*
Alignment is reasonably strong over the short region shown, but stops abruptly at base 549 of the *Culex*sequence. This correponds to an intron present in the genome at this point. CulexB1; Esterase B1 from *Culex pipiens*, pD913; 5' end sequence from clone pD913

sequence for the pD913 clone, there now remain only a few gaps in the final sequence which might be closed by further oligonucleotide sequencing.

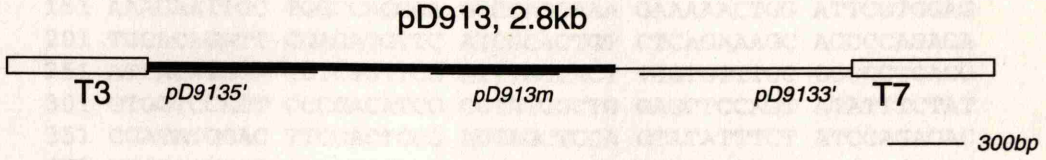
As soon as sequence data were available, databases were searched to look for homology to previously characterised genes. In one case (a FASTA DNA sequence search with a T3 contig of 503 bases), a slight match was seen to the esterase B1 gene from a mosquito (*Culex pipiens quinquefasciatus*; Mouches *et al.*, 1990). Although the score obtained would not have been notable in isolation, it was apparent that weak homology to at least one other esterase (a Juvenile Hormone Esterase from *Heliothis viriscens*, Hanzlik *et al.*, 1989) was also highlighted, along with homology to an expressed sequence tag of unknown function from *Caenorhabditis elegans*. Closer examination of the *Culex* sequence revealed good homology over a short (133 bp) stretch, followed by an abrupt halt (Figure 8.2). The *Culex* sequence providing the match was in fact a genomic sequence, and the abrupt halt in sequence similarity occured at an exon-intron boundary identified by Mouches *et al.*, (1990). Manipulation of the *Culex* sequence to remove intronic sequences allowed a stronger match to be identified between these two sequences. At this early stage therefore, it seemed clear that the pD913 clone represented a gene encoding a *Drosophila* esterase. Although several esterase genes have previously been isolated from *Drosophila*, the pD913 esterase seemed to bear more homology to the *Culex* esterase. This esterase had been originally isolated because mosquitoes resistant to Organophosphorous insecticides were found to have undergone amplification of the locus encoding this gene. Although their precise function is unknown, the products of these amplified loci are thought to be involved in the removal of Organophosphorous insecticides which would otherwise block the acetyl choline receptor. Further sequencing supported the initial findings with strong matches being found in both DNA and peptide searches.

A large number of esterase and related genes have been cloned from many different organisms. These esterases are responsible for the cleavage of esters in a variety of situations. By far the best known are the acetylcholinesterases (such as *Ace* in *Drosophila*; Hall & Spierer, 1986), which mediate the hydrolysis of acetyl choline to acetate and choline at cholinergic synapses. In addition to Acetylcholinesterases there are other cholinesterases (such as the

butyrylcholinesterases found in mammals) and esterases whose substrates are not choline related. Other groups of esterases such as the carboxylesterases, lysophospholipases and cholesterol esterases have a role in digestive processes. In *Drosophila* and other insects, a number of esterases have been isolated or defined genetically. Of the cloned ones, Esterase 6 (Oakeshott *et al.*, 1987), a male-specific protein found in adult testes is thought to be involved in hydrolysis of a female pheremone, whilst the Juvenile Hormone Esterase from the noctuid moth *Heliothis viriscens* (Hanzlik *et al.*, 1989) is thought to modulate levels of Juvenile Hormone at the initiation of metamorphosis in this insect. In addition to the B1 esterase cloned from *Culex* by Mouches *et al.* (1990), other esterases whose duplication is implicated in conferring insecticide resistance have been cloned from the peach potato aphid *Myzus persicae* (Field *et al.*, 1993) and recognised genetically in the housefly and other insects. All of these enzymes are highly related and are considered to belong to a superfamily, all derived from a single ancestor (Krejci *et al.*, 1991). All functional esterases in this family possess strongest homology around the active site which contains a serine residue. They are therefore known as serine esterases. There are a small number of molecules known which (though possessing the general structure), do not possess this active site serine and are therefore not considered to be functional esterases. Among these are the mammalian protein thyroglobulin (Merckent *et al.*, 1985) and the *Drosophila* neurotactin protein (de la Escalera *et al.*, 1990), which is thought to mediate cell-cell interactions in the *Drosophila* nervous system. Recently, the gene encoding another serine esterase-like protein has been cloned from the *Drosophila* nervous system (Auld *et al.*, 1995). The gene *gliotactin* is expressed in glia cells and is again thought to be involved in cell-cell interactions. All sequences obtained (as 3 contigs), along with predicted translation products are shown in Figure 8.3.

Figure 8.4 shows the phylogenetic relationship found between some of the genes described above and the pD913 clone. As can be seen from this phylogenetic tree, the clone characterised here seems to bear most resemblance to the other insect esterases described above, rather than to the previously described *Drosophila* esterases such as esterase-6 or acetylcholinesterase. It is interesting to note that among the sub-group containing pD913, the lengths of the branches of the tree suggests that these genes have diverged significantly from each other.

(a)



(b)

```
pD9135' fragment, 681 bases
  1 CAAGCGCTTG CGGTGGCGCC TCAAAACCAT CGAGCATAAA GTCCAGCAGT
 51 ATCGCCAGTC GACCAATGAA ACAGTTGTCG CCGACACGGA GTACGGCCAA
101 GTGAGGGGTT ATCAGCGTCT ATCTCTCTAC GATGTGCCCT ACTTCAGCTT
151 CGAGGGTATC CCGTACGCCC AGCCTCCGGT GGGGGAGTTG CGGTTTAAGG
201 CCCCTCAGAG GCCCATTCCC TGGGAGGGAG TTCGCGACTG CAGCCAGCCG
251 AAGATAAGGC CGTCCAGGTG GCAGTTCGTC TTCGATAAGG TAGAGGGCTC
301 CGAGGACTGC CTCTATCTCA ATGTGTACAC CAACAATGTG AAGCCCGACA
351 AGGCTCGCCC GGTTATGGTT TGGATTCACG GAGGAGGCTT CATTATCGGC
401 GAGGCCAATC GGGAATGGTA TGGCCCGGAT TACTTTATGA AAGAAGATGT
451 TGTTCTCGTC ACGATACAGT ACCGACTTGG TGCTTTGGGA TTTATGAGTC
501 TTAAGACCCC CGAGCTAAAT GTACCAGGAA ATGCTGGCCT CAAGGATCAG
551 GTGCTGGCCC TCAAGTGGAT CAAGAACAAT TGCGCTAGTT TCGGCGGAGA
601 TCCCAACTGC ATCACTGTTT TTGGAGAGAG TGCTGGAGGC GCCTCCACTC
651 ACTACATGAT GCTAACCGAT CAGACCCAAG G
```

(c)

```
pD9135' fragment, predicted peptide sequence 227aa (fr 2)
  1 KRLRWRLKTI EHKVQQYRQS TNETVVADTE YGQVRGYQRL SLYDVPYFSF
 51 EGIPYAQPPV GELRFKAPQR PIPWEGVRDC SQPKIRPSRW QFVFDKVEGS
101 EDCLYLNVYT NNVKPDKARP VMVWIHGGGF IIGEANREWY GPDYFMKEDV
151 VLVTIQYRLG ALGFMSLKTP ELNVPGNAGL KDQVLALKWI KNNCASFGGD
201 PNCITVFGES AGGASTHYMM LTDQTQG
```

Figure 8.3 *Sequences and Predicted Polypeptides from the pD913 Subclone*

(a) schematic representation of the sequences obtained from the pD913 subclone. The thick line represents the proposed coding region within this subclone. The relative position of the pD913 middle fragment (pD913m) was determined by aligning predicted protein sequences from this region to other esterase sequences (not shown). (b) sequence of the contig pD9135', this is a 681 base contig. (c) predicted amino acid from the pD9135' fragment. An ORF of 227aa is found (frame 2). Alignment of this sequence to known esterases is shown in Figure 8.5.

(d)

```
pD913m fragment, 509 bases
  1 CCAAAGGAGA AGATGAAGTA CCGCCTGGAG GAACTAACTC CATTCCCCAT
 51 GTTTATAGGA AACTACTTCG TACGAGTGCC TGCTGTGGTC CAGAGGAAAG
101 CTTATGCCGC AGTGCTGCAG CAGCTGGATG CTGGCACACC TTTCATTCCC
151 AAAGAATTGC TGGCCACGGA GCCCAGGAAA GAAAAACTGG ATTCGTGGAG
201 TGCACAGATT CGAGATGTTC ATCGCACTGT CTCAGAAAGC ACCCCAGAGA
251 ATTACATGGA TCTCTGTTCG ATTTACTACT TCGTGTTTCC GGCCCTGAGG
301 GTGGTCCATT CCCGACATCG CGTATGGCTG GAGCTCCAGT ATATTTCTAT
351 CGATATGGAC TTCGACTCCG AGGAGCTCCA GTATATTTCT ATCGAGAGAC
401 TTCGACTCCG AGGAGCTCCA GCCGCATACG CGTGTCGGGA ATGGACCACC
451 CTCAGGGCCG GAAACATCGA AGTAGTAAAT CGAACAGAGA TCCATGTAAT
501 TATCTGGGG
```

(e)

```
pD913m fragment, predicted peptide sequence 166aa (fr 1)
  1 PKEKMKYRLE ELTPFPMFIG NYFVRVPAVV QRKAYAAVLQ QLDAGTPFIPK
 51 ELLATEPRKE KLDSWSAQIR DVHRTVSEST PENYMDLCSI YYFVFPALRVV
101 HSRHRVWLEL QYISIDMDFD SEELQYISIE RLRLRGAPAA YACREWTTLRA
151 GNIEVVNRTE IHVIIW
```

(f)

```
pD9133', 818 bases
   1 TCTCGCTCTT TCCCTCTGGC TTTCTTTTCA AATGTATTTT GTATGAATCT
  51 CAAATGTTTT TCCACTTGCT TGCCTATTTT ATTTGCCATT CACAAACAAT
 101 ATTACAATTT GCCATGGATT GGTATCATGT TTCTTTTTAC TGTTTTCTAT
 151 GTTTGTTTAC AATTTGTCAC AACCTTACAA ATTGAAAAAT TGAAATAAAC
 201 CAATAAGTTT GTGGTTATTG CTGTTAGTTG GGTTTTGTTC TGCGATTGTT
 251 GCTAATATAT AATACTTGTT TATCGTCTCT GCGGTTCTAT TATCAATTTT
 301 GTTTTGTAAG GTAGCAGAGA AACGCACGGA ATTGGCCAAT TTTACCCTTG
 351 CCCGTTCCTA TACACAGTGC TAAATTTAAA AAATATGTTC TTCATTATCA
 401 TCTAAATGTT TAAAAATCAA AATATTATTC TTGAATATTT CCTTTTTACG
 451 GGAGGGGAAA ACTGTAATGC TATATTCAAT GTTTTTTGTT TTATATTTTA
 501 ATTCATATTT TATTTCAATT TGTGTTTTGC ATTTGCTCAT AAAAACAATG
 551 GTACAACTTA CAATTTGGCC ACATCTTAAA GTTAGAGTTT CAATATAATA
 601 TATATTGATA TACTTTATAT TTTATACTTA TGACGCCTAG GCATCTGATC
 651 TATGTGAAAG TAAAACAAAA CTGGCGCTGG CTTATGGTTA GACTACACAG
 701 ATTACAGAAT TTTTGGGGGT CACAAAATGG ATCGGAATAA ATAAGACATT
 751 CATCATATAC ATCATATATC GTATGTNATA TCATATAGCT GCGGTGTTTG
 801 ATTGAGTTGC GTGCTTG
```

Figure 8.3 *Sequences and Predicted Polypeptides from the pD913 Subclone (continued)*
(d) sequence of the contig pD913m, this is a 509 base contig. (e) predicted aa sequence from the contig pD913m, an ORF of 166aa is found (frame 1). (f) sequence of the pD9133' contig, this is a 818 base contig and is not thought to encode any peptide.

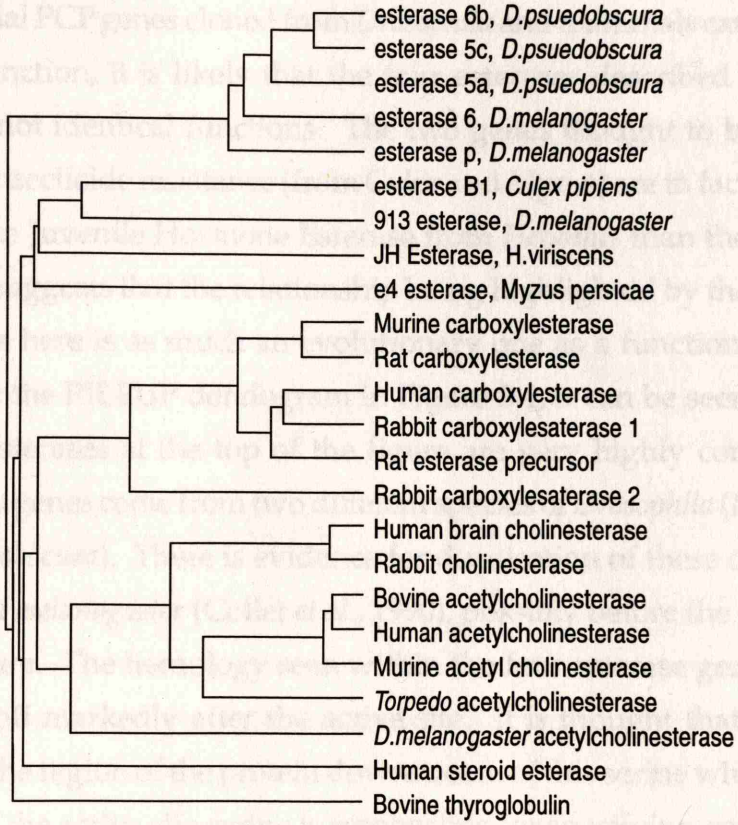


Figure 8.4 *Phylogenetic Relationship Between pD913 and Other Esterases.* PILEUP alignment between pD913 and 23 other esterases from *Drosophila* as well as other insects and mammals. Note that pD913 shows more homology to the esterases from *Myzus*, *Culex* and *Heliothis* than it does to other esterases from *Drosophila*. However it should also be pointed out that within this sub-group, conservation is less pronounced than in any of the other sub-groups suggested by this dendritic tree.

If the sequences themselves are lined up (as seen in Figure 8.5, using the PILEUP program from the GCG suite), it is clear that the pD913 esterase bears most homology to the *Culex*, *Myzus* and *Heliothis* esterases described above. Still however, these esterases are relatively poorly conserved at the protein level. They share approximately 42% identity (58% similarity) at the amino acid level, well below the 75% identity seen between the *Drosophila* and mammalian mitochondrial PCPs discussed in Chapter Six. Whilst it is the case that the mitochondrial PCP genes cloned from *Drosophila* and mammals carry out exactly the same function, it is likely that the four esterases described here perform similar but not identical functions. The two genes thought to be involved in conferring insecticide resistance (from *Culex* and *Myzus*) are in fact more closely related to the Juvenile Hormone Esterase from *Heliothis* than they are to each other. This suggests that the relationship being highlighted by the polypeptide comparisons here is as much an evolutionary one as a functional one. If we look back at the PILEUP dendogram in Figure 8.4, it can be seen that the five *Drosophila* esterases at the top of the figure are very highly conserved, even though these genes come from two different species of *Drosophila* (*D.melanogaster* and *D.pseudobscura*). There is evidence for duplication of these esterase genes in *Drosophila melanogaster* (Collet *et al.*, 1990), possibly before the divergence of *D.pseudobscura*. The homology seen within the four esterase genes compared here drops off markedly after the active site. It is thought that specificity is encoded in the region of the protein downstream of this serine whilst the region upstream of the active site serine is responsible for specifying general esterase activity.

From the PILEUP analysis of the pD913 sequence and those of the other three insect esterases, (Figure 8.5), it is apparent that the pD913 clone is not full length. It can be predicted from the lengths of the other three polypeptides shown here that this peptide is perhaps just one or two amino acids short of being full length. As discussed in Chapter Five, the pD913 cDNA clone represents a rare transcript (present at levels of less than 1 in 50,000 within the head library screened). In addition, the very long transcript size (at least 2.8kb) indicates that full length clones might be very rare within cDNA libraries. For this reason it was decided to screen a genomic DNA library to obtain a fragment containing the sequences at the extreme 5' end of this gene. 30,000 plaques from a wild

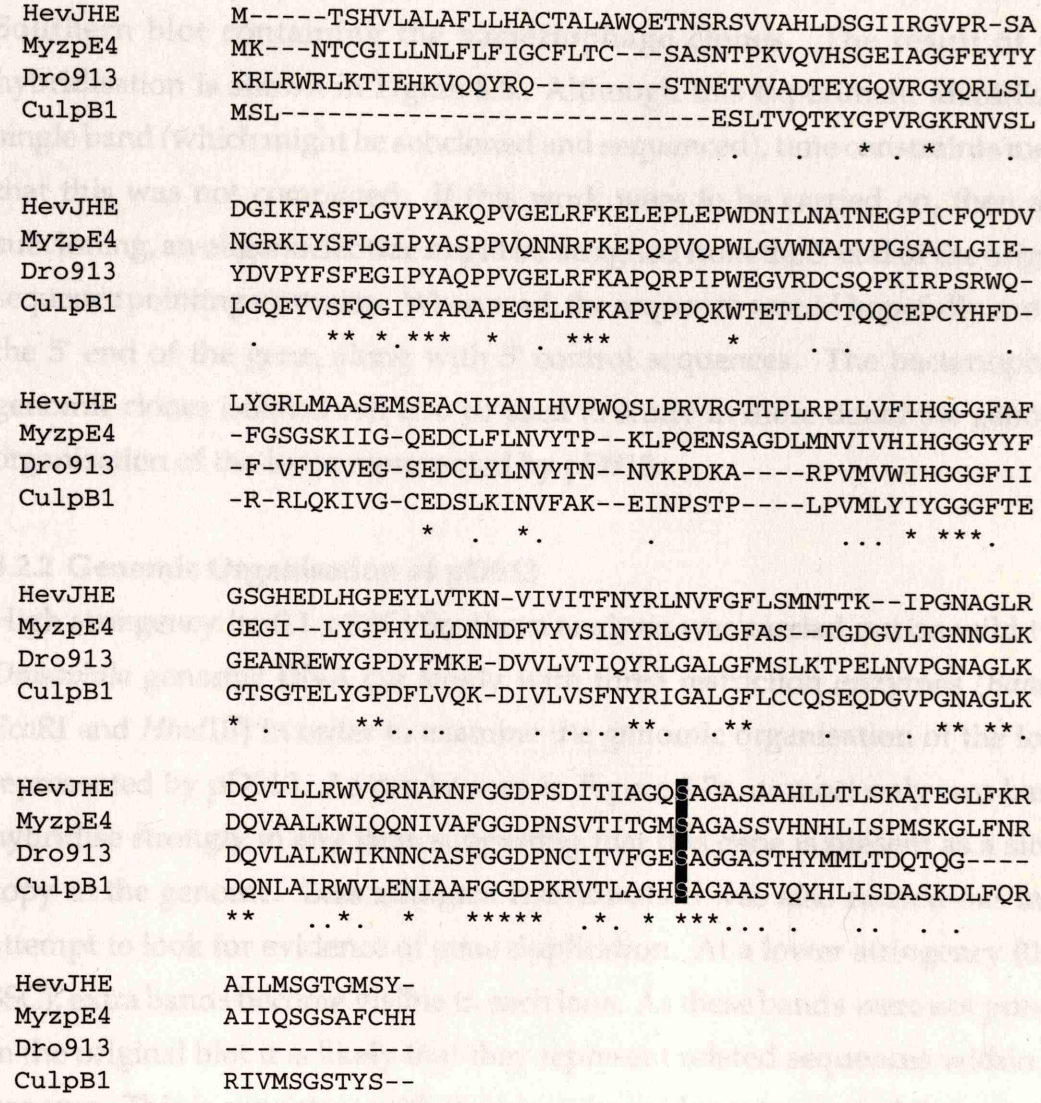


Figure 8.5 *Clustal Alignment of pD913 and the Other Esterases*

The sequence obtained from the 5' end of the pD913 subclone was translated and aligned to other related esterase sequences cloned from insects. Significant homology is seen up to and including the active site Serine residue (boxed). Homology falls off in the next region and it is impossible to reliably align the polypeptide predicted from the sequence of the middle fragment presented in Figure 8.3.

type *Drosophila* genomic library (made by Simon Tomlinson in the vector λgem11; see Chapter 2) were screened with the pD913 subclone. Three positives were obtained from which DNA was prepared. The three bacteriophage clones were clearly related (as judged by their common restriction patterns). To identify the fragment most likely to contain the 5' end of the gene, the oligonucleotide 913-3-1 (for position see Figure 8.3) was radiolabelled and used to screen a Southern blot containing the bacteriophage clones. The result of this hybridisation is shown in Figure 8.6. Although this experiment identified a single band (which might be subcloned and sequenced), time constraints meant that this was not completed. If this work were to be carried on, then after subcloning, an oligonucleotide might be designed from the 5' end of the original sequence pointing upstream. When read, the sequence would hopefully contain the 5' end of the gene, along with 5' control sequences. The bacteriophage genomic clones isolated can also be used to study in more detail the genomic organisation of the locus represented by pD913.

8.2.2 Genomic Organisation of pD913

High stringency (to 0.1 x SSC) Southern analysis was carried out on wild type *Drosophila* genomic DNA cut singly with three restriction enzymes (*Bam*HI, *Eco*RI and *Hind*III) in order to examine the genomic organisation of the locus represented by pD913. As can be seen in Figure 8.7a, at most only two bands hybridise strongly in any lane, suggesting that this gene is present as a single copy in the genome. Less stringent hybridisation was also carried out in an attempt to look for evidence of gene duplication. At a lower stringency (0.5 x SSC), extra bands become visible in each lane. As these bands were not present in the original blot it is likely that they represent related sequences within the genome. This is consistent with the previous evidence presented that esterase genes are often duplicated. These duplications are normally local. The lower stringency hybridisation experiments also gave some conflicting results and would have to be repeated before concrete conclusions may be drawn.

8.2.3 *in situ* Hybridisation to Polytene Chromosomes

The physical location of the pD913 subclone was determined by *in situ* hybridisation to polytene chromosomes. Figure 8.8 shows spread wild type chromosomes hybridised with clone pD913 DNA. Hybridisation can be seen

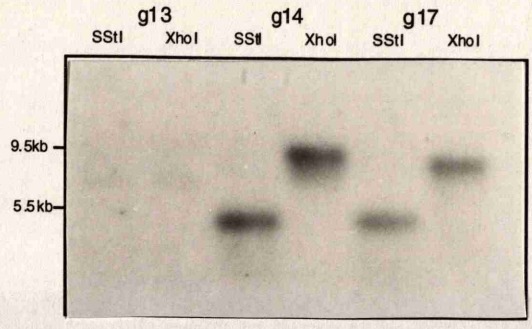


Figure 8.6 *Oligo Hybridisation of 913-3-3 to the λD913 Genomic Clones*
5µg of three genomic clones λD913g13, λD913g14 and λD913g17 were digested singly with the restriction enzymes *Sst*I and *Xho*I and run on a gel. This gel was blotted and hybridised with end-labelled 913-3-1 oligonucleotide at 37ºC, as described in Section 2.18. Filters were washed and autoradiographed. 2 day exposure. The oligonucleotide hybridises to a 5.5kb *Sst*I fragment and to a 9.5kb *Xho*I fragment present in clones λD913g14 and λD913g17. No hybridisation is seen to sequences present in λD913g13; presumably this clone does not contain the 5' end of the gene.

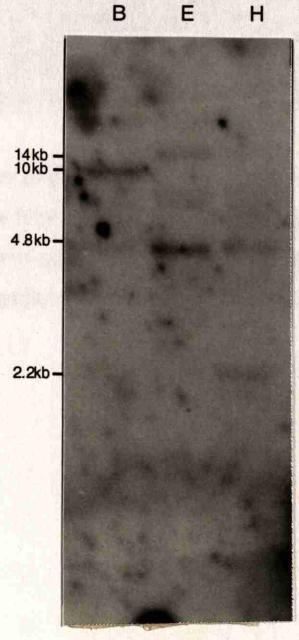


Figure 8.7 *Southern Analysis of the Region Represented by the pD913 Subclone.*
4µg *Drosophila* genomic DNA was digested with the three restriction enzymes *Bam*HI, *Eco*RI and *Hind*III. Digested DNA was separated on 0.8% TBE agarose gels, stained, photographed and transferred to nylon membrane for hybridisation with pD913 DNA. After hybridisation, filters were washed as described in Section 2.17 and autoradiographed for 72 hours. B; *Bam*HI; E, *Eco*RI; H, *Hind*III. A single band of >10kb was seen in the *Bam*HI digest, whilst the *Eco*RI digest generated a two hybridising bands of ~14kb and 4.7kb. In the *Hind*III digest, two bands of 4.8 and 2.2 kb are identified.

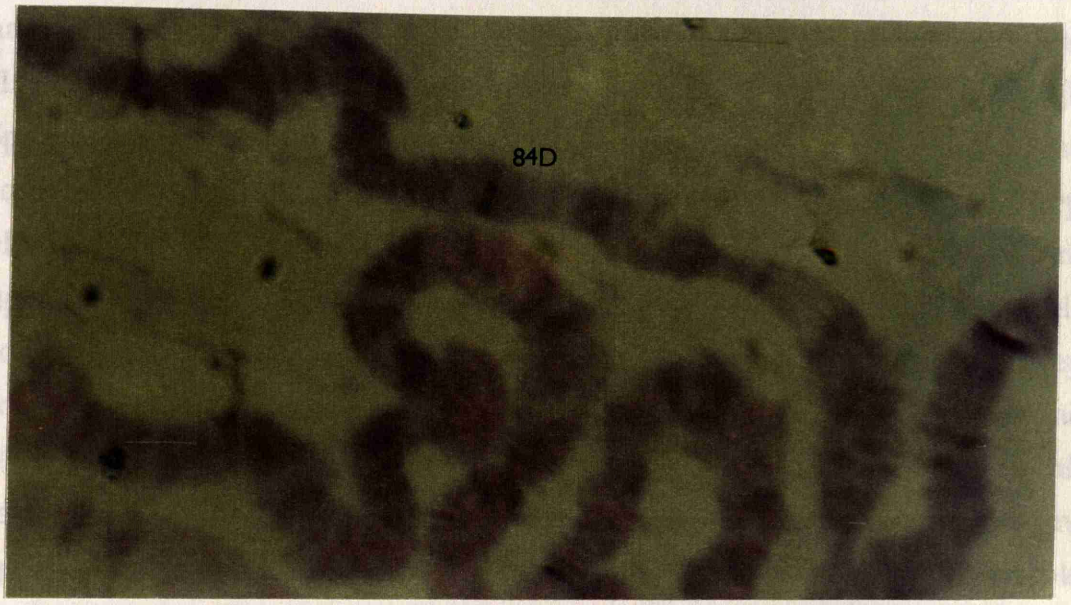


Figure 8.8 *in situ Hybridisation of pD913 to Spread Polytene Chromosomes.*
Spread polytene chromosomes from wild type flies were hybridised with plasmid DNA from clone pD913. Hybridisation is seen to band 84D, the predicted location of the *Drosophila* pD913 locus. This hybridisation was kindly carried out by Zong Sheng Wang.

at polytene band 84D on the right arm of the third chromosome. There is no evidence for hybridisation at any other position in the polytene chromosomes. Therefore, if this gene has undergone a duplication, then it must have been a local duplication. Again, this region of the *Drosophila* genome has been well characterised and there are a number of genes and lethal loci which are known to map to this location. Most importantly, a number of esterase genes are known to map at this site. Esterase 9 (also known as Esterase C; Cavener *et al.*, 1986) has long been known to map to this location. In addition, Spackman *et al.* (1994), finely mapped three esterase loci to the region 84D3-84E2. Although Spackman *et al.* infer that one of these genes (est-23) is homologous to a gene conferring insecticide resistance from *Lucilia cuprina* (the Australian sheep blowfly), they have no molecular data from either gene and are dependent upon phenotypic observations and genetic characterisation to make their conclusions. Their phenotypic observations would suggest that the gene they concentrated their work on (esterase 23) is not the gene represented by the pD913 subclone characterised here. The other three genes are however all reasonable candidates. *ali*-esterase is thought to encode an esterase with a non-specific activity. MCE (malathion carboxylesterase is the enzyme responsible for breakdown of the insecticide malathion. Esterase-9 (est-C) is the major *Drosophila* carboxylesterase. Further molecular analysis of these mutants would be necessary to elucidate which gene is represented by the pD913 subclone.

As with the other subclones, 'site-selected' mutagenesis was attempted without success. However in this case as there is reason to believe that there are mutations (or at least recognisable biochemical variants) in existence for the esterase genes at this locus, it may be best to use these strains as a starting point for further investigation. The deficiency mapping carried out by Spackman *et al.* (1994) identified deletion strains covering these loci. These could be used, along with our knowledge of the genomic organisation of pD913 to correlate the pD913 subclone with one particular esterase gene. The presence of four esterase genes at this site is perhaps evidence for a recent gene duplication in this region of the chromosome. Perhaps this might correlate with the presence of extra bands seen on Southern blots under conditions of reduced stringency.

8.2.4 Transcriptional Analysis of Clone pD913

As biochemical variants are available, it might be possible to correlate the protein distribution of a particular esterase at 84D to the distribution of the transcript represented by the subclone pD913. Analysis of the transcription of clone pD913 on Northern blots, or by *in situ* hybridisation might be especially informative here. Unfortunately, Northern analysis using the pD913 subclone as probe (though attempted a number of times) was never successful. This might be due in part to the rarity of the transcript (Northern analysis was never attempted on good quality poly A^+ RNA). Similarly, although *in situ* hybridisation to whole mount embryos and adult head Sections was attempted, no positive hybridisation was obtained.

8.3 Conclusions

Although it is clear that the locus represented by the pD913 subclone encodes an esterase, its specific function is as yet unclear. The significant biochemical studies already performed on *Drosophila* and the possibility that a biochemical mutant already exists for this gene make an *in vivo* approach to further dissection of this gene attractive. Genetic dissection of this locus using the newly generated molecular marker will help elucidate the organisation of esterase genes at 84D. This will facilitate further analysis of the evolution of this cluster and allow comparison with the other genes found in the *Drosophila* species group. Biochemical analysis (for example with reintroduction of a fusion gene construct) may be used to investigate the function of this gene product.

Chapter Nine

Conclusions

9.1 Introduction

The screening approach outlined in this thesis efficiently isolated a large number of genes expressed in the head of *Drosophila melanogaster.* Although time, and the possible presence of artefacts meant that this study examined only four genes, preliminary information gathered here would allow others to start work on these other clones, or to use the same screening strategy to screen another library where artefacts were not present.

Of the four clones chosen for further study, two remain unidentified, and a third represents a gene cloned in a number of other other organisms though not, as yet, in *Drosophila.* Only one (pD913) is thought likely to represent a variant previously known in *Drosophila melanogaster* (one of the cluster of esterase genes on the right arm of chromosome 3). Of the four genes, the pD913 esterase might seem the most interesting and of course it is exactly the sort of gene expected to be generated in this study. Both this esterase, and the Phosphate Carrier Protein (described in Chapter Six) were recognised initially by DNA sequence similarities to genes previously cloned and characterised in other organisms. The number of sequences in the GenBank databases is growing quickly and will continue to do so for the foreseeable future. Therefore, in the short term it should become even more fruitful to carry out sequence analysis as an aid to finding the function of a gene. By way of a contrast, the two subclones described in Chapter Seven bore no apparent similarity to any previously cloned genes, and their function remains unknown.

With the exception of pD682, the genes isolated in this study seem to have a head enriched rather than head specific pattern of transcription. This might suggest a role for these genes in the nervous system as a whole. Expression profiles for pC133 and pD974 certainly suggest a general neuronal pattern of expression. In the case of pD913, no expression data (apart from the crude expression profile deduced from the Reverse Northern analysis) are available for this clone. Esterases are found in all parts of the body, though individual esterases seem to have restricted patterns of expression (to the digestive midgut, the male reproductive tract, or indeed to the nervous system itself). Pressure of time has left much of the analysis initiated in this study incomplete, although work has been advanced far enough for each clone to be considered on its own

merit as the basis for future study. Ignoring the difficulties introduced by the presence of artefacts it is clear that the approach used in this work was efficient in its goal of finding new genes expressed in the *Drosophila* head.

It is encouraging that none of the original clones seemed to have been cloned previously; this vindicates the use of the novel approach presented here. The more conventional approach, originally used by Simon Tomlinson was successful only in re-identifying an already well characterised eye specific gene.

Of the four clones chosen for further analysis, three seem to have neuronal roles or expression patterns. Only one of the four loci studied seems to have been characterised genetically prior to this study, again suggesting that this approach may be successful in identifying genes missed by more conventional approaches.

9.2 Further Work

Time constraints have limited the experiments carried out on the clones isolated in this study to a relatively basic characterisation of each cDNA subclone. It seems clear however, that were these studies to be continued, it would be vital to examine the disruption of these loci and their function *in vivo*. The first step should be to find or attempt to create mutations at these loci. As described in Chapter Three, and attempted for all four subclones, 'site-selected' mutagenesis provides a theoretically efficient means of introducing new targeted mutations into the *Drosophila* genome. This method is particularly suited to mutagenesis of genes for which sequence may be the main source of information about a gene. In addition, as highlighted in Chapter One, the generation of new single P-element insertion lines creates a vast library of randomly created insertions which can be catalogued by *in situ* localisation of the P-element sequence. Even the relatively small figure of 10,000 independent P-element lines is sufficient to give one P-element insertion every 16kb throughout the genome. Furthermore, remobilisation of a single P-element may lead to the generation of many new insertions in the immediate vicinity of the locus under study due to the phenomenon of 'local jumping'.

Mutants would be useful for a number of reasons. First, mutations may be characterised at the phenotypic level and the collected observations formulated into testable hypotheses of the function of the gene under investigation. Second, mutant fly lines may be rescued by P-element transformation with the wild type cDNA fragment. Rescue of a mutant phenotype is perhaps the most potent evidence available that a cDNA fragment represents the gene under investigation. Third, once rescue of a mutant phenotype has been demonstrated, P-element transformation can be used to dissect the structure-function relationship of the gene. Transformation may be used to introduce truncated or augmented constructs derived from this gene to examine their function *in vivo*. Fourth, many of the P-element mutageneses referred to above have used engineered P-elements vectors such as P-*lacZ* or P-*Gal4* elements. The presence of an engineered element at the locus under study allows the same types of analysis one would carry out for a normal enhancer trap line. With a P-*Gal4* element under the control of the gene under study, expression of *Gal4* should mimic (temporally and spatially) the distribution of that gene. Any *Gal4* responsive construct may then be introduced, allowing the expression of almost any gene construct in a pattern mimicking that of the gene under study.

9.3 Current Technology

At the moment, and for the forseeable future, P-element based tools represent the most powerful tools available for studying *Drosophila* genes *in vivo*. New insertion lines are continually being generated, meaning that the main difficulty associated with this type of analysis (getting the P-element there in the first place) is no longer a significant barrier. In addition, the switch to P-*Gal4* elements now means that only one P-element insertion is needed as this opens the door to analyses with any one of a myriad of reporter constructs which (being under the control of the diffusible *Gal4* gene product) can be situated at remote sites in the genome, rather than in the immediate vicinity of the gene as before.

A side effect of the large number of researchers currently using P-element based technology in *Drosophila* is that new tools are constantly being developed and becoming available. This increases the usefulness of P-element insertions as

markers for genes and encourages more researchers to use the technology. The P-element has become an indespensable tool.

Likewise, the genome mapping and sequencing projects initiated in a large number of organisms in the last few years can only make in depth sequence analysis increasingly vital to the study of any new gene. Databases are growing in size very quickly. In tandem the advent of automated sequencing now means that the non-coding sequences surrounding coding regions are often sequenced as part of a sequencing 'project'. Soon, analysis of such non-coding sequences might be as fruitful as analysis of coding regions, providing information about evolution and evolutionary pressures, as well as transcriptional control. It is fortuitous that advances in the rate at which sequence data is generated has been mirrored by advances in the power and speed of computers needed to analyse it. Careful structuring of databases means that the information generated is efficiently disseminated within the *Drosophila* research community.

9.4 Future Prospects

As discussed before, to fully understand an organism, we must understand it at an organismal level, looking not at how individual genes function, but at how groups of genes interact within cells. We must examine an animals interaction with its environment, asking how cells respond to stimuli and how that response is communicated throughout the animal. As more genes are isolated and characterised, the information generated becomes more easily synthesised into the context of the organism as a whole.

Bibliography

Abrams, T.H., and Kandel, E.R., (1988)
Is contiguity detection in classical conditioning a system or cellular property: learning in *Aplysia* suggests a possible molecular site.
Trends in Neurosciences *11*, 128-135.

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

Artavanis-Tsakonas, S., (1988)
The molecular biology of the *Notch* locus and the fine tuning of differentiation in *Drosophila.*
Trends in Genetics *4*, 95-100.

Ashburner, M., (1989)
Drosophila. A Laboratory Manual (1st edition).
Cold Spring Harbor Laboratory Press.

Ashley, C.T., Pendleton, C.G., Jennings, W.W., Saxena, A., and Glover, C.V., (1989)
Isolation and sequencing of cDNA clones encoding *Drosophila* chromosomal protein D1. A repeating motif in proteins which recognize DNA.
J. Biol. Chem. *264*, 8394-8401.

Auld, V.J., Fetter, R.D., Broadie, K., and Goodman, C.S., (1995)
Gliotactin, a novel transmembrane protein on peripheral glia, is required to form the blood-nerve barrier in *Drosophila.*
Cell *81*, 757-767.

Baker, B.S., Nagoshi, R.N., and Belote, J.M., (1987)
Molecular genetic aspects of sex determination in *Drosophila.*
BioEssays *6*, 66-70.

Ballinger, D. G., and Benzer, S., (1989)
Targeted gene mutations in *Drosophila*.
Proc. Natl. Acad. Sci. (USA). *86*, 4055-4059.

Banerjee, U., Renfranz, P.J., Pollock, R.A., and Benzer, S., (1987)
Molecular characterisation of *sevenless*, a gene involved in neuronal pattern formation in the *Drosophila* eye.
Cell *49*, 281-291.

Beissmann, H., (1981)
Changes in the abundance of polyadenylated mRNAs in the development of *Drosophila melanogaster* analysed with cloned DNA fragments.
Dev. Biol. *83*, 62-68.

Bellen H.J., and Kiger, J.A., (1987)
Sexual hyperactivity and reduced longevity of *dunce* females of *Drosophila melanogaster.*
Genetics *115*, 153-160.

Benzer, S., (1967)
Behavioural mutants of *Drosophila melanogaster* isolated by countercurrent distribution.
Proc. Natl. Acad. Sci. (USA) *58*, 1112-1119.

Benzer, S., (1971)
From gene to behaviour.
J. Am. Med. Assoc. *218*, 1015-1022.

Benzer, S., (1973)
Genetic dissection of behaviour.
Sci. Am. *229*, 24-37.

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

Berg C. A., and Spradling, A.C., (1989)
New oogenesis mutants generated by controlled P-element insertional mutagenesis.
Amer. Conf. *Dros.* Research *30*:11.

Bier, E., Vaessin, H., Sheperd, S., Lee, K., McCall, K., Barbel, S., Ackerman, L., Carretto, R., Uemara, T., Grell, E., Jan, L.Y., and Jan, Y.N., (1989)
Searching for pattern and mutation in the *Drosophila* genome with a P-*lacZ* vector.
Genes Dev. *3*, 1273-1287.

Birnstiel, M.L., Busslinger, M., and Strub, K., (1985)
Transcription termination and 3' processing: the end is in site!
Cell *41*, 349-359.

Blackman,R.K., and Gelbart, W.H., (1989)
The transposable element *hobo* of *Drosophila melanogaster.*
in Mobile DNA (ed Howe, M.M.) ASM Press, Washington D.C.
pp523-529.

Bloomquist, B.T., Shortridge, R.D., Schneuwly, S., Perdew, M.H., Montell, C., Stellar, H., Rubin, G., and Pak, W.L., (1988)
Isolation of a putative Phospholipase C gene from *Drosophila norpA*, and its role in phototransduction.
Cell *54*, 723-733.

Bonini, N.M., Leiserson, W.M., and Benzer, S., (1993)
The *eyes absent* gene: Genetic control of cell survival and differentiation in the developing *Drosophila* eye.
Cell *72*, 379-395.

Boynton, S., and Tully, T., (1992)
latheo: a new gene involved in associative learning and memory in *Drosophila melanogaster*, identified from P-element mutagenesis.
Genetics *131*, 655-672.

Brand, A.H., and Perrimon, N., (1993)
Targeted gene expression as a means of altering cell fate and generating dominant phenotypes.
Development *118*, 401-415.

Bucheton, A., (1979)
Non-mendelian female sterility in *Drosophila melanogaster:* influence of ageing and thermic treatments. III Cumulative effects induced by these factors.
Genetics *93*, 131-142.

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

Bullock, W. O., Fernandez, J. M., and Short, J. M., (1987)
XL1-Blue: A high efficiency plasmid transforming *recA Escherichia coli* strain with β-galactosidase selection.
Biotechniques *5*, 376-379.

Burtis, K.C., and Baker, B.S., (1989)
The *Drosophila doublesex* gene controls somatic sexual differentiation by producing alternative spliced mRNA's encoding related sex-specific polypeptides.
Cell *56*, 997-1010.

Busseau, I., Pelisson, A., and Bucheton, A., (1989)
I elements of *Drosophila melanogaster* generate specific chromosomal rearrangements during transposition.
Mol. Gen. Genet. *218*, 222-228.

Castrillon, D.H., Gonczy, P., Alexander, S., Rawson, R., Eberhart, C.G., Viswanathan, S., DiNardo, S., and Wasserman, S. A., (1993)
Toward a molecular genetic analysis of spermatogenesis in *Drosophila melanogaster:* characterization of male-sterile mutants generated by single P-element mutagenesis.
Genetics *135*, 489-505.

Cavener, D.R., Otteson, D.C., and Kaufman, T.C., (1986)
Rehabilitation of the genetic map of the 84B-D region in *Drosophila melanogaster*.
Genetics *114*, 111-123.

Cavener, D.R., and Ray, S.C., (1991)
Eukaryotic start and stop translation sites.
Nucl. Acids. Res. *19*: 3185-3192.

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

Chang, P.L., Mueller, O.T. , Lafrenie, R.M., Varey, P.A., Rosa, N.E., Davidson, R.G., Henry, W.M., and Shows, T.B., (1990)
The human arylsulfatase-C isoenzymes: two distinct genes that escape from X inactivation.
Am. J. Hum. Genet. *46*, 729-737.

Chen, C. N., Denome, S., and Davis, R. L., (1986)
Molecular analysis of cDNA clones and the corresponding genomic coding sequences of the *Drosophila dunce*$^{+}$ *gene*, the structural gene for cAMP phosphodiesterase.
Proc. Natl. Acad. Sci. (USA). *83*, 9313-9317.

Chisholm, D., (1989)
A convenient moderate scale procedure for obtaining DNA from bacteriophage lambda.
Biotechniques *7*, 21-27.

Chomczynski, P., and Sacchi, N., (1987)
Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction.
Anal. Biochem. *162*, 156-159.

Clark, A.G., Silveria, S., Meyers, W., and Langley, C.H., (1994)
Nature screen: an efficient method for screening natural populations of *Drosophila* for targeted P-element insertions.
Proc. Natl. Acad. Sci. (USA) *91*, 719-722.

Clark, R.F., and Elgin, S.C., (1992)
Heterochromatin Protein-1, a known suppressor of position effect variegation is highly conserved in *Drosophila*.
Nucl. Acids Res. *20*, 6067-6074.

Collet, C., Nielsen, K.M., Russell, R.J., Karl, M., Oakeshott, J.G., and Richmond, R.C., (1990)
Molecular analysis of duplicated esterase genes in *Drosophila melanogaster*.
Mol. Biol. Evol. *7*, 9-28.

Cooley, L., Kelley, R., and Spradling, A.C. , (1988)
Insertional mutagenesis of the *Drosophila* genome with single P-elements.
Science *239*, 1121-1128.

Cozens, A.l., Runswick, M.J., and Walker, J.E., (1989)
DNA sequences of two expressed nuclear genes for human mitochondrial ADP/ATP translocase.
J. Mol. Biol *206*: 261-280.

Crozatier, M., Vaury, C., Busseau, I., Pelisson, A., and Bucheton, A., (1988)
Structure and genomic organisation of I elements involved in I-R hybrid dysgenesis in *Drosophila melanogaster.*
Nucl. Acids Res *16*, 9199-9213.

D'Alessio, J.M., and Gerard, G.F., (1988)
Second strand cDNA synthesis with *E. coli* DNA polymerase I and RNAse H: the fate of information at the mRNA 5' terminus and the effect of *E.coli* DNA ligase.
Nucl. Acids Res. *16*, 1999-2014.

Davis, R.L., (1993)
Mushroom bodies and *Drosophila* learning.
Neuron *11*, 1-14.

de la Escalera, S., Bockamp, E-O., Moya, F., Piovant, M., and Jimenez, F., (1990)
Characterisation and gene cloning of neurotactin, a *Drosophila* transmembrane protein related to cholinesterases.
EMBO. J., *9*, 3593-3601

Demerec, M (ed.) (1950)
The Biology of *Drosophila*.
Wiley, New York.

Devereux, J., Haeberli, P., and Smithies,O., (1984)
A Comprehensive set of sequence Analysis programs for the VAX.
Nucl. Acids Res. *12*, 387-395.

diBenedetto, A.J., Lakich, D.M., Kruger, W.D., Belote, J.M., Baker, B.S., and Wolfner, M.F., (1987)
Sequences expressed sex-specifically in *Drosophila melanogaster* adults.
Dev. Biol. *119*, 242-251.

Dolce, V., Fiermonte, G., Messina, A., and Palmieri, F., (1991)
Nucleotide sequence of human heart cDNA encoding the mitochondrial phosphate carrier.
DNA Sequence *2*,133-135.

Dorssers, L., and Postmes, A.M.E.A., (1987)
A simplified orientation-specific cDNA cloning strategy.
Nucl. Acids Res. *16*, 3629.

Dudai, Y., Jan, Y.-N., Byers, D., Quinn, W. G., and Benzer, S. (1976)
dunce, a mutant of *Drosophila* deficient in learning.
Proc. Natl. Acad. Sci. (USA) *73*, 1684-1688.

Dura, J.M., Preat, T., and Tully, T., (1990)
Identification of *linotte*, a new gene affecting learning and memory in *Drosophila melanogaster*.
J. Neurogenet. *9*, 1-14.

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

Fargnoli, J., Holbrook, N.J., and Fornace Jr, A.J., (1990)
Low-ratio hybridisation subtraction.
Anal. Biochem. *187*, 364-373.

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

Feany, M.B., and Quinn, W.G., (1995)
A neuropeptide gene defined by the *Drosophila* memory mutant *amnesiac*.
Science *268*, 869-873.

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

Ferreira, G.C., Pratt, R.D., and Pederson, P,L., (1989)
Energy linked anion transport.
J. Biol. Chem *264*, 15628-15633.

Field, L.M., Williamson, M.S., and Devonshire, A.L., (1993)
Cloning and analysis of the esterase genes conferring insecticide resistance on the peach potato aphid, *Myzus persicae* (Sulzer).
Biochem. J. *294*, 569-574.

Finnegan, D.J., (1989)
The I factor and I-R hybrid dysgenesis in *Drosophila melanogaster*.
in Mobile DNA (ed Howe, M.M.) ASM Press,Washington D.C
pp 503-517.

Fischer, J.A., Giniger, E., Maniatis, T., & Ptashne, M., (1988)
GAL4 activates transcription in *Drosophila.*
Nature *332*, 853-855.

Foster, J.M., Yudkin, B., Lockyer,A.E. and Roberts,D.B., (1995)
Cloning and sequence analysis of GmII, a *Drosophila melanogaster* homologue of the cDNA encoding murine Golgi alpha-mannosidase II
Gene *154*, 183-186.

Freeman, M., and Glover, D. M., (1987)
The *gnu* mutation of *Drosophila* causes inappropriate DNA synthesis in mutant and fertilised eggs.
Genes Dev. *1*, 924-930.

Frigerio, G., Burri, M., Bopp, D., S. Baumgartner & M. Noll (1986)
Structure of the segmentation gene *paired* and the *Drosophila PRD* gene set as part of a gene network.
Cell *47*, 735-746.

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

Fryxell, K.J., and Meyerowitz, E.M., (1987)
An opsin gene that is expressed only in the R7 photoreceptor cell of *Drosophila*
EMBO J., *6*,443-451.

Fuchs, R. (1991)
MacPattern: Protein pattern searching on the Apple Macintosh.
CABIOS *7*, 105-106.

Fujita, S. C., Zipursky, S.L., Benzer, S., Ferrus, A., and Shotwell, S.L., (1982)
Monoclonal antibodies against the *Drosophila* nervous system.
Proc. Natl. Acad. Sci. (USA) *79*, 7929-7933.

Fyrberg, E.A., Bond, B.J., Hershey, N.D., Mixter, K.S., and Davidson, N., (1981)
The Actin genes of *Drosophila*: protein coding regions are highly conserved but intron positions are not.
Cell *24*, 107-116.

Gerber, H-P., Seipel, K., Georgiev, O., Hofferer, M., Hug, M., Rusconi, S., and Schaffner, W., (1994)
Transcriptional activation modulated by homopolymeric glutamine and proline stretches.
Science *263*, 808-811.

Giniger, E., Wells, W., Jan, L.Y., and Jan, Y.N., (1993).
Tracing neurions with a kinesin-β-galactosidase fusion protein.
Roux's Arch. Dev. Biol. *202*, 112-122.

Gloor,G.B., Nassif, N.A., Johnson-Schlitz, D.M., Preston, C.R., and Engels, W.R., (1991)
Targeted gene replacement on *Drosophila* via P-element induced gap repair.
Science *257*, 1110-1117.

Goodwin, S.F., Tully, T., and Kaiser, K., (1995)
Disruption of learning in mutants of the Drosophila Protein Kinase type II regulatory subunit gene.
Neuron *(submitted)*

Gubler U., and Hoffman, B.J., (1983)
A simple and very efficient method for generating cDNA libraries.
Gene *25*, 263-269.

Hall, J.C., (1979)
Control of male reproductive behaviour by the central nervous system in *Drosophila*: dissection of a courtship pathway by genetic mosaics.
Genetics *92*, 437-457.

Hall, L., and Spierer, P., (1986)
The *Ace* locus of *Drosophila*: structural gene for acetylcholinesterase with an unusual 5' leader.
EMBO.J. *5*, 2949-2954.

Hamilton, B.A., Palazzolo, M.J., Chang, J.H., VijayRaghavan, K., Mayeda, C.A., Whitney, M.A., and Meyerowitz, E.M., (1991a)
Large scale screen for transposon insertions into cloned genes.
Proc. Natl. Acad. Sci. (USA) *88*, 2731-2735.

Hamilton, B.A., Palazzolo, M.J., and Meyerowitz, E.M., (1991b)
Rapid isolation of long cDNA clones from existing libraries.
Nucl. Acids. Res. *19*, 1951-1952.

Hanzlik, T. N., Abdel-Aal, Y. A. I., Harshman, L. G., Hammock, B. D., (1989)
Isolation and sequencing of cDNA clones coding for the Juvenile Hormone Esterase from *Heliothis virescens*.
J. Biol. Chem. *264*, 12419-12425.

Harbison, D.T., (1995)
Male-specific transcripts from *Drosophila melanogaster.*
Ph.D thesis, University of Glasgow.

Hartley, D.A., Xu, T. and Artavanis-Tsakonas, S., (1987)
The embryonic expression of the *Notch* locus of *Drosophila melanogaster* and the implications of point mutations in the extra cellular EGF-like domain of the predicted protein.
EMBO J. *6*, 3407-3417.

Hazelrigg, T., and Tu, C., (1994)
Sex-specific processing of the *Drosophila exuperantia* transcript is regulated in male germ cells by the *tra-2* gene.
Proc. Natl. Acad. Sci. (USA) *91*, 10752-10756.

Henikoff, S., (1984).
Unidirectional digestion with Exonuclease III creates targeted breakpoints for DNA sequencing.
Gene *28*, 351-359.

Higgins, D. G., and Sharpe, P. M., (1988)
CLUSTAL: a package for performing multiple sequence alignments on a microcomputer.
Gene *73*, 237-244.

Hoffman, K., (1993)
BOXSHADE 2.7, a program for shading multiple aligned sequence files.
available from ftp.embl.heidelberg.de (boxshade.zip).

Hosoya M., Onda H., Ogi K., Masuda Y., Miyamoto Y., Ohtaki T., Okazaki H., Arimura A., and Fujino M., (1992)
Structure of the human pituitary adenylate cyclase activating polypeptide (PACAP) gene.
Biochem. Biophys. Acta. *1129*, 199-206.

Hotta, Y., and Benzer, S., (1970)
Genetic dissection of the *Drosophila* nervous system by means of mosaics.
Proc. Natl. Acad. Sci. (USA) *67*, 1156-1163.

Hyde, D.R., Mecklenburg, K.L., Pollock, J.A., Vihtelic, T,S., and Benzer, S., (1990)
Twenty *Drosophila* visual system clones: one is a homolog of human arrestin.
Proc. Natl. Acad. Sci. (USA) *87*, 1008-1012.

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

Jurgens, G., Wieschaus, E.; Nusslein-Volhard, C. and Kluding,H.(1984)
Mutations afecting larval cuticle in *Drosophila melanogaster.*
Roux Arch. Dev. Biol. *193*, 83-295.

Kaiser, K., and Goodwin, S. F. (1990)
'Site-selected' transposon mutagenesis of *Drosophila.*
Proc. Natl. Acad. Sci. (USA) *87*, 1686-1690.

Kalfayan, L., and Wensink, P.C., (1982)
Developmental regulation of the *Drosophila* α-tubulin genes.
Cell *29*, 91-98.

Kamb, A., Weir, M.A., Rudy, B., Varmus, H., and Kenyon, C., (1989)
Identification of genes from pattern formation, tyrosine kinase, and potassium channel families by DNA Amplification.
Proc. Natl. Acad. Sci. (USA) *86*, 4372-4376.

Karpen, G. H., and Spradling, A.C., 1992
Analysis of subtelomeric heterochromatin in the *Drosophila* minichromosome Dp1187 by single P-element insertional mutagenesis.
Genetics *132*, 737-753.

Katzen,A.L., Montarras, D., Jackson, J., Paulson, R.F., Kornberg, T.B., and Bishop, J.M. (1991)
A gene related to the proto-oncogene *fps/fes* is expressed at diverse times during the life cycle of *Drosophila melanogaster*.
Mol. Cell Biol. *11*, 226-239.

Kemphues, K.J., Kaufman, T. C., Raff, R. A., and Raff, E.C., (1982)
The testis-specific beta-tubulin subunit in *Drosophila melanogaster* has multiple functions in spermatogenesis.
Cell *31*, 655-670.

Kidwell, M.G., (1987)
A survey of the success rates using P-element mutagenesis in *Drosophila melanogaster*.
Dros. Inf. Serv. *66*, 81-86.

Kimura, M., (1980)
A simple method for estimating evolutionary rate of base substitutions through comparitivce studies of nucleotide sequences.
J. Mol. Evol. *16*, 111-120.

Klingenberg, M., (1990)
Mechanism and evolution of the uncoupling protein of brown adipose tissue.
TIBS *15*, 108-112.

Konopka, R., Wells, S. and Lee, T., (1983)
Mosaic analysis of a *Drosophila* clock mutant.
Molec. Gen. Genet. *190*, 284-288.

Kotewicz, M. L., D'Alessio, J.M., Driftmeir, K. M., Blodgett, ,K. P., and Gerard, G.F., (1985)
Cloning and overexpression of Moloney-Murine Leukemia Virus in *Escherichia coli.*
Gene *35*, 249-258.

Kotewicz, M. L., Sampson, C.M., D'Alessio, J.M., and Gerard, G.F. (1988)
Isolation of cloned Moloney Murine Leukaemia Virus reverse transcriptase lacking ribonuclease H activity.
Nucl. Acids. Res. *16*, 265-277.

Krejci, E., Duval, N., Chatonnet, A., Vincens, P., and Massoulie, J., (1991)
Cholinesterase-like domains in enzymes and structural proteins: Functional and evolutionary relationships and identification of a catalytically essential aspartic acid.
Proc. Natl. Acad. Sci. (USA) *88*, 6647-6651.

Levin, L. R., Han, P. L., Hwang, P. M., Feinstein, P. G., Davis, R. L., and Reed, R. R., (1992)
The *Drosophila* learning and memory gene, *rutabaga*, encodes a Ca_2^+/ Calmodulin-responsive adenylyl cyclase.
Cell *68*, 479-489.

Levy, W.B., and McCarthy, B.J.,(1975)
Messenger RNA complexity in *Drosophila melanogaster*
Biochemistry *14*, 2440-2446.

Levy, L. S., and Manning, J. E. (1981)
Messenger RNA sequence complexity and homology in developmental stages of *Drosophila.*
Dev. Biol. *85*, 141-149.

Levy L.S.,and Manning,J.E., (1982)
Expression of a set of head specific genes during *Drosophila* development.
Dev. Biol. *94*, 465-476.

Levy, L. S., Ganguly, R., Ganguly, N., and Manning, J. E. (1982)
The selection, expression, and organisation of a set of head-specific genes in *Drosophila* .
Dev. Biol. *94*, 451-464.

Lin, H., and Spradling, A.C., (1992)
Analysis of germline stem cell (GSC) division in *Drosophila*.
A. *Dros*. Ref. Conf. *33*, 8.

Lindsley, D.L., and Zimm, G.G., (1992)
The Genome of *Drosophila melanogaster.*
CSHL Press, New York.

Littleton, J.T., Stern, M., Schuize, K., Perin, M., and Bellen, H.J., (1993)
Mutational analysis of *synaptotagmin* demonstrates its essential role in Ca_2^+ activated neurotransmitter release.
Cell *74*, 1125-1134.

Livingstone, M. S., Sziber, P. P., and Quinn, W. G., (1984)
Loss of Calcium Calmodulin responsiveness in adenylate-cyclase of *rutabaga*, a *Drosophila* learning mutant.
Cell *37*, 205-215.

Louvi, A., and Tsitilou, S.G., (1992)
A clone encoding the ADP/ATP translocase of *Drosophila melanogaster* shows a high degree of similarity with the mammalian ADP/ATP translocases.
J. Mol. Evol *35*, 44-50.

Macdonald, P.M., (1992)
The *Drosophila pumilio* gene: an unusually long transcription unit and an unusual protein.
Development *114*, 221-232.

Mackenzie, C.R., Stewart, B., and Kaiser, K. (1989)
A device for creating and replicating ordered arrays of plaques or colonies.
Technique *1*, 49-53.

McClean, C., Bucheton, A., and Finnegan, D.J., (1993)
The 5' untranslated region of the I factor, a long interspersed nuclear element like retrotransposon of *Drosophila melanogaster,* contains an internal promoter and sequences that regulate expression.
Mol. Cell Biol. *10*, 1042-1050.

McLelland, M., and Nelson, M., (1988)
The effect of site-specific DNA methylation on restriction endonucleases and DNA methyltransferases - a review.
Gene *74*, 291-304.

Matsuo, Y., and Yamakazi, T., (1989)
tRNA derived insertion element in histone gene repeating unit of *Drosophiola melanogaster*.
Nucl. Acids Res. *17*, 225-238.

Mercken, L., Simons, M.J., Swillens, S.J., Massaer, M., and Vassart, G., (1985)
Primary structure of bovine thyroglobulin deduced from the sequence of its 8,431-base complementary DNA.
Nature 316, 647-651.

Miller, C.A., and Benzer, S., (1983)
Monoclonal Antibody cross reactions between *Drosophila* and human brain.
Proc. Natl. Acad. Sci. (USA) *80*, 7641-7645.

Milligan, C. D., and Kaiser, K., (1993)
'Site-selected' mutagenesis of a *Drosophila* gene using the I factor retrotransposon.
Nucl. Acids Res. *21*, 1323-1324.

Montell, C., Jones, K., Hafen, E., and Rubin, G.M., (1985)
Rescue of the *Drosophila* mutation *trp* by germline transformation.
Science *230*, 1040-1043.

Mouches, C., Pauplin, Y., Munna, A., Lemieux, L., Herzog, M., Abadon, M., Beysatt-Arnaouty, V., Hyrein, O., Robert de Saint Vincent, B., Georghiou, G.P., and Pasteur, N., (1990)
Characterisation of amplification core and esterase B1 gene responsible for insecticide resistance in *Culex*.
Proc. Natl. Acad. Sci. (USA) *87*, 2574-2578.

Murray, N.E., (1983)
Phage Lambda and Molecular Cloning.
in Lambda II, (eds Hendrix, R.W., Roberts, J.W., Stahl, F.W., & Weisberg, R.A.) CSH Press.

Natzle, J.E., Fristrom, D.K., and Fristrom, J.W., (1988)
Genes expressed during imaginal disc morphogenesis: IMP-E1, a gene associated with epithelial cell rearrangement.
Dev. Biol. *129*, 428-438.

Neuman-Silberberg, F.S., Schejter, E., Hoffmann, F.M., and Shilo, B.Z., (1984)
The *Drosophila* ras oncogenes: structure and nucleotide sequence.
Cell *37*, 1027-33.

Nikolnikov, S., Posfai, G., and Sain,B., (1984)
The construction of a versatile plasmid vector that allows direct selection of fragments cloned into six unique sites of the *c*I gene of coliphage *434*.
Gene *30*, 261-265.

Nusslein-Vollhard, C., and Weischaus, E., (1980)
Mutations affecting segment number and polarity in *Drosophila*.
Nature *287*,795-801.

Oakeshott, J. G., Collet, C., Phillis, R. W., Nielsen, K.M., Russell, R. J., Chambers, G. K., Ross, V., and Richmond, R.C., (1987)
Molecular cloning and charactersisation of esterase-6, a serine hydrolase of *Drosophila*.
Proc. Natl. Acad. Sci. (USA) *84*, 3359-3363.

O'Connell, P., and Rosbash, M. (1984)
Sequence, structure, and codon preference of the *Drosophila* ribosomal protein 49 gene.
Nucl. Acids Res. *12*, 5495-5513.

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.

Ohsako, S., Nishida, Y., Ryo, H., and Yamauchi, T. (1993)
Molecular Characterisation and Expression of the *Drosophila* Ca_2^+/Calmodulin-dependent protein kinase II gene.
J. Biol. Chem. *268*, 2052-2062.

O'Kane, C.J., and Gehring, W.J., (1987)
Detection *in situ* of regulatory elements in *Drosophila*.
Proc. Natl. Acad. Sci. (USA) *84*,9123-9127.

O'Tousa, J.E., Baehr, W., Martin, R.L., Hirsh, J., Pak, W.L., and Applebury, M.L., (1985)
The *Drosophila ninaE* gene encodes an opsin.
Cell *40*, 839-850.

Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S., and Sharp, R.A., (1986)
Splicing of messenger RNA precursors.
Ann. Rev. Biochem. *55*, 1119-1150.

Pak, W.L., Grossfield, J., White, N.V., (1969)
Non-phototactic mutants in study of vision in *Drosophila*.
Nature *222*, 351-356.

Palazzolo, M. J., Hyde, D. R., VijayRaghavan, K., Mecklenburg, K., Benzer, S., and Meyerowitz, E., (1989)
Use of a new strategy to isolate and characterise 456 *Drosophila* cDNA clones corresponding to RNAs detected in head but not in early embryos.
Neuron *3*, 527-539.

Pardue, M. L. (1986)
in situ hybridization to DNA of chromosomes and nuclei.
in '*Drosophila*, a practical approach' Oxford, IRL Press.

Paro, R., and Hogness, D.S., (1991)
The Polycomb protein shares a homologous domain with a heterochromatin associated protein of *Drosophila.*
Proc. Natl. Acad. Sci. (USA) *88*, 263-267.

Patel, N. H., and Goodman, C. S. (1992)
Detection of *even-skipped* transcripts in *Drosophila* embryos with PCR/DIG-labeled DNA probes.
in 'Nonradioactive *in situ* hybridization application manual'
Boehringer Mannheim Biochemica.

Pearson, W. R., and Lipman, D.J., (1988)
Improved tools for biological sequence comparison.
Proc. Natl. Acad. Sci. (USA)*85*, 2444-2448.

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

Pflugfelder, G.O., Roth, H., Poeck, B., Kerscher, S., Schwarz, H., Jonschker, B., and Heisenberg, M., (1992)
The lethal(1)optomotor blind gene of *Drosophila melanogaster* is a major organiser of optic lobe development: Isolation and characterisation of the gene.
Proc. Natl. Acad. Sci. (USA) *89*, 1199-1203.

Phelps, A., Schobert, C.T., and Wohlrab, H., (1991)
Cloning and characterisation of the mitochondrial phasphate transport protein gene from the Yeast *Saccharomyces cerevisiae.*
Biochemistry *30*, 248-252.

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

Puissant, C., and Houdebine, L.-M., (1990)
An improvement of the Single-Step method of RNA isolation by Acid-Guanidinium Thiocyanate-Phenol-Chloroform Extraction.
Biotechniques *8*, 148-149.

Quinn, W.G., Harris, W.A., and Benzer, S., (1974)
Conditioned Behaviour in *Drosophila melanogaster.*
Proc. Natl. Acad. Sci. (USA) *71*, 708-712.

Quinn, W. G., and Dudai, Y. (1976)
Memory phases in *Drosophila melanogaster.*
Nature *262*, 576-577.

Quinn, W. G., Sziber, P. P., and Booker, R., (1979)
The *Drosophila* memory mutant *amnesiac.*
Nature *277*, 214-216.

Rasmusson,K., Serr, M., Gepner, J., Gibbons, I., and Hays, T.S., (1994)
A family of dynein genes in *Drosophila melanogaster*.
Mol. Biol. Cell *5*, 45-55.

Ray, K., and Ganguly, R., (1992)
The *Drosophila* G protein γ subunit (D-$G_{\gamma}1$) produces three developmentally regulated transcripts and is predominantly expressed in the Central Nervous System.
J. Biol. Chem. *267*, 6086-6092.

Reinke, R., Krantz, D.R., Yen, D., Zipursky, S.L., (1988)
chaoptin, a cell surface glycoprotein requires for *Drosophila* photoreceptor cell morphogenesis, contains a repeat motif found in yeast and humans.
Cell *52*, 291-301.

Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P., (1977)
Labelling DNA to high specific activity *in vitro* by Nick Translation with DNA Polymerase I.
J.Mol.Biol. *113*, 237-251.

Robertson, H.M., Preston, C.R., Phillis, R.W., Johnson-Schlitz, D., Benz W.K., and Engels, W.R., (1988)
A stable genomic source of P-element transposase in *Drosophila melanogaster*.
Genetics *118*, 461-470.

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

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

Runswick, M.J., Powell, S.J., Nyren, P., and Walker, J.E., (1987)
Sequence of the bovine mitochondrial phosphate carrier protein: structural relationship to ADP/ATP translocase and brown fat mitochondria uncoupling protein.
EMBO. J. *6*, 1367-1373.

Russel, M., Kidd, S., and Kelley, M.R., (1986)
An improved filamentous helper phage for generating single stranded plasmid DNA.
Gene *45*, 333-338.

Russell, S.R.H., (1989)
Identification of male-specific transcripts from *D. melanogaster.*
PhD thesis, University of Glasgow.

Saiki, R., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R. G., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988)
Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase.
Science *239*, 487-491.

Sambrook, J., Fritsch, E. F., and Maniatis, T., (1989)
Molecular Cloning: A laboratory manual (Second ed.).
Cold Spring Harbor Laboratory Press.

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

Sargent, T.D., and Dawid, I.B., (1983)
Differential gene expression in the gastrula of *Xenopus laevis.*
Science, *222*, 135-139.

Schatz, G., and Butow, R.A., (1983)
How are proteins imported into mitochondria.
Cell *32*, 316-318.

Scherer, G., Telford, J., Baldari, C., and Pirrotta, V., (1981)
Isolation of cloned genes differentially expressed at early and late stages of *Drosophila* embryonic development.
Dev. Biol. *86*, 438-447.

Schuh, R., Aicher, W., Gaul, U., Cote, S., Preiss, A., Maier, D., Seifert, E., Nauber, U., Schroder, G., Kemier, R., and Jackle, H., (1986)
A conserved family of nuclear proteins containing structural elements of the finger protein encoded by *kruppel*, a *Drosophila* segmentation gene.
Cell *47*, 1025-1032.

Shieh, B-H., Stamnes, M.A., Seavello, S., Harris, G.L., and Zuker, C.S., (1989)
The *ninaA* gene required for visual transduction in *Drosophila* encodes a homologue of the cyclosporin A binding protein.
Nature *338*, 67-70.

Short, J.M., Fernandez, J.M., Sorge, J.A., and Huse, W.D., (1988)
λZAP: a bacteriophage λ expression vector with *in vivo* excision properties.
Nucl. Acids Res. *16*, 7583-7600.

Shortridge, R.D., Yoon, J., Lending, C.R., Bloomquist, B.T., Perdew, M.H., and Pak, W.L., (1991)
A *Drosophila* Phospholipase C gene that is expressed in the central nervous system.
J. Biol. Chem. 266, 12474-12480.

Soeller, W. C., Oh, C. E., and Kornberg, T. B., (1993)
Isolation of cDNAs encoding the *Drosophila* GAGA transcription factor.
Mol. Cell Biol. *13*, 7961-7970.

Spackman, M.E., Oakeshott, J.G., Smyth, K-A., Medveczky, K.M., and Russell, R.J., (1994)
A cluster of esterase genes on chromosome 3R of *Drosophila melanogaster* includes homologues of esterase genes conferring insecticide resistance in *Lucilia cuprina*.
Biochem. Genet. *32*, 39-62.

Spofford, J. B., (1967)
Single-locus modification of position-effect variegation in *Drosophila melanogaster*. I. White variegation.
Genetics *57*, 751-766.

St. John, P., and Davis, R.W., (1979)
Isolation of galactose-inducible DNA sequences from *Saccharomyces cerevisiae* by differential plaque filter hybridisation.
Cell *16*, 443-452.

Suzuki Y., Nogi, Y., Abe, A., and Fukasawa, T., (1988)
GAL11 protein, an auxiliary transcription factor for genes encoding galactose-metabolising enzymes in *Saccharomyces cerevisiae*.
Mol. Cell Biol. *8*, 4991-4999.

Sved, J., 1986
eyes absent (*eya*)
Dros. Inf. Serv. *63*, 169.

Swanson, K. D. and Ganguly, R., (1992)
Characterisation of a *Drosophila melanogaster* gene similar to the mammalian genes encoding the tyrosine tryptophan-hydroxylase activator and protein-kinase-C inhibitor proteins.
Gene *113*, 183-190.

Tautz, D., and Pfeiffle, C. (1989)
A non-radioactive *in situ* hybridisation method for the localization of specific RNAs in *Drosophila* embryos reveals a translational control of the segmentation gene *hunchback*.
Chromosoma (Berl.) *98*: 81-85.

Tautz, D. (1992)
Localisation of the expression of the segmentation gene *hunchback* in *Drosophila* embryos with DIG-labeled DNA probes.
in 'Nonradioactive *in situ* hybridization application manual'
Boehringer Mannheim Biochemica.

Taylor, R.G., Walker, D.C., and McInnes, R.R., (1993)
E. coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing.
Nucl. Acids Res. *21*, 1677-1678.

Tomlinson, S.R., (1994)
A screen for *Drosophila* genes relevant to the nervous system.
PhD thesis, University of Glasgow.

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., and Reuter, G., (1994)
The protein encoded by the *Drosophila* position effect variegation gene *Su(Var)3-9* combines domains of antagonistic regulators of homeotic gene complexes.
EMBO. J., *13*, 3822-3831.

Van Vactor, D., Krantz, D.E., Reinke, R., and Zipursky, S.L., (1988)
Analysis of mutants in *chaoptin*, a cell specific glycoprotein in *Drosophila* reveals its role in cellular morphogenesis.
Cell *52*, 281-290.

Wharton, K.A., Johansen, K.M., Xu, T. and Artavanis-Tsakonas, S., (1985)
Nucleotide sequence from the neurogenic locus *Notch* implies a gene product that shares homology with proteins containing EGF-like repeats,
Cell *43*, 567-581.

Whittaker, P.A., Campbell, A.J.B., Southern, E.M., and Murray, N.E. (1988)
Enhanced recovery and restriction mapping of DNA fragments cloned in a new λ vector.
Nucl. Acids Res. *16*,6725-6736.

Wickens, M., and Stephenson, P., (1984)
Role of the conserved AAUAAA sequence: four AAUAAA point mutations prevent mRNA 3' end formation.
Science *226*, 1045-1051.

Yannish-Perron, C., Viera, 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.

Zipursky, S.L., Venkatesh, T.R., Teplow, D.B., and Benzer, S., (1985)
Neuronal development in the *Drosophila* retina: monoclonal antibodies as molecular probes.
Cell *36*, 15-26.

Zuker, C.S., Cowman, A.F., and Rubin, G.M., (1985)
Isolation and structure of a Rhodopsin gene from *D. melanogaster*
Cell *40*, 851-858.

FlyBase: http://morgan.harvard.edu/
A database of information on *Drosophila melanogaster* initiated at the University of Indiana, School of Biological Sciences.

Promega '*Protocols and Applications Guide*'.
2nd edition, (1991)
Promega Corp., USA